ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN PSICOLOGIA

Curriculum Neuroscienze Cognitive

Ciclo XXXVI

Settore Concorsuale: 11/E1

Settore Scientifico Disciplinare: M-PSI/02

The role of the ventromedial prefrontal cortex in self- and event-related schemata

Presentata da: Debora Stendardi

Coordinatrice Dottorato

Prof.ssa Elisabetta Crocetti

Supervisore

Prof.ssa Elisa Ciaramelli

Esame finale anno 2024

Acknowledgements

A PhD is always a journey. It could feel like strolling by the riverside, hand in hand with your loved one, listening to the calm melody of the most colourful chirping birds. But sometimes it could feel like climbing the Everest, having lost you walking stick and hiking map, surrounded by a cold fog, and unable to see what's in front of you. And maybe your tent and backpack are on fire. Most of the time though, a PhD will feel like both of those things, and everything in between. So, here, I feel the need to thank and acknowledge all the people who helped in dispelling the fogs I've encountered throughout this long journey.

First and foremost, I want to thank my supervisor, Professor Elisa Ciaramelli. You have always showed me the way when I needed it, but also allowed me to choose my own path when I asked for it. You have always treated me with intellectual respect and kindness, guiding me though this journey, which has been much more sunny than foggy, thanks to you.

I want to thank Alessandro Treves. Never in my life I've felt more welcome as when I came to SISSA, by both you and the LIMBO group. You have taught me so many things, always with the uttermost gentleness, wisdom, and brilliance. I could never imagine a better place than SISSA to learn and grow as a scientist, and for me, you have been largely responsible for it.

A special thank you to my colleague and companion through this long journey, Giovanni.

I wish to thank Michela Candini and Caterina Bertini, for their precious help in these final stages of my PhD.

I want to thank the brilliant scientists from the LIMBO group at SISSA, that I now have the immense pleasure of calling my friends: Anindita and Kwang Il.

Thanks to Chiara, whose conversations with me have always been invaluably enlightening.

Thanks to Gioele. It seems impossible for me to even choose something specific to thank you for, but know that I will forever be grateful for everything you have done for me. Thank you.

Thanks to my Mum, my rock and anchor to sanity.

Thanks to my brother: making you proud of me has been (and still is) one of my strongest driving forces.

I wish to end this section by thanking someone that is seldom ever recognized in neuropsychological research, but here I feel the need to acknowledge them: all the patients I have ever tested during my PhD. It is fairly obvious that without you there would be no research, no paper, no thesis. My sessions with you have been funny, interesting, long, and sometimes even emotional. I wish to thank all of you, and sorry if sometimes I might have bored you to death.

To Raja,

"*Until one has loved an animal, a part of one's soul remains unawakened.*"

Anatole France

Abstract

The ventromedial prefrontal cortex (vmPFC) is one of the largest regions of the human brain. Its activity has been linked to a plethora of cognitive functions, such as self-referential cognition, mind wandering, memory recollection, and many more. Yet, we still lack an exhaustive understanding of the essence of its functional specialization, if any. The schema theory proposes that vmPFC's overarching role lies in the activation of schematic knowledge in neocortex, which is then used by other brain regions, according to environmental demands. Therefore, the present thesis addresses the question of whether a vmPFC damage degrades schema-mediated cognition. In Chapter 1 I begin to investigate the role of vmPFC in imparting the memory advantage for self-referential information in memory. Whilst healthy and brain-damaged controls exhibit superior recall for self- (vs other-) related items, such advantage is proven absent in vmPFC patients, suggesting a degradation of the self-schema. In Chapter 2 I aim to clarify whether this lack of self-referential prioritization stems from vmPFC patients' memory deficits, or from an impairment in self-knowledge itself. Whilst healthy and brain-damaged controls exhibit more consistent self- rather than other-referential ratings over time, vmPFC patients do not, again suggesting a disturbance of the self-schema. In Chapter 3 we investigate vmPFC's involvement in activating (reinstatement) and using (instantiation) event-schemata, demonstrating that a vmPFC damage hinders schema reinstatement, wherein vmPFC patients reinstate incomplete, nebulous schemas. Finally, in Chapter 4 we model frontal and posterior cortical interactions in a hybrid Potts model of cortical dynamics, revealing a capacity of the frontal cortex to act as the source of predominant influence on latching dynamics. We interpret the result considering the schematic influence exerted from the frontal lobe on posterior brain regions observed in experimental practice. Finally, by modelling a frontal lesion, we reproduce event construction deficits of vmPFC damaged patients.

Table of contents

General introduction

The human brain contains approximately 86 billion neurons, communicating to each other at any given time through roughly 100 trillion connections (Zimmer, 2011; Lent et al., 2012). It is almost ironic that neuroscientists are sometimes accused of *reducing* human behaviour to mere biology (Ayala, 1987). The word "reducing" implies that we successfully pass from something extremely complex, to something simpler and more manageable. I do not know about every neuroscientist's opinion on the matter, but I am fairly confident that no neuroscientist ever, whilst studying the brain, has thought, "this is simple enough, we will solve it in no time". One of the most unsettling and disconcerting thoughts that at some point have touched my (and probably the reader's) mind is that each and every experience, emotion, perception and feeling that we have ever lived, and will ever live, is the result of the activity of those 86 billion neurons. No wonder studying the brain is far from being a simple task. However, once we start to move from the uncomfortability of that feeling, the itch to understand something more about what those extremely complicated cells are doing in our mind takes place.

Since we have established that, in this thesis, we are staying away from simple matters, I will be focusing my dissertation on the (arguably) most complicated portion of the brain: the frontal lobes. The frontal lobes make up approximately a third of the human cerebral cortex (Clark et al., 2010), and are the last to fully develop, reaching maturation by the age of 25 (Teffer & Semendeferi, 2012; Sharma et al., 2013). Specifically, I will be focusing my investigation on a specific part of the frontal lobes, the ventromedial prefrontal cortex (vmPFC). As we will see, vmPFC is one of the biggest brain regions of the human cortex, and understanding its functions has proven to be a titanic challenge for neuroscience (Myers-Schulz & Koenigs, 2012; Delgado et al., 2016; Hiser & Koenigs, 2018). I will thus start by giving an

overview of the anatomy and connectivity of this big, complicated brain region, and then I will move onto the cognitive processes linked to it. Finally, I will close this section by reviewing the theory that constitute the general theoretical framework of the present elaborate, and that in my (and many other researchers) opinion best explains what the vmPFC is fundamentally concerned with: schemata.

1. Anatomy and connectivity of vmPFC

1.1. Anatomy of the ventromedial prefrontal cortex

The ventromedial prefrontal cortex (vmPFC) is situated along the inferior part of the medial wall of the frontal lobe, incorporating the medial segment of the orbital frontal cortex. Medially, its posterior boundary is defined by the subcortical septal region, extending anteriorly to the frontal pole. On the ventral part, vmPFC stretches onto the orbital surface of the frontal cortex, reaching the medial orbital sulcus. The orbital section reaches its posterior limit by the primary olfactory cortex (Mackey & Petrides, 2014). Notably, anatomical classifications of vmPFC are varied and not always agreed upon in the literature. It is not rare for authors to state how the exact boundaries of vmPFC are not always clearly defined, and often depend on which classification one refers to (Mackey & Petrides, 2014; Schneider & Koenigs, 2017; Hiser & Koenigs, 2018; Alexander et al., 2023).

However, measurable variations in the density of cortical layers IV and Va in post-mortem human brains have allowed the identification of several sub-regions within the ventromedial prefrontal cortex (Delgado et al., 2016; Bhanji et al., 2019). Öngür and Price

(2000) delineated two networks within the human orbital and medial prefrontal cortex: the orbitofrontal network, encompassing Brodmann Areas (BAs) 12 and 13, and the medial frontal network, which includes BAs 9, 10, 11, 24, 25, and 32 (see also Nieuwenhuis & Takashima, 2011). Note that, in literature, some variations can be traced in the exact Brodmann'a areas that are considered part of vmPFC, sometimes even regarding parcellations within BAs (as an example, see Fig.1).

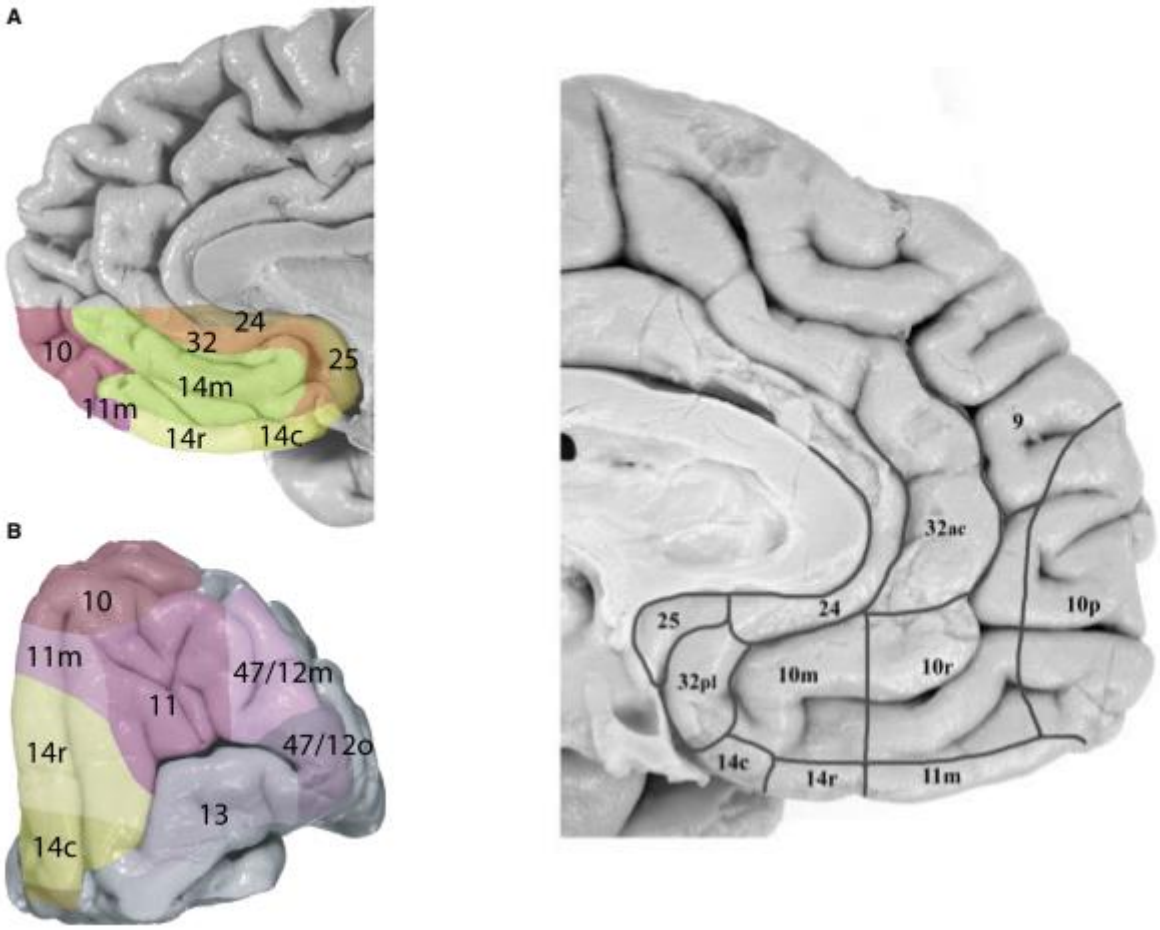

Figure 1. Architectonic parcellation of vmPFC. On the left, the ventral medial frontal surface (A) and (B) the orbital frontal surface (Mackey & Petrides, 2010). On the right, the nomenclature of vmPFC subdivisions by Ongur et al. (2003).

Findings from previous research have produced a valuable map that extends the classification employed in the commonly used Brodmann map. Distinct spatial patterns in

histology are evident within the vmPFC zone. Specifically, within vmPFC, there is a notable variation in histological features, in which areas closer to the medial aspect display a higher density of layer IV granule cells, while medial regions exhibit greater layer Va pyramidal cell density compared to their lateral counterparts on the orbital surface (Bhanji et al., 2019). As we will see in the next section, different anatomical subdivisions also correspond to different patterns of connectivity, both within vmPFC, and between vmPFC and other brain regions (Jackson et al., 2020).

1.2. **Connectivity of the ventromedial prefrontal cortex**

A comprehensive nomenclature of the structural connectivity of vmPFC stems from neuroanatomical tract-tracing techniques applied to homologous areas in nonhuman primate brains (Barbas, & Pandya, 1989; Croxson, 2005; Yeterian et al., 2012) . However, when such results are compared to the human brain through non-invasive imaging methods, substantial evidence strongly supports broad similarities (Ongur & Price, 2000; Mackey & Petrides, 2010; Wallis, 2012; Jbabdi et al., 2013; Schaeffer et al., 2020). Observably, the ventromedial prefrontal cortex exhibits distinctive connectivity characteristics. Unlike the lateral sections of the orbitofrontal cortex, vmPFC receives minimal direct inputs from primary sensory regions, and in contrast to lateral prefrontal regions, it demonstrates weak connections with the motor cortex (Ongur & Price, 2000; Bhanji et al., 2019). Noteworthy are the major outputs from vmPFC to various brain regions, including the hypothalamus, periaqueductal grey, amygdala, hippocampus, nucleus accumbens, and superior temporal cortex (Vianna & Brandão, 2003; Wallis, 2012; Gluth et al., 2015; Motzkin et al., 2015; Zhang et al., 2018). Also, through the

use of diffusion tensor imaging, researchers have been able to identify long-range connections between vmPFC and the posterior cingulate cortex (Greicius et al., 2009).

The vmPFC is anatomically connected to the hippocampus through three primary reciprocal pathways: the uncinate fasciculus, the fornix, and the cingulum bundle (Concha et al., 2005; Malykhin et al., 2008; Catani et al., 2013). In addition, a fourth indirect pathway links the vmPFC to the hippocampus through the mammillo-thalamic tract and anterior thalamic projections (McCormick et al., 2018a).

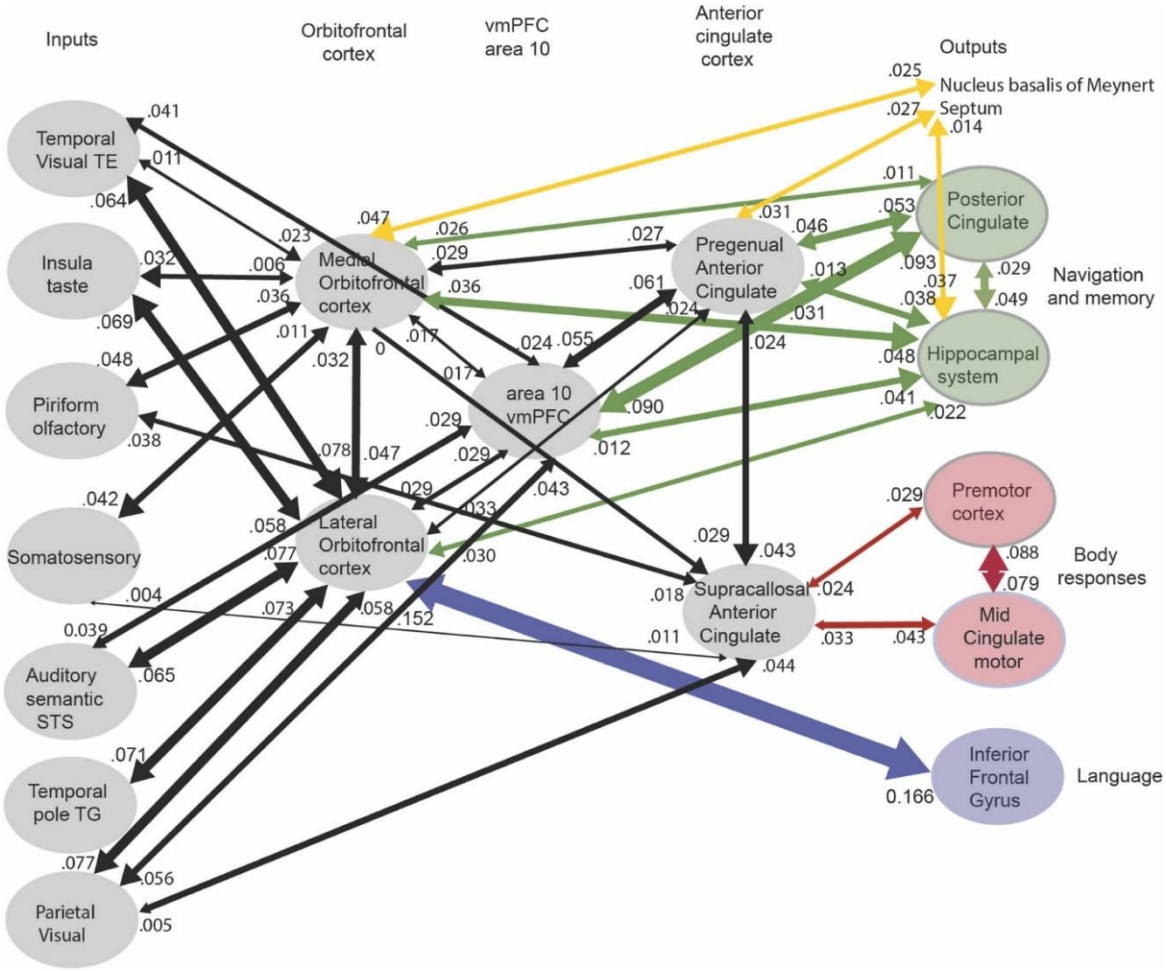

Figure 2. Effective connectivity of the vmPFC as measured by the Human Connectome Project (image from Rolls, 2022).

As mentioned above, differences in structural connectivity within sub-regions of vmPFC are also evident, with more pronounced projections from the amygdala to posterior

areas (more precisely areas 24 and 25) compared to the more anterior regions (Price & Drevets, 2010). Ventral areas of vmPFC exhibit stronger connections with ventral and medial areas of the striatum, such as the nucleus accumbens, while more dorsal areas of vmPFC connect with anterior and dorsal regions of the striatum (Lehéricy et al., 2004; Haber & Knutson, 2010). Moreover, the projections from vmPFC to the hypothalamus are most prominent from its posterior areas, particularly area 25 (Price & Drevets, 2010). Thus, a discernible pattern emerges, indicating variations in connectivity between anterior and posterior areas of vmPFC, aligning with the observed differences in histology along the anterior-posterior axis.

Now that we are starting to better qualify the complexity of the anatomy and connectivity patterns of the ventromedial prefrontal cortex, in the next section, we will see that, as everything else in the human brain, vmPFC is constantly working in concert with other brain areas to give rise to a multitude of cognitive processes. Specifically, vmPFC is part of one of the most mysterious but fascinating networks of the human brain: the Default Mode Network.

1.3. The Default Mode Network

The vmPFC is part of a complex and interconnected set of brain regions, called the Default Mode Network (DMN), which is by now as famous amongst neuroscientists as it is poorly understood. The DMN derives its name from the fact that when the brain is not concerned with an external task, its activity defaults to it (Shulman et al., 1997; Raichle, & Snyder, 2007; Sestieri et al., 2011; Raichle, 2015; Smallwood et al., 2015). Its discovery marked a pivotal moment in neuroscience, and as some of the most exciting scientific

breakthrough, has been deemed to be completely accidental (Gusnard et al., 2001; Raichle et al., 2001; Gusnard & Raichle, 2001; Mazoyer et al., 2001; Buckner et al., 2008). Nevertheless, even before these seminal papers that marked the actual discovery of the DMN, other researchers had delved into the notion that spontaneous thinking plays a crucial role during resting state (Ingvar, 1985; Andreasen et al., 1995; Shulman et al., 1997). The observations made by Ingvar (1985) and Andreasen et al. (1995) indicated that resting state was not a passive phase but rather it was associated with dynamic mental activity (Buckner et al., 2008). This early exploration hinted at the involvement of DMN regions in processes such as memory and planning (Andreasen et al., 1995; Binder et al., 1999). Subsequent research revealed that the brain regions comprising the DMN were not isolated entities but were functionally interconnected, operating cohesively as a system (Greicius et al., 2003). This collective body of work has significantly advanced our understanding of the DMN and its role in cognitive functions during what, before its discovery, was quite literally considered "resting state".

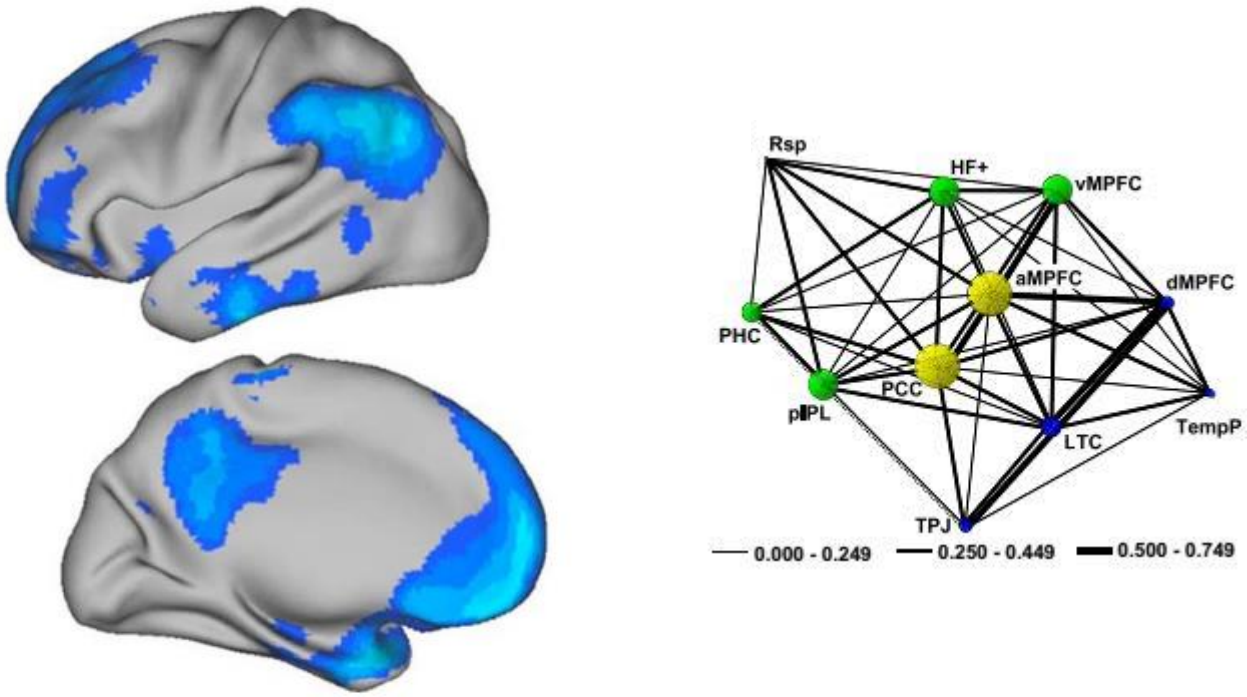

Figure 3. On the left, the medial and lateral surface of the left hemisphere from the PET data of the meta-analysis of Shulman et al. (1997); re-analysed in Buckner et al. (2005). On the right, connectivity of the subsystems and central hubs of the DMN based on their functional coupling as measured by fMRI (Andrews-Hanna et al., 2010a).

The Default Mode Network is characterized by the interaction of two subsystems and a midline core. The *medio-temporal subsystem* includes the hippocampus, parahippocampal cortex, retrosplenial cortex (RSC), posterior inferior parietal lobe, and ventromedial prefrontal cortex (vmPFC). On the other hand, the *dorso-medial subsystem* is composed of the dorsal medial prefrontal cortex (dmPFC), temporoparietal junction (TPJ), lateral temporal cortex, and the temporal pole (Buckner et al., 2008; Greicius et al., 2009; Andrews-Hanna et al., 2010a,b; Andrews-Hanna et al., 2014). Positioned along the cortical midline, both the anterior medial prefrontal cortex (amPFC) and the posterior cingulate cortex (PCC) demonstrate robust functional coupling with both subsystems, acting as crucial functional hubs that facilitate the transfer of information between them (Buckner et al., 2008; Fransson & Marrelec, 2008; Laird et al., 2009; Leech et al., 2011; Andrews-Hanna et al., 2012; Leech, & Sharp, 2014; Raichle, 2015).[1]

Previous research has associated DMN activity with two primary functions: monitoring the external environment (the so-called "sentinel hypothesis") or engaging in internal mentation (Buckner et al., 2008; Andrews-Hanna, 2012). According to the sentinel hypothesis, the DMN might play a role in monitoring the external environment, something that implies a stark contrast with a focused attention on a specific task. This poses that the DMN is more active during periods of passive or diffuse awareness, where individuals are vigilantly observing and processing stimuli from their surroundings (Shulman et al., 1997; Gusnard and Raichle, 2001; Kong et al., 2010; Gronchi & Giovannelli, 2018).

[1] To clarify what we mean anatomically when reporting regions that comprise the DMN, I would like to specify that researchers concerned with the DMN sometimes tend to parcellate mPFC into its anterio-medial (aMPFC), dorsal (dmPFC) and ventral sector (vmPFC), calling mPFC what I (and Ongur et al., 2003) call "vmPFC". Note, however, that aMPFC, which is presented as one of the DMN central hubs, corresponds to almost the exact centre of what we call here vmPFC (see Andrews-Hanna et al., 2010a,b; 2014; Bathelt et al., 2020). Its coordinates reported in numerous fMRI studies are in fact centred on BA10 of the MNI space (see Schmitz & Johnson, 2006; Lieberman et al., 2019; Wen et al., 2020), even though some authors consider aMPFC to also comprise BA9, BA11 or part of them (Johnson et al., 2002; Zysset et al., 2003; Longe et al., 2009; Goldberg et al., 2011).

Support for this hypothesis was initially found in the observation that task-induced deactivation in the DMN is most prominent during tasks involving central (foveal), as opposed to peripheral stimuli (Shulman et al., 1997). Moreover, under certain circumstances, DMN activity positively correlates with performance on sensory processing tasks. For instance, Hahn et al. (2007) reported a positive correlation between DMN activation and high levels of performance on a target-detection task in a condition where attention was broadly distributed across multiple possible locations, rather than focused on a specific location. Moreover, support for the sentinel hypothesis comes from the neuropsychological observation that patients with bilateral lesions extending to the precuneus and cuneus region can present with Balint's syndrome (Mesulam, 2000). Patients with Balint's syndrome experience a form of tunnel vision, wherein they are only able to perceive a limited portion of the visual field at any given time, failing to notice objects outside their immediate focus of attention (Mesulam, 2000). This impairment aligns with the concept of a disruption of a brain system that supports global (instead of focused) attentional capacity (see also Tripathi & Garg, 2022).

An alternative perspective on the function of the DMN suggests its direct role in internal mental processes. Instances of self-reflective thinking and judgments based on inferred social and emotional content strongly activate the DMN, particularly mPFC (Craik et al., 1999; Gusnard et al., 2001; Kelley et al., 2002; Mitchell et al., 2005, 2006; Johnson et al., 2006; Heatherton et al., 2006; Davey et al., 2016; Mancuso et al., 2022). Notably, as mentioned above, the central hubs of the DMN are intricately connected with the hippocampus, making the DMN share common regions with those activated during the recollection of episodic memories and the imagination of future scenarios. The DMN is in fact consistently found to be active during autobiographical memory recollection and episodic future thinking (Buckner et al., 2005; Spreng et al., 2009; Spreng, & Grady, 2010; D'Argembeau et al., 2014; Schacter et al., 2017; Addis, 2020). Indeed, Svoboda and colleagues (2006) performed a meta-analysis of

neuroimaging studies focused on autobiographical memory (see also Maguire, 2001; Cabeza & St. Jacques, 2007; Philippi et al., 2015). Across all the studies included, participants were engaged in recalling experiences from their personal pasts. The synthesized findings across these studies reveal a consistent set of brain regions remarkably akin to the default network, encompassing the ventromedial prefrontal cortex (vmPFC), dorsal medial prefrontal cortex (dmPFC), posterior cingulate cortex/retrosplenial cortex (PCC/Rsp), inferior parietal lobule (IPL), lateral temporal cortex (LTC), and the hippocampal formation (HF).

Furthermore, multiple neuroimaging studies suggest the involvement of mPFC in Episodic Future Thinking (EFT; D'Argembeau & Mathy, 2011; D'Argembeau, 2011; Schacter et al., 2015; Xu et al., 2016; Schacter et al., 2017; Bellana et al., 2017; Yang et al., 2020). mPFC also plays a pivotal role in constructing experiences without a specific temporal context (Hassabis et al., 2007a). Specifically, neuroimaging evidence indicates that vmPFC is heavily recruited during the construction of complex scenarios (Hassabis & Maguire, 2007), supporting simulations of future emotional experiences and personal goals (Benoit et al., 2014; Stawarczyk & D'Argembeau, 2015), and during episodes of mind wandering (Christoff et al., 2009; Fox et al., 2015; Bertossi & Ciaramelli, 2016; Kucyi et al., 2016; Mittner et al., 2016; Poerio et al., 2017; Philippi et al., 2021).

It is noteworthy that a recent study found the mPFC and the PCC to serve as convergence points for different subsystems within the DMN that were observed to be characterized by opposing correlations with EEG alpha power. These distinct correlations suggest disparate functions, with one subnetwork associated with a sentinel role and the other linked to introspection, indicating that the DMN might be involved in switching between introspection and sentinel functions (Bowman et al., 2017; see also Smith et al., 2018).

Now that we have tried to lure the reader into (rightfully) thinking that vmPFC (and the network it is part of) are involved in a multitude of cognitive functions, what I would like to do is review each one of them individually. This is not to say those functions do not constantly interact with each other: the brain has no interest in maintaining the strict theoretical parcellation of cognitive processes that we, researchers, love so much. However, for this elaborate to maintain a coherent sense of what we are discussing, I need to somehow separate all those topics. Thus, I will follow this "divide and conquer" approach just for the sake of clarity, and to accompany the reader towards what I think will successfully bind together all the functions vmPFC is related to into a coherent functional principle: *schematic processing*.

In the following sections, I will retain a special attention towards neuropsychological studies, given this is essentially the focus of the present elaborate. Whenever possible, and at risk of appearing "less professional", I will also recount some personal experiences I had while testing patients with vmPFC damage. I decided to do so because it is my personal opinion that sometimes what strikes us the most is not exactly what you were testing, asking, or what you expected. In doing so, I also believe I will achieve two crucial goals of the present thesis: one, I would like to satisfy the reader's curiosity about what these patients are really like, even outside the testing sessions. I strongly believe these snippets of lived experiences often give insight into it. Two, as a more "scientific" justification, I would like to remind the reader that using an "episodic approach" can spur reasoning and imagination, perhaps, (who knows) for the sake of future studies. Indeed, the literature teaches us that the mechanisms at the bases of detailed episodic recollection allow for the flexible recombination of elements from different experiences into novel scenarios (Addis et al., 2007; Hassabis et al., 2007a,b; Moscovitch, 2008; De Luca et al., 2018a), something that is instrumental in various highly adaptive functions and behaviours, such as problem-solving (Sheldon et al., 2011), creativity (Warren et al., 2016), and decision-making (Kwan et al., 2015).

With all that being said, to maintain a cohesive logical continuum, we will start by discussing vmPFC and its role in autobiographical memory recollection.

2. Cognitive processes linked to vmPFC

2.1. vmPFC and autobiographical memory

I could certainly tell many stories that reveal vmPFC patients' memory deficits, but here I decided to pick just one of them. This episode does not just speak about memory impairment *per se*, but also about different levels of awareness that patients might (or might not) have about it. Once, I had to test the same patient in two separate sessions, a week apart. The second testing session was extremely similar to the first one, apart from some small detail. For that reason, I would sometimes start the session by asking whether the patient remembered what we had done the first time, to then briefly recap instructions and describe what was different that time around. On this particular occasion, as I asked the patient whether he remembered what we had done the previous week, I was met with a stern "no", and a look that made me question the absurdity of what I was asking. "No", *tabula rasa*. So, I started to explain the task in detail as it was the first time, all whilst the patient was looking at me with a bored expression, clearly thinking "again, really?". The patient did indeed remember the previous session, and could not understand why I was speaking to him like he was amnesic. What I did wrong here, which I only understood later, was that I asked a question that was too broad and unspecific. The patient was sincere in telling me that he did not remember what I asked him to do the previous week. However, had I asked "do you remember last time when we did *so and*

so [..]?" his answer would have probably been a lot different. We will see later that this exact behaviour was what tricked researchers decades ago into thinking that vmPFC damaged patients are indeed amnesic.

As mentioned in the previous section, vmPFC has been consistently implicated in autobiographical memory recollection, in both neuroimaging and lesion studies (Bonnici et al., 2012; Martinelli et al., 2013; Bertossi et al., 2016a,b; Bonnici & Maguire, 2018; Rolls, 2022). Its interaction with the hippocampus has convincingly revealed itself as a key node for a coherent organization of autobiographical memory, integrating previous experience into ongoing memory processes (Nieuwenhuis & Takashima, 2011; Preston & Eichenbaum, 2013; Spalding et al., 2015; Daviddi et al., 2023). But what exactly is vmPFC doing?

In a recent study using dynamic causal modelling, Nawa & Ando (2020) demonstrated that, during the (re)construction of episodic autobiographical memory, vmPFC, along with other posterior midline cortical regions (PCC, precuneus, and retrosplenial cortex), played a crucial role in initiating the search for specific autobiographical events. This network also coordinated the subsequent activation of the angular gyrus and the hippocampus, and, during the later stage of autobiographical memory elaboration, vmPFC continued to stimulate hippocampal activity throughout the entire recollection process. This is consistent with results from MEG studies, showing vmPFC and the hippocampus to exhibit the highest levels of activity during the retrieval of episodic autobiographical memories, an activity that remains strong throughout the later phases of the recollection process (McCormick et al., 2020). So, if vmPFC both initiates and monitor autobiographical memory recollection, much like a conductor, where does that leave patients that have suffered a vmPFC lesion?

As outlined above, more than 30 years ago, it was observed that individuals with vmPFC damage encountered substantial difficulties in recalling autobiographical memories

(Della Sala et al., 1993; Kopelman et al., 1999). For instance, a recurring finding was that patients with vmPFC lesions tended to retrieve fewer autobiographical memories compared to healthy controls, a deficit closely linked to impairments in executive functions (Della Sala et al., 1993).

It is of crucial importance here to highlight how these early investigations made a somewhat similar mistake as mine when I asked a broad and unspecific question to a vmPFC patient. These researchers tended to investigate autobiographical memory by employing the Crovitz Test (Crovitz & Schiffman, 1974), which involves giving cue words and asking participants to retrieve a memory for each one, which is, again, broad and unspecific, and leaves patients with the burden of applying metacognitive search strategies.

More recently, other studies reported slightly similar results, describing how frontal patients displayed difficulties in recalling autobiographical events and lacked the use of appropriate metacognitive search strategies (Thaiss & Petrides, 2008). Nevertheless, once patients successfully recalled a specific event, they could provide as much details as healthy controls. Similarly, Kurczek et al. (2015) tasked participants with damage to vmPFC or to the medial temporal lobe (MTL) with generating autobiographical memories using a similar technique to that of the Crovitz Test. Patients were instructed to pinpoint one specific moment from a memory and provide a detailed description. Remarkably, individuals with vmPFC damage demonstrated an intact ability to describe these chosen snapshots, whereas patients with MTL damage struggled to provide detailed descriptions even of individual moments from events. Likewise, Bertossi et al. (2016a) prompted participants to recall nine autobiographical memories using a Crovitz-type approach. Once a memory was identified, participants were asked to describe the entire event in as much detail as possible. Patients with vmPFC damage exhibited difficulties in recalling details of these events, a deficit apparent across both recent and remote autobiographical memories. This suggests that whilst hippocampal patients may

struggle to picture even a single scene in their mind, those with vmPFC damage might face challenges in visualizing the unfolding of extended events.

To sum up, the memory deficits of patients with vmPFC damage seem to concern primarily meta-cognitive strategies related to autobiographical memory organization, and the coherent and detailed unfolding of extended events, which is consistent with the supposed role of vmPFC in both initiating and monitoring autobiographical memory recollection. As we will see in the next section, these impairments are also apparent (maybe even more so) when vmPFC patients are asked to imagine future events.

2.2. **vmPFC and episodic future thinking**

Once, I was asking a vmPFC patient to imagine and describe a fictitious event that could plausibly happen to him in the future in as much detail as possible. He replied with just one short sentence. Second trial: same thing all over again. Third trial, you guessed it: just one sentence, despite my encouragements to give more details. This was the first time I ever tested this patient, so I just though he was a man of few words. Since I could tell he was getting quite tired, I asked if he wanted to take a break. "Yes, please". So, I interrupted the testing and asked whether he was struggling with the task. "No, not at all", and then proceeded to *make conversation*. Minutes and minutes of simple conversation, non-stop, without any trouble at all. He was telling me about his life and his previous job. Nothing too specific, but he was going on and on, so much so that I felt sorry I had to resume the task. And then, of course,

when we got to the next trial and I asked him to imagine another event, I could only get a few short sentences out of him.

So, what was happening there?

He was not, of course "a man of few words". The difference between what I asked him to do and him making conversation, is that in the first case, he had to *create* something. He had to set a scene, make up an event, and imagine himself in it. As we will discuss later, this is exactly what he was struggling with, despite being completely unaware of it.

Episodic memory allows us to mentally transcend subjective time, transporting us back into the past and forward into the future (Tulving, 1985). This mental projection into the future, involving the pre-experiencing of events through simulation, has been termed "episodic future thinking" (EFT; Atance & O'Neill, 2001, 2005; Schacter et al., 2017) or "prospection" (Gilbert & Wilson, 2007; Szpunar & Spreng, 2014). Through this process, individuals can simulate personal future scenarios, envision contextual details, and anticipate their outcomes (Atance & O'Neill, 2001; Gilbert & Wilson, 2007; Schacter et al., 2015). These imagined scenarios exhibit high vividness, richness, and specificity, generating a sense of pre-experiencing the future event akin to the re-experience of a genuine past event (Tulving, 1985; Atance & O'Neill, 2001). Notably, the act of envisioning personal future experiences shares striking similarities and common neural bases with the recollection of personal past experiences (Maguire, 2001; Tulving, 2002; Martin-Ordas et al., 2012; Cole et al., 2016). Indeed, fMRI studies have consistently reported an overlap of activation in brain regions of the DMN that are typically associated with autobiographical memory recollection (mPFC, PCC, Rsp, the hippocampus; Okuda et al., 2003; Szpunar et al., 2007; Spreng, et al., 2009; Viard et al., 2011; Verfaellie et al., 2012; Schacter et al., 2012).

However, EFT is not just based on the same exact processes involved in episodic memory recollection. Recent studies have in fact started to implicate semantic memory in the process of simulating future events. These investigations reveal that, during EFT, individuals also activate semantic, abstract information, including personal knowledge and/or general events (D'Argembeau & Mathy, 2011; Irish et al., 2012; Ito et al., 2013; Wang & Huang, 2014, 2016; La Corte & Piolino, 2016). Thus, abstract knowledge derived from semantic memory serves as a contextual framework, guiding the retrieval and integration of episodic details (D'Argembeau & Mathy, 2011; Klein, 2013) and acts as a "scaffold," facilitating the (re)construction of both past and future events, and providing abstract representations to merge with episodic details, enriching them with personal significance (Irish et al., 2012; Irish & Piguet, 2013). These scaffolding structures, when idiosyncratically related to the self, have been elsewhere dubbed "personal semantics", and are known to be crucially linked to vmPFC activity (Renoult et al., 2012, 2016; Grilli, & Verfaellie, 2014; Coronel & Federmeier, 2016; Tanguay et al., 2023).

If vmPFC is heavily involved in both the production of episodic details, and in the semantic structures acting as their building blocks, are patients with vmPFC lesions capable of engaging in episodic future thinking?

Bertossi et al. (2016b) tested patients with vmPFC damage using a Crovitz-based technique that also included future scenarios. Patients displayed comparable impairments in imagining both future and past events. Similarly, Bertossi et al. (2016a) used a scene construction task in which participants described either fictitious scenes or anticipated future events. Once again, vmPFC patients displayed deficits in imagining both types of scenarios. Subsequently, the same authors sought to distinguish the mental construction of future scenarios from describing a visible image or recalling a recently viewed picture (Bertossi et al., 2017a; see also McCormick et al., 2018a). Findings revealed that patients with vmPFC damage

struggled to provide specific details across all conditions. Interestingly, even when accounting for performance in the description conditions, the deficit in the mental construction of future scenarios persisted. Moreover, Verfaellie et al. (2019) reported how vmPFC patients were unable to incorporate the self-schema into imagined future scenarios, an ability that was instead retained in MTL-damaged patients.

In summary, the EFT impairments of vmPFC patients, much like their difficulties in autobiographical memory recollection, seem to lie in the activation of the scaffolding necessary for event construction, and in transitioning from one scene to another to advance towards a cohesive mental representation of a prolonged event. In the next section, we will see that such impairment is also evident when asking vmPFC patients to construct scenes.

2.3. **vmPFC and scene construction**

Once, I was asking a vmPFC patient to mentally construct a scene starting from an object, and then imagine an event that could take place in the scene. Throughout the trials, I could tell that this patient was not really complying with the instructions: she was just giving me general semantic information about the objects. Since I was not sure about whether she was really trying or just wanted to get the task done as quickly as possible, I started to remind her that she should describe a scene in as much details as she could. She answered something along the lines of "this is exactly what I am doing". *Ok, I better try some other tactics*, I thought. On the next trial, she had to imagine and describe a scene starting from a swing. As always, she gave me some general information, telling me that there is probably a child on it, and if the

child is grown enough, he could ride it without the parent. This time, I was determined to understand whether she was really putting an effort in, so I explicitly asked "is there a scene surrounding the swing?". She said "yes", but would not continue. So, I prompted her "do you think you could describe it?" "It's a playground". *Nevermind*, I thought, and went on with the testing, as usual. Picturing a scene in the mind's eye did not make much sense to her. She was really trying, and thought she was doing exactly what I was asking. But why?

Scene construction pertains to the ability to envision and describe spatially coherent scenes (Mullally et al., 2014; Irish et al., 2015; Kim et al., 2015; Roberts et al., 2018; Madore et al., 2019). Scene construction theory (Hassabis & Maguire, 2007; Mullally et al., 2012; McCormick & Maguire, 2021) aims to elucidate the shared cognitive and neural foundations underlying episodic memory, episodic future thinking, and the visualization and maintenance of complex spatial scenes. This constructive process evolves from the reactivation, retrieval, and integration of semantic, contextual, and sensory information stored in sensory cortices (Wheeler et al, 2000; Hassabis et al., 2007b; Maguire & Mullally, 2013; Dalton et al., 2018; Schacter & Addis, 2020). Hassabis et al. (2007a) explored brain activity as measured by fMRI during the recall of past experiences and the construction of novel, fictitious scenarios, revealing a network of brain regions termed by Hassabis & Maguire (2009) "construction system" (see also Summerfield et al., 2009). In particular, this network comprises the hippocampus, parahippocampal gyrus, retrosplenial cortex, posterior parietal cortex, middle temporal cortices and mPFC (Hassabis et al., 2007a; Bird et al., 2010; Mullally & Maguire, 2014). Of note, the construction network is not exactly the same as the one concerned with object representation and manipulation, highlighting the differences between scene construction and simple visual imagery (Roland & Gulyas, 1994; Sugiura et al., 2015).

As per autobiographical memory recollection and future thinking, previous evidence suggests a role of vmPFC as the initiator and conductor of scene construction processes. In a

recent MEG study, Barry et al. (2019) found vmPFC to guide hippocampal activity during the construction of novel scenes. Alterations in theta power within the vmPFC occurred before similar changes were noted in the hippocampus, a finding suggesting a shared mechanism for episodic memory retrieval and the imaginative construction of scenes (see also Monk et al., 2020). Monk et al. (2021) replicated the finding of vmPFC driving hippocampal activity during the first stages of scene construction, also reporting mutual entrainment between the two brain regions even after the initial stages.

In the case of scene construction, the performance of vmPFC patients closely mirrors that of patients with hippocampal damage, albeit with some differences (Mullally et al., 2012, 2014; McCormick et al., 2017; Lynch et al., 2020). Specifically, vmPFC patients exhibit deficits in simulating personal past and future events, also struggling when imagining fictitious, atemporal scenarios (Bertossi et al., 2016a). De Luca et al. (2019) confirmed that vmPFC-damaged patients are indeed impaired at constructing coherent scenes, also showing a reduced boundary extension effect.

Challenges in scene construction can also impact spatial navigation. Ciaramelli (2008) explored the spatial navigation ability of a vmPFC patient within his hometown. Notably, even though he retained intact knowledge of landmarks and routes, he experienced difficulty in maintaining his goal destination during navigation, and was instead drawn towards familiar landmarks and previously attended locations along the route (Ciaramelli, 2008; McCormick et al., 2018a).

To sum up, we have seen how vmPFC patients demonstrate a general inability in directing their attention towards a mental image, but, as we will see in the next paragraph, their struggle also seems to expand in orienting their attention inward, towards their mental life.

2.4. **vmPFC and mind wandering**

Sadly, I do not really have any stories about vmPFC patients' mind wandering experiences. It is vanishingly rare for them to let you know something about their thoughts, emotions, or about their psychological state in general, and, as we will discuss later, this is most likely due to their mental life being impoverished overall.

Mind wandering refers to the "occurrence of thoughts that are not tied to the immediate environment, not related to a given task at hand" (Murray et al., 2020). In instances of mind wandering, individuals' thoughts veer away from the current task, redirecting towards inner reflections, fantasies, and emotions unrelated to the immediate objective (Smallwood & Schooler, 2006, 2015; Mooneyham & Schooler, 2013; Feng et al., 2013; Christoff et al., 2016; Seli et al., 2016). Mind-wandering takes place with the initiation of mental events (Gelbard-Sagiv et al., 2008; Callard et al., 2013; Smallwood & Schooler, 2015), and prominently encompasses the generation of mental images depicting scenes, including autobiographical reminiscences, future-oriented thoughts, and simulations of atemporal scenes and events (Watkins, 2008; Baird et al., 2011, 2012; Andrews-Hanna et al., 2010a,b; 2013, 2014; Smallwood & Andrews-Hanna, 2013; Smallwood et al., 2016). People spend approximately 50% of their waking life engaging in mind wandering (Kane et al., 2007; Killingsworth & Gilbert, 2010; but see Seli et al., 2018), so, even though I hope to capture the reader's attention all throughout the reading of this thesis, it is extremely likely that at some point, s(he) will have to (or already had to) read a paragraph twice, having suddenly realized their mind has drifted away, disrupting the understanding of what is written here. It is far from being a rare experience. But what are the neural bases of mind wandering?

McGuire et al. (1996) was the first to demonstrate that the frequency of mind wandering reported by participants during a low-demanding task strongly correlated with mPFC activity. Since then, the literature has convincingly established a close link between the mind wandering experience and DMN activity (Christoff et al., 2004; Mason et al., 2007; Sonuga-Barke, & Castellanos, 2007; Buckner et al., 2008; Gruberger et al., 2011; Kucyi et al., 2016; Christoff et al., 2016; Poerio et al., 2017). For example, Fox et al. (2015) performed a meta-analysis utilizing activation likelihood estimation (ALE), highlighting several pivotal areas within the DMN that consistently exhibited activation across studies (mPFC, PPC, MTL, and the inferior parietal lobule, IPL).

In an fMRI study, Christoff et al. (2009) had participants perform experience sampling during a sustained attention to response task (SART). The authors found that within the core default network, regions exhibited higher activity levels during task-unrelated thought as compared to task-related thought. However, different DMN sub-systems seemed to be involved in different types of mind wandering. For example, the medio-temporal subsystem demonstrated similar activity levels for both task-unrelated and task-related thought. The same subsystem also revealed an increased activity when participants were unaware of their task-unrelated thoughts compared to when they were aware of them. This lack of awareness also appeared associated with fewer constraints on thought, implying a distinctive connection between the medio-temporal subsystem and spontaneity. Conversely, regions within the core of the DMN displayed comparable activity levels for both unaware and aware task-unrelated thought (see also Smith et al., 2006; Stawarczyk et al., 2011; Mittner et al., 2016; Christoff et al., 2016).

Bar et al. (2007) identified a network for contextual associative processing that closely resembled the medio-temporal subsystem, including the PHC, retrosplenial cortex, medio-parietal cortex, and mPFC. Components of this network showed heightened activation when

individuals viewed pictures of objects that evoke strong contextual associations compared to pictures of objects lacking specific contextual uniqueness, thus having low associativity. Therefore, the authors suggested that the medio-temporal subsystem might contribute to spontaneous thought through its involvement in contextual associative processing (see also Aminoff et al., 2013). The importance of the hippocampus in self-generated spontaneous thoughts is also apparent when studying the temporal dynamics of mind wandering. Data derived from single-cell recordings indicates that the emergence of spontaneous memories is heralded by the activation of hippocampal neurons that were engaged during an initial encoding process (Gelbard-Sagiv et al., 2008). Likewise, recent research has identified the hippocampus to exhibit significant neural activity 2 seconds before the occurrence of a spontaneous thought (Ellamil et al., 2016; see also Girn et al., 2017). Neuropsychological studies on amnesic patients also provide insight for the role of the hippocampus in spontaneous cognition. McCormick et al. (2018c) conducted a study involving six patients with bilateral hippocampal damage. Over the course of two days, patients were prompted to share the contents of their thoughts at 20 different time points. Hippocampal patients did not exhibit a significant increase or decrease in mind wandering compared to the healthy control group. However, the quality of their mind wandering episodes was different, being more frequently associated with the present vs the past. Notably, patients reported a higher prevalence of atemporal and hypothetical thinking. Additionally, in comparison to the control group, patients reported fewer visual thoughts and a higher occurrence of verbal thoughts (see also McCormick et al., 2018a,b; Faber & Mills, 2018). This seems to suggest that hippocampal patients do engage in mind-wandering, but their mind wandering episodes reveal themselves as generally devoid of episodic content.

Therefore, it has been proposed that the hippocampus might play a role in generating the diversity of content experienced during mind wandering episodes (Christoff et al., Mills et al., 2018). In this context, the vmPFC-hippocampal axis acquires crucial importance. Given

the well-known role of vmPFC in initiating and orchestrating the activation of scaffolding structures necessary for autobiographical memory recollection and EFT, Ciaramelli & Treves (2019) proposed that during mind-wandering, a similar functional principle might apply. Specifically, vmPFC might trigger the formation of alternative events beyond direct perceptual experience by activating relevant *schemata*, which are then used by the hippocampus to construct a basic outline or sketch of the event (Gilboa & Marlatte, 2017; Ciaramelli et al., 2019). Then, vmPFC plays a crucial role in enriching the mental event by engaging in iterative retrieval and integration of elements congruent with the activated schemata, facilitated by feedback loops with the hippocampus (Benoit et al., 2014; Moscovitch et al., 2016; McCormick et al., 2018a; Ciaramelli & Treves, 2019; O'Callaghan et al., 2019; Konu et al., 2020; Spalla et al., 2021).

This proposal aligns with the observed correlation between the strength of functional connections between the hippocampus and vmPFC and the experienced degree of mental time travel during mind wandering (Karapanagiotidis et al., 2017). Also, recent evidence suggest reduced mind wandering episodes in vmPFC patients. Bertossi & Ciaramelli (2016) had participants engage in three tasks with varying cognitive demands, whilst their thoughts were intermittently sampled. Additionally, they provided self-reports on their daydreaming tendencies in daily life. vmPFC damaged patients demonstrated lower rates of mind-wandering across tasks and reported experiencing daydreaming less frequently compared to healthy and brain-damaged controls. Notably, vmPFC damage was associated with a reduction in off-task thoughts related to the future, while concurrently increasing thoughts about the present (see also Bertossi et al., 2017b; Giordani et al., 2023).

Now that we have discussed vmPFC patients' inner mental life, in the next section we will move onto how this mental life expresses itself in relation to others, by exploring the involvement of vmPFC in social cognition.

2.5. vmPFC and emotion regulation, social cognition, and theory of mind

This is the story of my first ever session with a patient with vmPFC damage. I had previously read a great deal about them: I obviously knew the story of Phineas Gage (Harlow, 1848, 1993), and I had also read papers about their poor emotional regulation, social inappropriateness and impulsivity, so much so that Blumer & Benson (1975) defined them "pseudopsychopathics". I'll admit I was a bit scared. Anyway, for this particular session, I was only supposed to assist: another student was testing, and he already knew the patient. He started as usual, by asking some personal data. The patient had the sheet of paper in front of him. He read the question "Sex", with the options "M" or "F". He looked at both of us, giggling, and asked "does the M stand for *much*?". The student simply brushed it off and said "come on G., we have work to do". The patient laughed, and the session started as normal.

I have to say, this particular patient was probably not the most socially appropriate person in the world, but was never harassing or malicious. Some people with a particular sense of humour could very well describe him as hilarious. However, not many people would choose to tell such a joke in such a context, so, what was going on?

Emotion regulation (ER) can be conceptualized as one's attempts to monitor and modulate their emotional experience (Gross & Thompson, 2007; see also Gross, 2008, 2014; Kok 2020). Drawing from behavioural and neuroimaging findings, researchers have identified two overarching forms of emotion regulation: "explicit" and "implicit" regulation (Gyuriak et al., 2011; Webb et al., 2015; Braunstein et al., 2017). Explicit regulation involves conscious effort for its initiation and needs active and deliberate monitoring during implementation, and is thus linked to a certain level of insight and awareness (Etkin et al., 2015; Dhaka & Kashyap,

2017). One extensively studied explicit regulation strategy is reappraisal, which involves consciously altering the self-relevant meaning (appraisal) of a stimulus that induces emotion (Goldin et al., 2008). Neuroimaging meta-analyses indicate that reappraisal is correlated with activation in various brain regions, including the frontoparietal executive network (such as the dlPFC, the vlPFC, and the parietal cortex), as well as the insula, supplemental motor area (SMA), and pre-SMA (Buhle et al., 2014; Kohn et al., 2014; Frank et al., 2014; Picó-Pérez et al., 2019).

Implicit regulation, on the other hand, is automatically triggered by the stimulus itself, unfolds without conscious monitoring, and can occur without insight and awareness (Mauss et al., 2007; Gyurak et al., 2011; Koole & Rothermund, 2011; Koole et al., 2015). Examples of implicit regulation include the inhibition of fear and the regulation of emotional conflict (Kerns et al., 2004; Sotres-Bayon & Quirk, 2010; Mocaiber et al., 2010; Tupak et al., 2014). In these scenarios, neural activation consistently occurs in the ventral anterior cingulate cortex (vACC) and the vmPFC (Stevens et al., 2011; Etkin et al., 2015; Braunstein et al., 2017; Silvers & Moreira, 2019; Berboth & Morawetz, 2021).

The vACC/vmPFC axis has also been demonstrated to modulate stimulus value during decision-making. Specifically, opting for a choice associated with a higher positive value is consistently linked to increased activation in the vACC and vmPFC (Rushworth & Behrens, 2008; Rushworth et al., 2011; Bartra et al., 2013; Vassena et al., 2014; Clithero & Rangel, 2014; Zhang & Gläscher, 2020). When a reward becomes linked to a particular stimulus or action, there is a spontaneous escalation in the perceived decisional value of that specific stimulus or action (Grabenhorst & Rolls, 2011; Kaping et al., 2011; Etkin et al., 2015). This heightened decisional value, as reflected in the increased vACC–vmPFC activity, contributes to an increased likelihood of selecting that particular stimulus or action in subsequent decisions (Harris et al., 2011; Rushworth et al., 2011; Rangel & Clithero, 2014). Additionally, evidence

from both neuroimaging and lesion studies indicate a role of vmPFC in moderating and inhibiting amygdala response for a successful regulation of fear and for extinction of fear conditioning (Bechara et al., 1999; Phelps et al., 2004; Sotres-Bayon et al., 2004; Mobbs et al., 2007; Klumpers et al., 2010; Motzkin et al., 2015; Gold et al., 2015, 2016; Hise & Koenigs, 2018; Andrewes & Jenkins, 2019; Battaglia et al., 2020).

vmPFC has also been implicated in social cognition by several lines of work (Adolphs, 1999; Amodio & Frith, 2006; Bicks et al., 2015; Gangopadhyay et al., 2021; Meisner et al., 2022). Specifically, its fundamental functional role might involve facilitating the assessment and portrayal of interpersonal and mental attributes related to both the self and social targets (Macrae et al., 2004; Van Overwalle, 2009; Murray et al., 2012; Hughes et al., 2012; Flagan & Beer, 2013; Delgado et al., 2016). Thus, vmPFC is believed to contribute to the assessment and depiction of interpersonal qualities, as well as the capacity to deduce the mental states of other individuals, commonly referred to as theory of mind (ToM; Beer et al., 2003; Gallagher & Frith, 2003; Delgado et al., 2016). Baron-Cohen et al. (1994) found BA11 to be engaged in a mental-state recognition task in which participants were instructed to recognize mind-related words (*e.g.* "think", "believe"). Kobayashi et al. (2007) found vmPFC to be preferentially activated in children performing a ToM (rather than a non-ToM) task, in which participants had to infer other people's thoughts. In an fMRI study, Sebastian et al. (2012) reported mPFC and vmPFC to display preferential activation during a task requiring affective ToM (*i.e.* the ability to infer other's feelings) rather than cognitive ToM (*i.e.* the ability to infer beliefs and motivations of others; see also Leopold et al., 2012). Also, Hooker et al. (2011) reported how theory of mind skills were correlated with grey matter volume of the vmPFC in schizophrenic patients. Lev-Ran et al. (2012) used repetitive transcranial magnetic stimulation (rTMS) to inhibit vmPFC activity whilst participants performed a ToM task, and found that, while the

control sham condition did not produce any effect, the active rTMS significantly disrupted ToM learning.

Lesion studies on the role of vmPFC in social cognition and emotional regulation have a long and rich history, going back to one of the most famous cases in neuropsychology, Phineas Gage, who suffered a traumatic brain injury and experienced dramatic changes in behaviour and personality (Harlow, 1848, 1993). Since then, emotional regulation and social cognition have been extensively studied in individuals with vmPFC damage. For instance, vmPFC patients are known to exhibit difficulties in recognizing emotional facial expressions (Heberlein et al., 2008; Tsuchida & Fellows, 2012; Vandekerckhove et al., 2014). Additionally, their attention towards emotionally salient information in faces is diminished, as evidenced by a reduced frequency of gaze towards such cues (Wolf et al., 2014, 2016). These patients also display attenuated skin conductance responses to emotional stimuli, such as images portraying emotional faces (Damasio et al., 1990; Koenigs et al., 2007). Moreover, assessments using the Iowa Scales of Personality Change (ISPC) consistently reveal that vmPFC-damaged patients exhibit diminished emotional expressiveness (Barrash et al., 2000, 2011, 2022). Patients with vmPFC lesions also show a greater disposition to negative mood induction, and sometimes react aggressively and with impulsivity (Koenigs et al., 2007; Gillihan et al., 2011; Blair, 2016). This disposition to negative emotions appears in stark contrast with the observation that vmPFC lesions are recognized to be protective against depressive disorders (Koenigs et al., 2008a,b; Hiser & Koenigs, 2018). McCormick et al. (2018a) suggested the tentative explanation that both the impulsive behaviour and the reduced rates of depression in vmPFC patients might be due to their general impoverishment of inner mental reflections.

vmPFC patients' poor meta-cognitive abilities also manifest themselves in social situations. For example, they sometimes tend to use inappropriate verbal behaviour directed towards strangers (Rolls et al., 1994; Roberts et al., 2019) or divulge personal information of

an inappropriate nature during conversations with unfamiliar individuals, seemingly devoid of the typical embarrassment associated with it (Tranel, 2002; Beer et al., 2003, 2006; Anderson et al., 2006). Additionally, these patients endorse more frequently than healthy and brain damaged controls behaviours that typically evoke interpersonal disgust (Ciaramelli et al., 2013). They also tend to exhibit abnormal preferences for interpersonal distance, often manifesting a propensity for closer proximity in social settings (Perry et al., 2016). Importantly, vmPFC damaged patients appear to retain semantic knowledge of social rules, since they experienced typical embarrassment when observing recordings of their socially inappropriate behaviour (Beer et al., 2006). Therefore, their impairment seems to lie in the capacity for self-insight and real-time mental reflection regarding appropriateness in social situations.

These deficits are apparent even when testing vmPFC patients in tasks related to ToM skills, particularly (but not exclusively) when testing the affective component of ToM (Stuss et al., 2001; Shamay-Tsoory et al., 2005, 2007; Leopold et al., 2012; Gupta et al., 2012). For instance, they are impaired in understanding the thoughts of a person considering the thoughts of another (Shamay-Tsoory et al., 2003). Additionally, they encounter difficulties in detecting instances where a character makes statements without considering the listener's preferences, an ability that requires the simulation of others' mental states (Stone et al., 1998; Leopold et al., 2012). However, individuals with vmPFC damage may still demonstrate intact abilities in emotion recognition and affective empathy when they depend on immediate emotional contagion and resonance mechanisms (Zaki & Ochsner, 2012).

The role of vmPFC in ToM tasks and tasks requiring social cognition has been related to the social proximity felt with the social target (Mitchell et al., 2002, 2005, 2006; Singer & Tusche, 2014; Welborn & Lieberman, 2015), and thus in their distance from the self-schema, something that brings us to the next section, self-referential cognition.

2.6. **vmPFC and self-referential cognition**

Here, I am not going to report a particular anecdote about an interaction with a vmPFC patient, since self-referential cognition is exactly the focus of the first two studies of the present elaborate. I will, however, highlight something I find important: despite the astonishing impairments we found in these patients (see Chapters 1 and 2), they are not necessarily evident during simple conversations with them. Quite the contrary. You could spend hours talking to them and not realize the profound disturbance they display in self-referential cognition. This is probably due to them being totally unaware of this deficit, and thus, if one does not specifically pay attention to it, (s)he might very well miss it. Probably, this is the reason why, despite the fact that these patients have been studied for more than a century, researchers have just recently started to study their self-schema. As we will see later, however, when experimentally tested, this degradation in self-referential cognition becomes more than evident.

Self-perception involves making attributions about one's own personality based on observations and memories (Bem, 1967, 1972; Goldstein & Cialdini, 2007). Research on the self in social cognition has primarily focused on the structure of individuals' mental representations of themselves and how these representations influence their perceptions of others (Markus et al., 1985; Silvia & Gendolla, 2001; Kenny & West, 2010). For example, personality traits that are considered central to individuals' mental self-representations, known as their self-schema, are more consistently accessible (Markus, 1977; Fenigstein & Levine, 1984; Bargh et al., 1986; Kihlstrom & Klein, 2014). As a result, these traits influence the attributions individuals make about the behaviour of others. Indeed, when individuals are asked to assess the similarity of other people, they often default to using their own selves as a reference point (Srull & Gaelick, 1983; Smith, 1984; Catrambone et al., 1986; Wagner et al.,

2012), which could be the explanation for vmPFC patients' deficits in ToM tasks (Lieberman, 2007; Krueger et al., 2009; Roy et al., 2012).

Early investigations into the cognitive mechanisms implicated in self-representation revolved around memory processes. For instance, initial research indicated that evaluating trait words based on their self-relevance enhances their subsequent recall (the so-called self-reference effect, SRE; Rogers et al., 1977; Klein & Kihlstrom, 1986; Symons & Johnson, 1997; Klein, 2012), which has consistently been linked to vmPFC activity (Craik et al., 1999; Kelley et al., 2002; Yaoi et al., 2015; Kim et al., 2015; see Chapter 1). Interestingly, more dorsal regions of mPFC (BA 8 and BA 9) seem to be preferentially implicated in attributions about others (Baron et al., 2011; Wagner et al., 2012; Kang et al., 2013), with a ventral-to-dorsal axis along the self-others continuum (Jenkins et al., 2008; Denny et al., 2012; Ferrari et al., 2016; Lieberman et al., 2019). The advantage in memory for the self-schema also extends to stimuli that are categorized in relation to oneself rather than to others (Conway, 2005; Cunningham et al., 2008; Turk et al., 2011a,b). Moreover, vmPFC is consistently found to be active when participants engage in self-reflection, even when reflecting on one's past and future self, but still retaining a stronger activation when participants think about their present self (Jenkins et al., 2008; D'Argembeau, 2008; Rameson et al., 2010; D'Argembeau et al., 2010a; Herwig et al., 2012). Recently, Yin et al. (2021) linked vmPFC activity to the processes allowing for the self-bias, *i.e*. the systematic biases in perception, memory, and attention that favour information associated with oneself as opposed to information related to other individuals (Sui et al., 2012; Sui & Humphreys, 2017). Specifically, authors observed vmPFC to exhibit heightened functional connectivity with working memory regions while participants maintained self-associated cues, and this connectivity was predictive of individuals' behavioural self-prioritization effects. Moreover, in a follow-up transcranial direct current stimulation (tDCS) experiment, the authors demonstrated that cathodal (inhibitory) tDCS completely abolished the

self-prioritization effect (Yin et al., 2021). mPFC has also been implicated in the tendency to self-enhance, *i.e.* "the tendency to maintain an often unrealistic, positive view of the self" (Taylor & Brown, 1988; Alicke & Sedikides, 2009; Yasin et al., 2022). Indeed, TMS studies have demonstrated that disrupting mPFC activity while participants rate themselves or their best friend causes them to perceive themselves as less "enhanced" (Kwan et al., 2007; Barrios et al., 2008; Luber & Lisanby, 2014). Furthermore, vmPFC has a crucial role in the performance enhancement caused by self-determined choice (Murayama et al., 2016). In an fMRI study, Murayama et al. (2015) had participants play a game-like task involving a stopwatch. Participants were either allowed to choose the stopwatch (self-determined choice) or not (forced-choice). Self-determined choice improved participants' performance, and neuroimaging revealed that in the forced-choice condition, failure feedback decreased vmPFC activation, while this effect was absent in the self-determined-choice condition. Moreover, vmPFC's resilience to failure in the latter correlated with improved performance. Lockwood et al. (2018) explored the neural correlates of the self-ownership bias, *i.e.* the tendency to attribute a higher value to self- rather than other-owned objects (Thaler, 1980; Pierce et al., 2003), and found vmPFC, along with the ACC, to respond more to self vs. stranger associations. In an fMRI study, Sui et al. (2013) implemented an associative learning paradigm where participants assigned self-relevant labels to three neutral shapes representing themselves, their best friend, or an unfamiliar person. Notably, participants exhibited a strong preference for self-tagged stimuli. Self-tagging correlated with heightened activity in vmPFC and regions involved in social attention (left posterior superior temporal sulcus, LpSTS), and responses in these brain regions predicted behavioural biases favouring self-relevance. Conversely, associations with others engaged a dorsal frontoparietal control network, whose activity was anticorrelated to vmPFC and LpSTS.

The self-schema and its reliance on medial prefrontal regions are apparent from the literature even when considering a different line of research, concerned with the so-called "personal semantics" (PS; Renoult et al., 2012, 2016). For instance, Maguire & Mummery (1999) found that while activations related to personal semantics and general semantics (GS) overlapped in lateral temporal and medial prefrontal regions, the PS condition exhibited higher activation compared to GS. This heightened activation was particularly evident in a left-lateralized network, encompassing the mPFC, retrosplenial cortex, temporal pole, and temporoparietal junction. PS is considered to lie in the middle along the continuum between episodic and semantic memory (Renoult et al., 2012, 2016; Coronel & Federmeier, 2016; Grilli et al., 2018; Tanguay et al., 2023). Accordingly, various brain regions, including mPFC, are differentially responsive to the three types of memory, with a decreasing pattern of activation from autobiographical memory to personal semantics to general semantics (D'Argembeau & Salmon, 2012; Tanguay et al., 2023; Teghil et al., 2024). Here, it is crucial to note that this continuum also overlaps with the distance from one's self-schema, since there is nothing more idiosyncratically related to the self than autobiographical memory (Conway et al., 2000; Nelson, 2015; Charlesworth et al., 2016; Conway & Rubin, 2019).

Indeed, patients with vmPFC damage have been found to struggle in incorporating the self-schema into imagined events (Verfaellie et al., 2019). Also, Ciaramelli et al. (2021) reported how vmPFC patients were incapable to adopt a hypothetical future self-location, failing to classify a series of events as past or future relative to it, which suggests an inability in projecting the self-schema into the future. Finally, Philippi et al. (2012a) reported patients with mPFC damage not to show the mnemonic advantage for self-related items (SRE). For what concerns patients with vmPFC damage, I will avoid spoilers here, and will just refer the reader to Chapters 1 and 2.

As we will discuss in the next section, the self-schema and its disturbance are a key factor in one of the most astonishing and mysterious phenomena known in neuropsychology: confabulation.

2.7. **vmPFC and confabulation**

During the same session I recounted in paragraph 2.4, I was asking the patient to mentally construct a scene and imagine an event starting from an object. This time, the object was a rear-view mirror, and if you read paragraph 2.4, you already know that the patient would just utter some brief sentences for each trial. However, once we got to the rear-view mirror, she started talking about the possibility of being involved in a car accident, and said "*you know, it's exactly what happened to me, and this is the reason I am now being tested by you*", and then went on to describe the dynamics of the accident, how she felt, and how she hoped she would make a complete recovery. Even though this was not what I was asking, I was honestly surprised and pleased at how rich and detailed her description was, although in the back of my mind I had a fuzzy memory of reading her case history, and it mentioning an aneurysm. I just thought "*I must have mixed her up with another patient*" and continued the testing. However, on the following trials, she went back to being as laconic as previously. I figured that probably the rear-view mirror just stirred up an emotionally salient memory, so I did not think much about it. The following days I met with the neuropsychologist that treated her from the beginning, and asked her when did the accident happen, and how the patient was doing in rehabilitation. She said there never was a car accident. She had a ruptured aneurysm.

I was taken aback by how confident and honest the patient seemed when she was telling me about the accident, so how was this possible?

Confabulation happens when a patient generates a false memory without the intention to deceit (Fotopoulou, 2008), hence the term "honest lying" coined by Moscovitch (1989). Confabulating patients normally remain oblivious to inaccuracies and, at times, clings to these false beliefs even when presented with the truth (Gilboa & Moscovitch, 2002). Patients' erroneous convictions are not solely expressed through verbal statements but are also often manifested in their actions (Kopelman, 1987; Kern et al., 1992; Rapcsak et al., 1998; Tallberg & Almkvist, 2001). Confabulations in retrieval contexts often involve false details within a genuine event, potentially representing true memories displaced in time, but sometimes confabulatory contents seem to be entirely fabricated (Talland, 1965; Berlyne, 1972; Moscovitch, 1989; Johnson & Raye, 1998; Nahum et al., 2012). The patient is unaware of confabulation and often experiences anosognosia and lacks awareness of any memory deficit (Weinstein, 1991; Gilboa & Moscovitch, 2002; Fotopoulou, 2010). Confabulations are not intentionally produced and likely do not result from compensatory mechanisms (Baddeley & Wilson, 1988; Johnson & Raye, 1998; Schnider, 2003). While autobiographical recollection triggers confabulations prominently, cases exist where spontaneous confabulation is unrelated to the patient's life. Under specific testing conditions, confabulations may also emerge in semantic memory tasks (Dalla Barba, 1993; Moscovitch & Melo, 1997; Kopelman et al., 1997).

Kopelman (1987) proposed a differentiation between spontaneous and provoked forms of confabulations. The rare spontaneous confabulations are associated with an amnesic syndrome or with frontal damage, whereas provoked confabulations are more common and considered a normal response to flawed memory (Schnider et al., 1996; Gilboa & Verfaellie, 2010). Spontaneous confabulation can occur in some amnesic patients, particularly those diagnosed with Korsakoff's syndrome (Talland, 1965; Benson et al., 1996; Borsutzky et al.,

2008; Van Damme & d'Ydewalle, 2010). However, it is not necessarily accompanied by amnesia, since it has been described in patients with minimal anterograde memory impairment (Kapur & Coughlan, 1980; Papagno & Baddeley, 1997; Feinstein et al., 2000; Nedjam et al., 2000). At a neural level, it is well-known that vmPFC damage alone is sufficient to cause confabulation (Moscovitch & Winocur, 2002; Metcalf et al., 2010; Mendez & Fras, 2011; Gilboa & Moscovitch; Schneider & Koenigs, 2017; Bateman et al., 2023).

Early studies suggested a conceptualization of confabulation as a manifestation of psychological defence mechanisms triggered by embarrassment or the necessity to conceal memory lapses or fill gaps in knowledge (Zangwill, 1953; Weinstein & Kahn, 1955; see also Mercer et al., 1977; DeLuca 2000). However, experimental investigations provided little empirical evidence to support this view (Dalla Barba, 1993; Kopelman et al., 1997; Mercer et al., 1977; Schnider et al., 1996). Furthermore, patients' lack of insight about their own deficit strongly argues against confabulation being somewhat intentional. An alternative hypothesis posits that confabulation may arise from a disrupted sense of chronology, where patients recall real events' contents but struggle with the correct order of occurrences. Consequently, they may misattribute elements of events from one time to those of another. Initially proposed by Korsakoff (Victor et al., 1955) and echoed by others (Talland, 1965; Victor et al., 1971), this hypothesis continues to feature prominently in current theories (Dalla Barba, 1993; Schnider et al., 1996; Ptak & Schnider, 1999; Nedjem et al., 2000; Dalla Barba et al., 2017, 2020). However, explanations exclusively rooted in temporal deficits face challenges when addressing confabulation linked to semantic memory. Many temporality hypotheses have focused on distortions within a personal temporal frame of reference, suggesting confabulations should be limited to the episodic domain. Experimental evidence supporting a temporality disorder in spontaneous behavioural confabulation comes from studies utilizing a continuous recognition paradigm with two runs separated by 1 hour (Schnider et al., 1996; Schnider & Ptak, 1999;

Schnider, 2000; Schnider et al., 2000). Patients exhibiting spontaneous confabulations display a specific disproportionate deficit in the second run, mistakenly selecting stimuli relevant to the previous list but not the current one. This was interpreted as indicative of temporal context confusion, leading to the conflation of memories from widely dispersed time periods. However, accumulating evidence suggests that confabulation can manifest in both semantic and episodic retrieval, something that contradicts the view of confabulation as being attributable purely to temporal confusions (Delbecq-Derouesne et al., 1990; Dalla Barba et al., 1998; Moscovitch & Melo, 1997; Kopelman et al., 1997; Gilboa et al., 2006; Kan et al., 2010). For instance, Moscovitch & Melo (1997) demonstrated that the frequency of confabulations related to content distortion can surpass those linked to temporal distortion, although these two aspects likely interact with each other.

A related perspective on confabulation proposes that it arises from patients' struggle to discern the source of different memories (source monitoring) or to differentiate between real events and imagined ones (reality monitoring; Johnson, 1991; Johnson et al., 1993, Johnson & Raye, 1998; Fotopoulou et al., 2007; Metcalf et al., 2007). Johnson et al. (1997) examined temporal memory (duration and order), source memory (temporal and speaker identification), and reality monitoring (actual vs. imagined autobiographical vignettes) in a confabulating patient (G.S.) and in non-confabulating frontal patients and healthy controls. G.S. exhibited source memory deficits comparable to non-confabulating frontal patients, but with preserved temporal ordering ability. Notably, in contrast to controls and other frontal patients, G.S. supplied more details for imagined vignettes than real ones, indicating impoverished autobiographical memory and a lack of temporal information for them. The recurrent confabulations received similar scores to real memories in terms of amount and type of details provided. Thus, the authors concluded that while deficits in source monitoring may contribute

to confabulation, they only play a role in conjunction with other factors, such as a vivid imagination and an inability to systematically retrieve autobiographical memories.

Finally, strategic retrieval accounts posit that it is in fact a confluence of numerous different factors that lead to confabulation, factors that are exclusively concerned with to the process of memory retrieval (Moscovitch, 1989; Kopelman et al., 1997; Johnson et al., 1997; Baddeley et al., 2003; Gilboa et al., 2006; Gilboa, 2010; Gilboa & Verfaellie, 2010; Ghosh et al., 2014; Hebscher et al., 2016). This arises from the neuropsychological observation that confabulation does not only impacts recent memories acquired after brain damage, but also remote memories obtained long before the occurrence of the injury (Conway & Tacchi, 1996; Gilboa et al., 2006; Gilboa & Verfaellie, 2010). Moscovitch (1989) proposed that confabulation is linked to deficiencies in strategic retrieval, involving impaired search and monitoring processes. Evidence supporting the strategic retrieval hypothesis comes from the performance of confabulators across a range of memory tasks (Moscovitch, 1989; Moscovitch & Melo, 1997). Specifically, confabulating patients tend to produce fewer memories overall, which is indicative of poor search strategies (Della Sala et al., 1993; Kopelman et al., 1999). Also, they benefit more from prompts that aid retrieval, thereby confirming that their search strategy is defective (Moscovitch & Melo, 1997). In instances where a memory is retrieved, deficient monitoring in confabulating patients often leads to a higher frequency of confabulations as compared to other amnesic patients.

A crucial component of the strategic retrieval account is the proposal of a disrupted schematic processing (Gilboa et al., 2006; Hebscher & Gilboa, 2016). Indeed, schemas have long played a prominent role in confabulation theories (Burgess & Shallice, 1996; Burgess & McNeil, 1999; Schnider, 2001; Gilboa, 2004; 2010; Attali et al., 2009). Schemas may contribute to confabulation in two ways: by influencing the content of erroneous memories and by shaping the sense of conviction associated with confabulations (Elliott et al., 2000;

Moscovitch & Winocur, 2002; Gilboa, 2004; 2010; Gilboa et al., 2009; Kan et al., 2010; Hebscher & Gilboa, 2016). Schematic structures are extensively mediated by vmPFC activity, but here, I will refrain from any spoiler on what essentially constitutes the theoretical framework of the present thesis, and I will just refer the reader to the next section.

3. Schemas and schematic processing

3.1. What is a memory schema

3.1.1. History of schemas

The term schema refers to an "adaptable associative networks of knowledge extracted over multiple similar experiences"[2] (Ghosh & Gilboa, 2014; Ghosh et al., 2014). There is notable consensus regarding the fundamental functions attributed to schemas in the existing literature. Various studies consistently attribute three main roles to schemata: guiding behaviour (Head & Holmes, 1911; Rumelhart et al., 1972; Cooper et al., 1995; Kumaran et al., 2009; Rumelhart, 2017); facilitating the encoding of new information, including inferential elaboration (Head & Holmes, 1911; Piaget, 1926; Bartlett, 1932; Carmichael et al., 1932;

[2] To avoid any potential confusion, I will specify here that the words "*schemas*" and "*schemata*" are used as synonyms throughout the present elaborate, and are essentially both the plural forms of "*schema*" (Corcoran, 2006; Berlucchi & Aglioti, 2010).

Bransford & Johnson, 1972; Preston & Eichenbaum, 2013; Emmott & Alexander, 2014; Anderson, 2018); and initiating retrieval processes in the context of memory search and reconstruction (Anderson, 1978; Anderson & Pichert, 1978; Arkes & Freedman, 1984; Nuthall, 2000; Rumelhart, 2017).

Head & Holmes (1911) first used the term "schema" within the context of the body schema (see also Frederiks, 1969). Specifically, they introduced the concept of "postural recognition," which concerns the awareness of one's body position and relies on a schema, acting as a standard against which postural changes are compared to. In this view, schemas, as cognitive structures, aid in interpreting new information, modifying impressions from sensory impulses. This early conceptualization already emphasized the impact that prior information exerts on perception. Piaget (1926) later expanded the scope of schemas to developmental psychology, characterising them as general cognitive structures linking multiple representations of phenomena. Unlike the narrower focus of Head and Holmes, Piaget applied schemas across various cognitive domains, but also highlighting their influence on interpreting new information, particularly in linguistic and perceptual contexts. Bartlett (1932) reintroduced schemas explicitly in the realm of memory, and emphasized their dynamic nature as constantly evolving cognitive structures rather than static arrangements.

Schank & Abelson (1977) coined the concepts of "scripts" and "plans", both resembling some aspects of Bartlett's theorisation of schemas. Scripts outlined event sequences (*e.g.*, a wedding), while plans referred to actions for goal achievement. General knowledge structures like scripts were argued to optimize behaviour evolutionarily (Klein et al., 2002b; Kroes & Fernández, 2012; Shaw & Hazelett, 2014). Within the differentiation between semantic and episodic memory proposed by Tulving (1972), semantic memory networks aligned with schema definitions, since they were conceived as not being limited to factual

information. Tulving also retained the idea that new episodes are interpreted through existing semantic networks.

Also, schema theory significantly impacted educational psychology. Anderson & Pearson (1984) adapted the notion of schema and applied it to reading comprehension, highlighting schema-related functions in both memory encoding and retrieval. In cognitive science, schemas influenced the literature on action selection, and particularly on the ways schemas build and retain associations with goals (Norman & Shallice, 1986; Cooper et al., 1995). In the artificial intelligence literature (Schank & Abelson, 1977; Schank, 1983), computational modelling has provided insights into neural mechanisms guiding behaviour. Specifically, neural networks have applied the notion of schemata to model object recognition and scene analysis, language comprehension, skill learning, attention, and many more (John & McClelland, 1990; Leow & Miikkulainen, 1993, 1997; Lane et al., 2000; Wrigley & Brown, 2004; see Chapter 4).

In the next section we will see how, based on the recent advancements in neuroscience and cognitive psychology, the modern literature has started to operationalize the construct of schemata.

3.1.2. Characteristics of schemas

Ghosh & Gilboa (2014) pinpointed four fundamental attributes that characterise schematic structures: 1) they are an associative network structure, 2) they are based on multiple episodes, 3) they lack unit details, 4) they are adaptable.

Schemas are associative in the sense that they consist of interconnected units, forming an associative network structure. These schema units are referred to as element, events, variables, schema nodes, features, or paired-associates (Schank & Abelson, 1977; Anderson, 1984; Cooper et al., 1995; Halford & Busby, 2007; Tse et al., 2007, 2011; Wang et al., 2012; Van Kesteren et al., 2013; Rumelhart & Ortony, 2017). It is noteworthy that the interrelationships among these units are often considered more critical than the units themselves (Ghosh & Gilboa, 2014). For example, the schematic representation of a scene would include its elements and the spatial relations between them. In neural networks, the relationships amongst elements of a schema have been conceptualized to drive latching dynamics, *i.e.* the probability of the network to jump from a particular configuration to another (Ciaramelli & Treves, 2019; Viol et al., 2021; Spalla et al., 2021; see Chapter 4). The associative network structure is deemed essential because, without units, a schema would lack information, and without their interrelations, the information would be isolated, severely limiting its meaning and functionality. While the associative network structure is a defining feature, it is not sufficient by itself, as it would also include patterns experienced only once. Typically, schemas are portrayed as being built upon multiple variable episodes that share a basic structure.

Indeed, schemas represent broad, high-level constructs that incorporate the shared features or commonalities among events, rather than focusing on the unique details of each event, an idea already expressed by Bartlett (1932). In fact, we would not say we know what normally happens to a conference if we have ever been to just one of them. The features and context of a specific event would give rise to what Robin & Moscovitch (2017) defined as "*gist*" rather than to a schema. The necessity for schemas to be based on multiple episodes is underscored by their role in facilitating the encoding of new information and guiding behaviour

in novel situations, which would be hindered if they were defined by unique episodes (Bartlett, 1932; Rumelhart & Ortony, 2017).

The absence of unit details in schemas directly stems from their reliance on multiple experiences, given the inherent variability between different episodes. Indeed, we would not say that normally, the event schema of a lecture comprises that *professor X* gives a lecture, rather, we would say that *a professor* gives a lecture, something that is only possible because we have attended more than one lecture. According to Rumelhart and Ortony (2017), each unit (or "variable") within a schema possesses constraints, which are conceptualized as distributions rather than strict boundaries. The flexibility of these constraints allows schemas to accommodate significant deviations from the norm. Schemas need to be general to organize new information and provide additional meaning, yet they also need to reserve space for the specific details of new individual episodes.

The fourth and final crucial feature of schemas is their adaptability, which is an idea that appeared in literature more than a century ago (Head & Holmes, 1911, Bartlett 1932). Piaget (1952) identified two ways schemas can be altered: 1) assimilation, *i.e.*, the integration of elements without challenging existing relationships, and 2) accommodation, *i.e.* the alteration of schemas according to new elements. Adaptability has been recognized as a crucial aspect of schemas in the neuroscience literature (Ghosh & Gilboa, 2014). For instance, Tse et al. (2007) found that assimilating new information into existing schemas makes it rapidly hippocampal-independent (see also Tse et al., 2011; Van Kesteren et al., 2010a,b, 2012; Wang et al., 2012; Hebscher et al., 2019). Adaptability is necessary for schemas to efficiently store vast information from diverse experiences and update it in an evolutionary perspective. Schemas must be flexible to support the acquisition of new information associated with similar past contexts and behaviours. Without adaptability, schemas would contain specific

information rather than extracted commonalities, limiting their role in facilitating encoding and guiding behaviour.

Finally, Ghosh & Gilboa (2014) also defined the characteristics that influence schemas, namely: 1) chronological relationships, 2) hierarchical organization, 3) cross-connectivity, and 4) embedded response options.

Some schemas (*e.g.* event schemata, or scripts) are sensitive to temporal order in the sense that, for the schema to maintain its coherence and function, the chronology of its elements must be respected. However, even chronology is a feature sensitive to adaptation. Rumelhart & Ortony (1976) depicted schemas as having a hierarchical structure, also comprising sub-schemas. The authors proposed that this organization enables both top-down and bottom-up activation of cognitive structures, based on whether a schema activates a sub-schema or vice-versa. Cooper et al. (1995) integrated hierarchical organization into the realm of automatic action selection, specifying that attention can be directed at high-level schemas or at a sequence of lower-level schemas. However, Ghosh & Gilboa (2014) suggested that having a hierarchical organization does not mandate for all schemas to comprise sub-schemas or belong to larger encompassing schemas. Hence, a hierarchical organization might be inherent in the way schema units are connected, allowing for the storage of more complex information, but should not be considered a necessary feature of schemas.

Cross-connectivity denotes the existence of shared units between schemas. For example, the action "eat" is part of a "going out for dinner" schema, but is also part of a "family gathering" schema. Also, schemas can communicate with one another, and different sub-schemas can be part of more than one ensemble of higher-order schemata (Bartlett, 1932; Rumelhart & Ortony, 1976). While not deemed essential for functionality by Ghosh & Gilboa (2014), cross-connectivity likely emerges in schemas due to the potential for the same concepts

and sub-schemas to hold different meanings in various contexts, making it a feature to which schemas are sensitive to. Cross-connectivity can also lead to competition among schemas. Cooper et al. (1995) noted that if several higher-level schemas share a sub-schema linked to a specific goal, these schemas compete, and selection would occur based on activation surpassing a pre-defined threshold.

Having embedded response options refers to the fact that schemas also retain information about how the schematic knowledge needs to be used, which links a schema with its goal (Rumelhart, 1980; Goodman, 1980; Cooper et al., 1995; Humphreys & Forde, 1998). However, Ghosh & Gilboa (2014) still consider embedded response options as a sensitive feature rather than a necessary one. Specifically, this sensitivity acquires pivotal importance when a schema guides behaviour, but not all schemas serve this function. For instance, the schema of a scene does not necessarily require associated response options, since a scene is not always linked to a goal. However, according to the authors, the schema would still facilitate the perception of the scene and its encoding and retrieval.

3.2. Neural correlates of schematic knowledge

3.2.1. Acquisition of schemas

The ecological investigation of schema acquisition poses significant challenges, mainly due to the considerable time required to construct these complex cognitive structures

(Tenenbaum et al., 2011; Gilboa & Marlatte, 2017). However, animal studies have begun to elucidate the neural mechanisms involved in the acquisition of schemas. For example, Tse et al. (2007) trained rats in a flavour-place paired associate paradigm arena, which is well-known to be a hippocampal-dependent task. The learned associations resulted in the establishment of a schema, a shared spatial set of trained associations, facilitating rapid acquisition of novel schema-congruent pairs. Remarkably, these new associations became hippocampal-independent within two days, highlighting how systems consolidation can occur extremely rapidly when there is a pre-existing schema into which new information is incorporated (see also Tse et al., 2011; Wang et al., 2012). McKenzie et al. (2014) used an electrophysiological approach to study the representations of related memories in rats that learned associations between locations and rewarded/unrewarded objects. Hippocampal networks established hierarchical structures that organized associated elements from separately acquired memories within the same context. Indeed, distinct organizational patterns emerged for memories in situations where contextual cues played a role in differentiating object-reward associations. Importantly, introducing new items within previously learned contexts revealed the hierarchical schema structure, underscoring how prior knowledge rapidly influences hippocampal neural coding during new learning (see also Baraduc et al., 2019; Zhou et al., 2021; Farzanfar et al., 2023).

Resembling the methodologies used in rats, Sommer (2017) taught (human) participants ten distinctive arrays of object-location associations over the course of nine months. During this phase, there was a transition from hippocampal-mediated retrieval to ventrolateral prefrontal cortex (vlPFC)-mediated retrieval, aligning with the process of semanticization. Subsequently, participants encoded new, related information, and vmPFC orchestrated the integration of this novel information into the already acquired knowledge. However, by 3 months, vmPFC activity was absent, implicating other regions with the retrieval

of highly over-learned associations (vlPFC, anterior temporal lobe, TPJ, the angular gyrus; see also Guo & Yang, 2020; Branzi et al., 2021; Takeuchi et al., 2022; Audrain & McAndrews, 2022).

Wagner et al. (2015) had participants acquire and retrieve two controlled, rule-based schema structures over consecutive days while undergoing fMRI scans. The retrieval process was linked to activation in vmPFC, along with medio-temporal and parietal regions, indicating the successful establishment of a schema during (or after) encoding. During retrieval, the authors observed the convergence of schema components within the angular gyrus, suggesting its role as a confluence zone for low-level visual features and high-level decision rules (see section 3.2.2).

In another fMRI study, Masis-Obando et al. (2022) presented participants with stories with both story-specific and schematic-specific representations. Then, participants recalled each narrative. The authors observed the anterior mPFC to exhibit a significant correlation between the activation of schema representations during encoding and the subsequent behavioural recall performance. Interestingly, this mPFC region, although implicated in schema representation during encoding, did not play a role in schema representation during retrieval.

After acquisition, schematic knowledge serves its function by influencing perception and learning. Hence, the schema has to be reinstated first (*i.e*. activated) and then instantiated (*i.e.* sustained and used; see Gilboa & Marlatte, 2017; Hebscher et al., 2019; Yeshurun et al., 2021; Wing et al., 2021). In the next section, we shall explore how these two processes are implemented and the neural bases they rely on.

3.2.2. Schema reinstatement and instantiation

Several lines of research implicate vmPFC, posterior neocortical cortices, and the hippocampus in both schema reinstatement and instantiation (Gilboa & Marlatte, 2017; Sekeres et al., 2018; Yeshrun et al., 2021; Moscovitch & Gilboa, 2022). However, both in ecological and experimental situations, reinstatement and instantiation often go hand in hand, as schemas, to be instantiated, must be reinstated first. Nonetheless, the literature converges on ascribing a pivotal role to hippocampal-neocortical connections in schema reinstatement and instantiation (Preston & Eichenbaum, 2013; Weilbächer & Gluth, 2016; Robin & Moscovitch, 2017; Bowman & Zeithamova, 2018). In an fMRI study, Schlichting & Preston (2014) used an associative learning paradigm in which participants were exposed to triads of stimuli (e.g., ABC) where only two out of the three possible associations were explicitly taught (*e.g.*, AB, BC), so that knowledge of the third association (AC) reflected associative inference. The reactivation of pre-encoded AB pairs, along with hippocampal interactions with the neocortex, not only predicted the explicitly learned associations but also the inferred AC knowledge. Also, authors found that the mnemonic advantage conferred by previously learned associations extended beyond the initial memories themselves to influence the subsequent encoding of schema-congruent information. The mechanism through which offline reactivation was based on was located in the hippocampal-neocortical connectivity (both functional and structural) and was found to drive to the strengthening of memory traces, guiding integrative encoding (see also Schlichting et al., 2015; Schlichting & Preston, 2016).

Audrain & McAndrews (2022) investigated the behavioural and neural correlates of retrieving schema-congruent and incongruent object-scene associations. When information was congruent with the learned schema, memory exhibited a trend toward generalisation over

time, facilitated by post-encoding coupling between the anterior hippocampus and mPFC. Notably, only schema-congruent representations were integrated in the mPFC after 72 hours, and they were organized based on the schematic context (see also Sekeres et al., 2018; Guo et al., 2023a,b).

Schemata, and more specifically, event schemas (scripts) are also reinstated and instantiated during the perception of events (Zadbood et al., 2017; Zacks, 2020; Lee et al., 2020). Baldassano et al. (2018) presented healthy participants with 16 stories, which comprised four schematic events derived from two distinct scripts (dining at a restaurant or going through the airport). Despite sharing a common script structure, the stories diverged in terms of characters, plotlines, and were presented in two different formats (audio-visual clips or spoken narration). Also, whilst one group of participants were presented with the stories in their original temporal sequence, the other (control) group were presented with the same stories in a temporally scrambled order. mPFC, the postero-medial cortex and superior frontal gyrus displayed schematic event patterns that transcended individual stories, participants, and presentation modalities. Furthermore, mPFC patterns demonstrated sensitivity to the script structure, as highlighted by the fact that temporally scrambled events elicited weaker schematic representations and mPFC activity (see also Reagh & Ranganath, 2023).

Several studies have also emphasized the involvement of posterior cross-modal cortices, such as the angular gyrus and modality-specific cortices during schema-instantiation (van Buuren et al., 2014; Reggev et al., 2016; de Caso et al., 2017; Liu et al., 2017). Indeed, the angular gyrus plays a pivotal role in schema processing, potentially because of its functions in representing multimodal information (Seghier, 2013; Wagner et al., 2015; van der Linden et al., 2017; Hebscher et al., 2019; Giuliano et al., 2021), processing contextual information (Ramanan et al., 2018), and facilitating goal-directed behaviour (Gallivan & Goodale, 2018). Furthermore, across studies, connectivity analyses consistently reveal coactivation patterns

between posterior neocortical structures and the vmPFC and hippocampus (van Kesteren et al., 2013; Brod et al., 2016; Liu et al., 2017; Gilboa & Marlatte, 2017).

Electrophysiological investigations have also started to explore the neural bases of schema instantiation processes, providing insight into its temporal dynamics. For example, Mudrik et al. (2010) studied the event-related potentials (ERPs) implicated in scene-congruent (*i.e.* schema congruent) and scene-incongruent object processing, finding different evoked potential responses for the two categories. Specifically, the scalp distribution of both an early (N300/400) and a late (N650/850) component, reflecting perceptual and semantic processing (respectively), were able to discriminate between scene-congruent and incongruent items. Other studies have described schema-related effects as early as 170ms after stimulus presentation. For instance, Rourke et al. (2016) had cardiologists and pneumologists rapidly evaluate Chest X-rays (CXRs) and electrocardiogram (EKG) whilst EEG was being recorded. Cardiologists, who demonstrated a significantly higher expertise with EKGs than CXRs, also displayed an increased amplitude of the N170 ERPs while reading EKGs compared to CXRs, suggesting that the influence of established knowledge structures can act as early as in the perceptual stages. Also, Gilboa & Moscovitch (2017) had participants view pictures of faces of people with various degrees of familiarity (acquaintances, famous and non-famous people). Their task was to respond positively only to pictures of people they had personally met, thus contrasting new information with their previous experience (*i.e.* their self-schema). ERPs analysis revealed early frontal ERPs that distinguished between self-schema congruent and incongruent trials. Specifically, healthy participants exhibited the posteriorly distributed N170 ERP, with higher N170 amplitudes for personally familiar faces compared to familiar and unfamiliar faces, again suggesting early impacts of schemas on the neural signatures of incoming information.

Now that we have seen how schemas are reinstated and instantiated to influence the perception of new information, we shall explore how schemata affect the way in which this new information becomes memory by guiding (or biasing) encoding, retrieval and consolidation.

3.2.3. Influence of schemas on learning: encoding

Prior knowledge can impact memory processes, influencing both higher-order evaluative and associative binding, as well as lower-level perceptual processing. This impact can enhance or hinder performance, depending on the task and experimental condition. The Schema-Linked Interactions between Medial Prefrontal and Medial Temporal Regions (SLIMM) model proposed by van Kesteren et al. (2012) attributes the mPFC a prominent role in signalling the congruency between existing schemas and current information, referred to as resonance (see also van Kesteren et al., 2013; Durrant et al., 2015; Gilboa & Marlatte, 2017; van Moort et al., 2020; Zacharia et al., 2022). Increased resonance would prompt a shift from hippocampal-dependent memory processing to a neocortical-based learning (particularly, vmPFC-dependent learning). Thus, in this context, vmPFC would activate relevant schematic information and suppress irrelevant details, facilitating but also constraining encoding. For schema-congruent information, vmPFC would thus inhibit hippocampal-dependent binding of arbitrary or unique event features. The SLIMM model has received substantial evidence from behavioural, neuroimaging, and electrophysiological studies (for a review, see van Kesteren & Meeter, 2020).

Sweegers et al. (2015) administered participants a memory task in which faces were paired with homes. Half of the faces allowed responses guided by a schema, while the other half lacked such a schema. The schema was comprised of several pre-learned rules about the combinations of facial features and homes. The authors assessed memory and the depth of processing at encoding, which was quantified using the parietal ERP effect between 500 and 800ms post-stimulus presentation, indicative of successful recollection. Schema-congruency led to significant impairments in item memory and even more substantial impairments in context memory (*i.e.* the associations). Additionally, the parietal old/new ERP effect suggested an enhanced recollection for schema-incongruent memories compared to schema-congruent ones. These combined results imply that when goals can be achieved using existing schemas, the in-depth processing of novel inputs might be hindered, thereby impairing the formation of perceptually detailed and contextually rich memory traces. This result also underlies the function of schemas to provide a scaffolding that generalises across experiences, rather than to process a specific schema-congruent item in depth. Also, encoding information in a schema-congruent (vs schema-incongruent) context can result in increased hits but also increased false alarms for perceptually similar items, something that reinforces the view of schemas being capable to guide the encoding of new information (Spalding et al., 2015). One such example is the effect observed at the Deese-Roediger-McDermott (DRM) paradigm (Deese, 1959). In this paradigm, participants are exposed to word lists containing the most potent associates of a crucial non-presented word, determined by word association norms *(e.g.* presenting "pie, slicer, pear", but not "apple"). During subsequent recall and recognition assessments, participants frequently mistakenly include the non-presented critical word, believing it was part of the initial study material, thereby manifesting the creation of a false, schema-congruent memory (see Graham, 2007; Cann et al., 2011; Jou & Flores, 2011; Pardilla-Delgado et al., 2017).

Berkers et al. (2017) crucially linked this effect to mPFC, demonstrating that disrupting mPFC activity with TMS immediately before the DRM task reduced false memory formation.

In an fMRI study, Bein et al. (2014) presented participants with pairs of semantically related (schema-consistent) and semantically unrelated (schema-inconsistent) words, and measured the functional coupling at encoding between mPFC, parietal posterior regions and the hippocampus. Interestingly, successful recollection was modulated by posterior and hippocampal-mPFC connectivity only in schema-inconsistent events, which might indicate that schema-congruent items are incorporated into an existing schema, and do no need interactions between mPFC, posterior cortical regions and the hippocampus for a successful recollection (see also van Kesteren et al., 2013).

Other studies suggest complementary roles for mPFC and the hippocampus during schema-mediated memory formation (Preston & Eichenbaum, 2013; McKenzie et al., 2013, 2014). For instance, in the associative inference paradigm of Schlichting & Preston (2014), in which participants inferred the non-explicitly taught association in a triad of stimuli, the increased vmPFC-hippocampal functional coupling during rest could reflect the replay of previously learned information. Also, the strength of this functional coupling predicted the success in inferring associations, which suggests that vmPFC drives and monitor hippocampal activity in order to create an integrated schematic representation during encoding or immediately after it (see also Weilbächer & Gluth, 2016; Sommer et al., 2022; Guo et al., 2023a,b).

In an fMRI study, van Kesteren et al. (2010a) had participants watch the first half of a movie in either its original form or with its temporal sequence scrambled. The subsequent day, participants viewed the other half while undergoing scanning. The results indicated that when the initial narrative was presented intact, there was an increased inter-subject

synchronization in vmPFC activity during the encoding of the second part, along with a diminished functional connectivity between mPFC and MTL regions during both the encoding phase and the rest period following encoding. This finding suggests that the alteration of pre-existing schemas results in changes within and among memory-associated brain structures, both during the acquisition of new information and in the subsequent offline period.

However, some studies did not find any mPFC activation or interactions with other regions during schema-related encoding. For instance, McAndrews et al. (2016) employed a continuous recognition paradigm and measured hippocampal and mPFC activation to first and second presentations of scene-object pairs as a function of their semantic congruence. Congruency sped up reaction times, and engaged mPFC whilst also causing inhibition of the hippocampus during recognition. Conversely, when recognising incongruent targets, hippocampal activation was heightened. However, mPFC was not involved in object recognition during the second presentation. Nonetheless, the literature seems to suggest reliable mPFC (and particularly, vmPFC) activation when congruency with an existing schema is *perceived* by participants (van Kesteren et al., 2010a,b, 2020; Brod et al., 2016; Liu et al., 2017; Romero et al., 2019; Raykov et al., 2020).

Once information is encoded, be it coherent or incoherent with an existing schema, it must be retrieved and consolidated appropriately. In the next section, we will delve into how schemata also influence these processes, and the neural correlates associated with them.

3.2.4. Influence of schemas on learning: retrieval and consolidation

The retrieval of stored memories is an active process that can initiate reconsolidation, thus modifying and integrating memories with current information (Mckenzie & Eichenbaum, 2011; Lee et al., 2017). Associative retrieval, particularly when performed iteratively, reveals neural signatures related to both generalization in the mPFC and episodic-specific details in medio-temporal and parietal regions (Gais et al., 2007; Wing et al., 2013; Antony et al., 2017; Ferreira et al., 2019). As mentioned above, one prominent research area that has consistently studied encoding, retrieval and consolidation of schema-mediated memory is that of animal studies (see section 3.2.1). Tse (2007) showed that, in rats, learned flavour-place associations quickly established a schema. When tested at retrieval (after 48 hours), newly acquired schema-congruent associations were found to be hippocampal-independent, which demonstrates an acceleration of consolidation processes for schema-congruent memories.

Guo & Yang (2020) employed a human version of the rodent spatial schema task to investigate brain activity during the immediate retrieval of paired associations in both schema-consistent and schema-inconsistent conditions. The authors reported a heightened anterior hippocampal involvement in retrieving associations in the schema-consistent condition compared to the schema-inconsistent condition. Furthermore, connectivity analyses revealed stronger coupling between the anterior hippocampus and vmPFC when participants successfully retrieved newly learned associations in the schema-consistent condition, while the coupling of the posterior hippocampus with the vmPFC exhibited the opposite pattern.

Guo et al. (2023b) had participant learn paragraphs describing features of unfamiliar words from both familiar and unfamiliar categories, thus, with strong or weak prior schematic representations. Using fMRI, authors observed stronger activation in the anterior-medial

hippocampus when participants correctly retrieved sentences with strong (vs. weak) schematic influence. Conversely, the posterior hippocampus and vmPFC exhibited stronger activation when correctly rejecting new sentences with strong (vs. weak) schematic representation. Moreover, functional connectivity analysis demonstrated stronger coupling between the vmPFC and the anterior-medial hippocampus in the strong schematic representation condition. Taken together, these results underscore the importance of schematic influences exerted by the vmPFC-hippocampal axis during the retrieval of schema-congruent information.

Zeithamova et al. (2012) investigated the effectiveness of retrieval-mediated learning, which involves recalling previous event details during the encoding of related experiences. These conditions for retrieval enhanced participants' ability to deduce connections between distinct events with shared content. Also, changes in activation within a functionally coupled circuit involving the hippocampus and vmPFC corresponded to the development of integrated memories and to a successful inferential memory performance. Thus, the vmPFC-hippocampal axis was demonstrated to have a crucial role in facilitating the formation of schematic memory networks to then support inferential memory.

Memory consolidation involves the gradual transformation of memories and their neural signatures, that are, at first, experience-dependent (Dudai et al., 2015; Squire et al., 2015). Thus, through consolidation, memories alter their susceptibility to interference and forgetfulness over time (Wixted, 2004; Sara & Hars, 2006; Robertson, 2012). However, recent evidence suggest that different types of memory can originate from the same event and can persist for extended durations. Given their reliance on distinct functional and structural neurobiological substrates, memory consolidation of even a single event might be a dynamic process, also comprised of the relations between such types of memory, and might depend on the neural representational correspondence amongst them (Gilboa & Moscovitch, 2021; see also Moscovitch & Gilboa, 2022). However, some studies that specifically focused on the

consolidation of schema-based memories can shed light onto the influence of schematic knowledge on the consolidation process.

Since many of the aforementioned studies on the influence of schematic knowledge on memory encoding also investigated both retrieval and consolidation processes (see section 3.3.3), we have already partially touched on the role of vmPFC in the consolidation of memories. For example, in rats, Tse (2007) demonstrated an accelerated consolidation process for schema-congruent items, which was mediated by cortico-hippocampal interactions. Consistent with findings from rodent studies, several neuroimaging investigations in humans have reported a heightened vmPFC activity during memory consolidation (Gais et al., 2007; Takashima et al., 2006, 2007; Sterpenich et al., 2009). For instance, Takashima et al. (2006) presented participants with a large set of pictures and assessed recognition memory immediately, one day, one week, and three months later. Results indicated a decline in hippocampal activity, accompanied by an increase in vmPFC activity, as consolidation progressed. This augmentation in vmPFC activity also correlated with the decrease in hippocampal activity. In a subsequent study, an increase in vmPFC activity was observed during the retrieval of face-location association pairs that were trained a week before, as compared to those learned on the same day (Takashima et al., 2007).

Gais et al. (2007) employed a sleep deprivation paradigm and had participants learn word pairs. The authors found a consolidation-dependent increase in functional connectivity between the hippocampus and vmPFC three days after encoding, along with a general rise in vmPFC activity six months later (see also Sterpenich et al., 2009 for similar results using emotional pictures).

Based on these pieces of evidence, Nieuwenhuis & Takashima (2011) proposed that the transfer of information between the hippocampus and vmPFC that takes place through the

communication between the two regions enables vmPFC to take on the role (initially hippocampally-dependent) of linking neocortical representations in remote memory. As introduced above, the consolidation process is influenced by the degree of congruency between a memory and prior schematic knowledge (Tse et al., 2007; Hebscher & Gilboa, 2016; Gilboa & Marlatte, 2017; Sommer, 2017). Indeed, it is well-established by the literature that the acceleration of neocortical consolidation is significantly augmented when existing knowledge is concurrently activated alongside incoming information (van Kesteren et al., 2012; Coutanche & Thompson-Schill, 2014, 2015; see also McClelland et al., 1995 and McClelland, 2013 for replicating the effect computationally in neural networks). This simultaneous activation of existing schemata and the perception of schema-congruent information results in synchronized activity among vmPFC and posterior cortical regions (Takashima et al., 2007; Bakker et al., 2015; Gilboa & Moscovitch, 2017). Accordingly, Sommer et al. (2022) reported an increased vmPFC activity and a heightened functional coupling between vmPFC and the precuneus for the consolidation of schema-congruent (vs incongruent) information (see also van Kesteren et al., 2012; Coutanche & Thompson-Schill, 2014, 2015; Bakker et al., 2015; Gilboa & Moscovitch, 2017).

Numerous pieces of evidence supporting rapid neocortical consolidation processes in humans come from sleep studies demonstrating unique patterns of sleep-related memory reactivation and transformation (Gilboa & Marlatte, 2017). Indeed, sleep is known to facilitate the abstraction of rules to form new schemas (Wagner et al., 2004; Pace-Schott et al., 2012; Stickgold & Walker, 2013), the integration of knowledge into existing schemas (Paller & Voss, 2004; Dumay & Gaskell, 2007; Tamminen et al., 2010) and creativity that requires the destruction of schemas that are not useful anymore (schema disintegration; Stickgold et al., 1999; Walker et al., 2002; Ritter et al., 2012). Landmann et al. (2014) suggested that the emergence of schemas and the integration of information into existing schemata might be

preferentially promoted by slow wave sleep, whereas the dismantling of schemas might be primarily promoted by REM sleep (Cai et al., 2009; Diekelmann & Born, 2010; Lewis & Durrant, 2011; Sio et al., 2013). Durrant et al. (2015) studied the role of vmPFC in consolidation processes during REM sleep. Participants were tasked with memorising 32 melodies, half conforming to a tonal schema which was culturally shared, and half deviating from this schema. After a 24-hour consolidation interval, participants had to learn 32 additional melodies, and then underwent a recognition test featuring melodies from both sessions, alongside previously unheard lures. Results showed that participants exhibited better recall for schema-congruent melodies, particularly after consolidation, suggesting a preference for consolidating schema-conformant items over a 24-hour period. The authors also monitored overnight sleep between sessions and found that the extent of consolidation benefit for schema-conformant items correlated with the amount REM sleep obtained and EEG theta power in frontal regions during REM sleep. These findings are consistent with the view that mPFC regions signals the congruency between existing schemas to promote consolidation processes, thus align with the SLIMM model (Van Kesteren et al., 2012).

Now that we have tapped into the characteristics of schemata, their functional significance, their neural bases, and the way in which they bias incoming information and memory formation, we shall see how a lesion of the area that seems to orchestrate and exert this schematic influence can impact schematic processing. Thus, in the next section, I will review neuropsychological evidence that inextricably links vmPFC to schemata.

3.3. **Schema processing in patients with vmPFC lesions**

Due to the somewhat recent advancements in lesion mapping techniques, early neuropsychological studies on schematic processing primarily focused on individuals with lesions spanning the whole frontal (and later, prefrontal) lobe (Sirigu et al., 1995; Godbout & Doyon, 1995; Allain et al., 1999, 2001; Godbout et al., 2004; Wood et al., 2005). Specifically, these studies centred their investigation on event-schemas (scripts) knowledge in frontal patients. Although the precise nature of their impairment varied across studies, likely due to differences in tasks and scoring procedures, all investigations indicated some degree of degradation of event-script knowledge amongst frontal patients. This included issues such as sequencing or boundary errors, selection of incoherent headlines, or impairments in correctly categorising actions within their respective script (see also Grafman et al., 1993; Zanini, 2008).

In the recent literature, numerous studies have supported the idea that a damage to the vmPFC leads to deficiencies in schema reinstatement and instantiation, and in the inhibition of currently irrelevant schemata. Ghosh et al. (2014) asked healthy participants and vmPFC damaged patients to determine whether a word was related to an event schema (*e.g.*, "dine at a restaurant") and then, ten minutes later, to a second, distinct schema (*e.g.*, "attend a wedding"). vmPFC patients that were prone to confabulation (vs non-confabulators) exhibited difficulties in correctly pair a word to its relevant schema. Of note, Giuliano et al (2021) later reported the same impairment even in non-confabulating vmPFC patients. These results suggest that even in situations where memory is not explicitly challenged, vmPFC lesions hinder the ability to maintain an active schema as a template for processing incoming information (Gilboa & Marlatte, 2017).

Cameron et al. (2018) asked participants to make moral judgments of a word (morally wrong, non-moral negative, or neutral) following a distractor word from one of the same three categories. Patients with vmPFC lesions displayed a stronger tendency to make incorrect judgments after a morally wrong distractor, consistent with schema intrusion from the moral category when the distractor was incongruent with the target word. vmPFC patients were also found to exhibit a weakened schematic benefit on memory encoding. Spalding et al. (2015) showed participants word-picture pairings that were either schema-congruent (*e.g.* toothbrush and bathroom) or incongruent (*e.g.* bathing suit and office). In a subsequent recognition phase, healthy controls were more likely to correctly recognise items encoded in the schema-congruent condition, indicating that the schema had promoted memory encoding and/or retrieval (see section 3.2.3). Conversely, vmPFC patients failed to show such an influence from schematic knowledge, instead displaying a similar recall performance in both congruent and incongruent conditions.

However, in some circumstances, this lack of schematic influence on memory encoding and retrieval can paradoxically lead to an improved performance in vmPFC damaged patients than in healthy participants. In the Deese-Roediger-McDermott (DRM) paradigm, participants are presented with a list of semantically associated words and are later administered a recognition task. Whereas healthy participants are more likely to commit false alarm mistakes for critical lures (see section 3.2.3), individuals with vmPFC damage are not, indicating weakened schema-related processing (Warren et al., 2014; see also van Kesteren & Brown, 2014; Schlichting & Preston, 2015; Berkers et al., 2017). Nonetheless, Ciaramelli et al. (2009) demonstrated that a high cognitive load (*i.e.* divided attention) during retrieval suppressed false recognition of lures in confabulating patients, but increased false recognition in non-confabulating patients and healthy controls.

vmPFC lesions have also been strongly linked to the degradation of schematic processes involved in the successful recollection of inferred associations. Spalding et al. (2018) had participants learn pairs of objects structured as AB pairs and BC pairs, with A and C never shown together, but whose associations could be inferred from the item B, presented with both of them separately (see also Schlichting et al., 2015; Schlichting & Preston, 2016). Patients with vmPFC damage performed comparably to healthy controls in recalling direct associations (AB and BC) but were impaired in identifying the inferred AC pair. Koscik & Tranel (2012) studied the role of vmPFC in the process of making transitive inferences, such as the logical operation that if A > B and B > C, then A > C. Participants first learned the relationships between patterns (*i.e.* A, B and C) and were then presented with novel pairings, some of which necessitated transitive inference. Patients with vmPFC damage demonstrated a selective impairment in transitive inference, indicating that vmPFC is essential for its normal execution. Such a result reinforces the view that complex inference processes involved in schema-like mechanisms do indeed depend on vmPFC integrity (see also Yu et al., 2020; Wing et al., 2021).

Other investigations of vmPFC patients' performance in schema-related tasks have also tapped into the temporal dynamics of schematic processing following vmPFC damage. For instance, Gilboa & Moscovitch (2017) presented healthy controls and vmPFC patients with images of acquaintances, along with famous and non-famous individuals whilst recording EEG activity (see section 3.2.2). Participants were instructed to respond positively only to pictures of individuals they had personally met (personal familiarity). Here, the self-schema operated as a superordinate cognitive schematic structure, aiding in the accurate endorsement of acquaintances and the exclusion of non-personal but familiar faces. Control participants exhibited pre-stimulus theta coherence desynchronization between mPFC, inferotemporal, and lateral temporal cortices. However, these oscillatory coherence patterns were notably diminished in patients with vmPFC damage, particularly those with an history of confabulation.

Crucially, the authors found that the pre-cue cortico-cortical desynchronizations modulated the posterior cortical N170 (indexing automatic memory processes). Furthermore, this pre-cue desynchronization predicted an early post-cue frontal positive component (P230) and response accuracy. Thus, the authors suggested that vmPFC plays a pivotal role in biasing posterior neocortical long-term memory representations, thereby enhancing automatic memory cue processing and driving frontally-mediated rapid memory monitoring (P230). Therefore, a vmPFC damage results in inaccurate, context-irrelevant activation of schemas which in turn cause an impairment in monitoring signals (see also Gilboa, 2004; Hebscher et al., 2016; Hebscher & Gilboa, 2016).

Giuliano et al. (2021) investigated the role of vmPFC in the reinstatement and instantiation of schemas and semantic categories by examining network-level oscillatory dynamics with EEG. Healthy controls and vmPFC patients were tasked with classifying words based on a given schema or category. Given that reinstatement is a preparatory process, to examine its neural signatures the authors specifically focused on oscillations occurring 500 milliseconds before stimulus presentation. On the other hand, for instantiation, which takes place at stimulus presentation, the authors examined oscillations happening between stimulus presentation and 1000 milliseconds post-stimulus. Crucially, reinstatement was linked to pre-stimulus theta and alpha desynchrony between vmPFC and the posterior parietal cortex for schemas, and between vmPFC the lateral temporal lobe and inferotemporal cortex for both schemas and categories. Notably, damage to the vmPFC affected both schemas and categories, but individuals with damage to the subcallosal vmPFC exhibited schema-specific deficits. The instantiation phase exhibited similar oscillatory patterns in the post-stimulus time frame, but in the alpha and beta frequency bands. These findings seem to suggest a partial overlap for the reinstatement and instantiation of schemas and categories, whilst underscoring the role of vmPFC and its interaction with the temporal cortex in the two domains. The authors further

suggested that the involvement of vmPFC and the lateral temporal cortex in both schema and category processing might indicate that prior knowledge exists on a spectrum, which is congruent with the idea that semantic category information is a large component of schemas, since it constitutes the schema elements (Gilboa, 2004, 2010; Ghosh & Gilboa, 2014; Gilboa & Marlatte, 2017).

Now that we have reviewed the neuropsychological evidence of the pivotal role of vmPFC in schematic processing, in the next and final section of this general introduction, we shall delve into how vmPFC patients' impairments in schematic processing might be at the bases of the deficits they typically show in other cognitive domains (see section 2).

3.4. **The role of schemata in cognitive functions linked to vmPFC**

To account for the deficits observed in vmPFC patients in autobiographical memory recollection, episodic future thinking and scene construction (see sections 2.1, 2.2 and 2.3), Ciaramelli et al. (2019) proposed a model on the role of vmPFC, the hippocampus, and their interactions during such processes. The model stems not only from neuropsychological investigations, but also from fMRI studies (Okuda et al., 2003; Addis et al., 2007; Szpunar et al., 2007; Campbell et al., 2018; Lieberman et al., 2019), and has been supported by MEG studies that explored the temporal dynamics of neocortical-hippocampal interactions during scene construction and autobiographical memory recollection (Barry et al., 2019; McCormick et al., 2020; Monk et al., 2020, 2021; see also Nawa & Ando, 2019; Roehri et al., 2022). Specifically, the authors suggested a role of vmPFC as the initiator for activating schematic

and relevant knowledge associated with a specific event within the neocortex. Simultaneously, vmPFC would inhibit elements that are deemed irrelevant, a suggestion that resembles Gilboa et al. (2006, 2009) proposal of a monitoring role of vmPFC in schema-mediated cognition. This selective information is then communicated to the hippocampus, which, in turn, constructs a comprehensive scene snapshot based on the elements of the event. Subsequently, vmPFC engages in iterative processes through feedback loops with other neocortical structures and the hippocampus. These dynamic interactions facilitate and orchestrate schema-based retrieval, monitoring, and sequencing, all of which are indispensable for the construction of successive scenes that collectively form the unfolding mental event (Benoit et al., 2014; Bertossi et al., 2016a,b, 2017a; McCormick et al., 2018a, 2020; Verfaellie et al., 2019; see Fig. 4). According to this view, vmPFC may not directly support scene construction, autobiographical memory recollection and episodic future thinking *per se* but rather the implementation of schematic structures essential for them (Ciaramelli et al., 2019; Barry et al., 2019; Monk et al., 2021)[3].

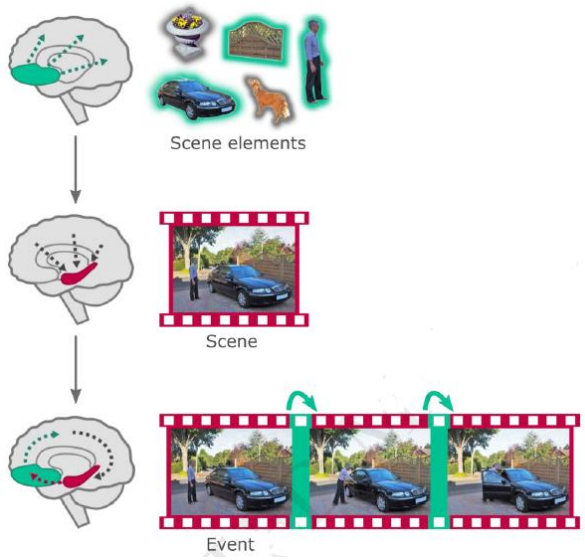

Figure 4. Schematic representation of the model proposed by Ciaramelli et al., (2019) on the role of the vmPFC-hippocampal axis during scene construction, autobiographical memory recollection and episodic future thinking. The vmPFC is depicted in green, the hippocampus in red (figure from Ciaramelli et al., 2019).

[3] Of note, some authors refer to these kinds of schematic structures that enable scene and event construction as "scaffolding", that are a strikingly similar concept (Irish, 2020; Masís-Obando et al., 2022). However, within the schema theory, providing scaffolding is one of the functions of schemata (Ghosh & Gilboa, 2014; Guo & Yang, 2023), but here we note that in the context of scene construction, autobiographical memory recollection and episodic future thinking, the term "scaffolding" is often used with a similar meaning to that of "schemata".

Moreover, the more pronounced deficits that vmPFC patients display in constructing future (vs. fictitious) scenarios (Bertossi et al., 2016a) might stem from the higher demands on schematic structures imposed by imagining one's personal future. Indeed, the Construal Level Theory (CLT; Lieberman & Trope, 1998; Trope & Liberman, 2003, 2010) posits that imagining an event that is distant in the future requires more mental construal, drawing less from direct experience of the event. As a consequence, people tend to imagine distant future events in more abstract, rather than concrete terms (Lieberman et al., 2002; Trope & Liberman, 2003; Förster et al., 2004). Thus, this mental construal requires the activation of numerous schemata, including the self-schema, and the schemas of one's future goals and motivations, which are then projected into a future scenario, something that heavily relies on vmPFC activity (D'Argembeau et al., 2010; O'Doherty, 2011; D'Argembeau, 2013; Stawarczyk & D'Argembeau, 2015; Reber et al., 2017; see Chapter 1). Accordingly, Rosenbaum et al. (2023) reported that some of the vmPFC patients tested in an event construction task did not show the typical construal-level effects when imagining future scenarios.

A similar process of vmPFC orchestrating the activation of schemata in neocortex has been proposed to also take place during the initiation of mind wandering (Ciaramelli & Treves, 2019; see section 2.4). In this view, vmPFC might be responsible for triggering the reinstatement and instantiation of schematic structures that allow the hippocampus to construct the perceptually-decoupled experience that is mind wandering. Then, vmPFC would also monitor the ongoing mind-wandering process through feedback loops with the hippocampus. Thus, at a computational level, schemata act as a collection of local attractor networks that drives "latching" dynamics, wherein the neocortex does not just fixate into a single attractor, but instead continually hops from one attractor to the next (Treves, 2005; Ciaramelli & Treves, 2019; Spalla et al., 2021; Viol et al., 2021).

Yu et al., (2020) recently proposed that the schema-related deficits observed in vmPFC patients could also constitute the bases of their impairments in social cognition, emotional regulation, and theory of mind (see section 2.5). In particular, the ability to discern emotions from subtle cues involves making inferences about partially concealed states, something that requires the activation and usage of schemata of emotional states (Heberlein et al., 2008; Tsuchida & Fellows, 2012; Jenkins et al., 2014; Willis et al., 2014; Wolf et al., 2014; Andrewes & Jenkins, 2019). Notably, neuroimaging evidence suggests that vmPFC is indeed engaged in grouping or separating fearful stimuli into distinct schemas for extinction (Kalisch et al., 2006; Gershman et al., 2013). Furthermore, impaired schematic processing may offer an explanation for vmPFC patients' deficits observed in the social domain. Indeed, patients with vmPFC lesions tend to downplay context and instead concentrate on observable outcomes when making moral judgments or social decisions (Ciaramelli et al., 2012, 2013; Moretto et al., 2013). Yu et al. (2020) suggested that these pieces of evidence might indicate that vmPFC damaged patients display difficulties in bringing forth relevant social schemas. Moreover, issues related to social appropriateness, such as the inability to recognise social blunders or generate appropriate solutions to social problems, could stem from a degradation of schemas governing social norms (Pullen et al., 2006; Peters et al., 2017). Hence, an impairment in social schema representations would also hinder the performance in theory of mind tasks (ToM; Yu et al., 2020). Nevertheless, a separate body of research might suggest a link with ToM-related functions and the self-schema, which, as it is one of the richest and most complex schemas we have, is supposed to be degraded in vmPFC-lesioned patients (see section 2.6 and Chapters 1 and 2). Indeed**,** when participants make judgments about other people, they seem to default to using their own selves as a reference point (Srull & Gaelick, 1983; Smith, 1984; Catrambone et al., 1986; Wagner et al., 2012).

Finally, the phenomenon of confabulation has historically laid the foundations for the proposals put forward by the schema theory and the supposed role of vmPFC in schema-mediated cognition (Moscovitch & Melo, 1997; Gilboa & Moscovitch, 2002; Moscovitch, 2005; Gilboa et al., 2006; Gilboa & Verfaellie, 2010; see section 2.7). Indeed, the process of memory reconstruction relies on an organizing structure (*i.e.*, a schema), to systematically assemble details into a cohesive narrative (Nieuwenhuis & Takashima, 2011; Preston & Eichenbaum, 2013; Daviddi et al., 2023). The content of confabulations can in fact be influenced by pre-existing schemas (Burgess & Shallice, 1996), especially when these schemas are closely tied to the self-schema (Moscovitch & Melo, 1997; Gilboa, 2004, 2010; Gilboa et al., 2006). Indeed, Gilboa et al. (2006) reported vmPFC patients to falsely recognise as true statements that were blatantly inconsistent with their self-schema and life history, something that inextricably links confabulation and self-referential cognition.

Within the context of the schema theory, vmPFC lesions are proposed to lead to confabulation by damaging the ability to monitor and control the activation of inappropriate schemas. Consequently, these inappropriate schemas may serve as a template for memory reconstruction, contributing to the creation of inaccurate or distorted memories. Thus, this perspective suggests that the vmPFC plays a crucial role in orchestrating the activation of relevant schemas and preventing the intrusion of inappropriate ones during the memory reconstruction process (Gilboa et al., 2006; Ghosh & Gilboa, 2014; Hebscher et al., 2016).

Now that we have provided a general introduction on the nature, functions, and neural bases of schemata and their critical reliance on vmPFC, we shall start to explore the effects that a damage to vmPFC might have on the most cardinal, identity-defining schemas we will ever have: the self.

Chapter 1. Memories of myself: the status of self-related information in memory

In this chapter, I will examine the neural bases of the mnemonic advantage for self-related information in memory. Reading the literature, one can easily stumble upon endless discussions and speculations on whether the self-schema retains a special status in the brain. However, the word "special" is vague and undefined: what does it mean to be special?

A better, and more precise question we could ask is whether information that pertains to the self is treated differently than information related to someone else, for example in memory. In doing so, we could also advance the investigation and ask whether this *special* treatment is imposed by a specific part of the brain, and whether a damage to that specific region would reduce the self to a *not-so-special* schema.

Study 1. Present and future self in memory: the role of vmPFC in the self-reference effect*

*This work has been published in Stendardi, D., Biscotto, F., Bertossi, E., & Ciaramelli, E. (2021). *Social Cognitive and Affective Neuroscience*, *16*(12), 1205-1213.

Introduction

We often find ourselves thinking about who we are: whether we are introvert and why, what are our music preferences, favourite clothes, places, philosophers, what is that drives us crazy or that instead we wish for the future. Instances of self-knowledge such as these revolve around the self-schema, an articulated set of beliefs about oneself, generally deriving from the repeated categorization and subsequent evaluation of one's behaviour, which defines our

identity and biases the way we process incoming information (Markus, 1977). Self-knowledge (*e.g.* 'I am an introvert person') is at the border between episodic memory, our ability to recollect personal experiences within their unique spatio-temporal context (*e.g.* 'yesterday at the party I only talked to Francesca'), and semantic memory, our (culturally shared) knowledge of facts and concepts by now detached from the context of acquisition (*e.g.* 'Introvert does not mean shy'), as it is at the same time personal and devoid of context (Renoult et al., 2012). Self-knowledge is dissociated from episodic and semantic memory. For example, patients with episodic amnesia due to medial temporal lobe (MTL) damage typically have preserved self-knowledge (Klein et al., 1996; Klein and Lax, 2010; Picard et al., 2013), and self-relevant semantic concepts can be preserved in semantic dementia (Westmacott et al., 2001). Self-knowledge is also dissociated from other personal semantic information, such as repeated events characterizing lifetime periods (*e.g.* 'In high school, I would hang out only with Francesca'), which are associated with greater contextual detail, and often impaired in MTL amnesia (St-Laurent et al., 2009).

What are the neural bases of self-knowledge? One way to investigate this is to study the footprints self-knowledge leaves on new learning. Rogers et al. (1977) found that trait adjectives processed in relation to the self (*e.g.* 'are you an introvert person?') are remembered better than trait adjectives processed for their phonetic or structural properties (*e.g.* 'does introvert rhyme with disconcert?'), their meaning (*e.g.* 'does introvert mean the same as shy?'), or even in relation to another individual (*e.g.* 'is she an introvert person?'; Klein & Kihlstrom, 1986; Symons & Johnson, 1997; Kelley et al., 2002) - a phenomenon called 'self-reference effect' (SRE; Rogers et al., 1977). In a functional magnetic resonance imaging (fMRI) study, Kelley et al. (2002) found that the medial prefrontal cortex (mPFC) was selectively activated in the self-related condition and not in the other-related or lexical conditions (see also Craik et al., 1999). Macrae et al. (2004) showed that activity in the mPFC predicted both judgments of

self-relevance for trait adjectives and the SRE in memory, and Kim & Johnson (2012) extended the finding to objects owned by the participants, which were associated with increased subjective value, memorability and mPFC engagement compared to other people's objects. Together, these findings point to medial prefrontal regions as implicated in self-related processing - a finding that has been corroborated by several meta-analyses (Denny et al., 2012; see also Northoff et al., 2006; Lieberman, 2010; Murray et al., 2012 for reviews), which point to Brodmann's area (BA) 10 as the most prominent cluster of self-related activity (Lieberman et al., 2019). Consistent with this, patients with mPFC lesions (centred on BA 10) were found to not show the SRE in memory (Philippi et al., 2012a).

We do not just reflect on how we are currently, but also on how we were in the past or predict how we will be in the future. Thinking about the future shares component processes with remembering the past (Buckner & Carroll, 2007; Schacter et al., 2012) and is as fractionated a process as is remembering the past (Addis et al., 2007; D'Argembeau & Mathy, 2011; D'Argembeau, 2020). For example, patients with MTL amnesia cannot imagine specific future events but can report semantic (including autobiographical) information about the future (Race et al., 2011) and can think about (Kwan et al., 2013) and self-project into the future in abstract terms (Arzy et al., 2009). An important question is how we represent our past and future selves. D'Argembeau et al. (2008) asked participants to reflect on their current traits, their traits in the past, and on the current and past traits of another individual. They found that both ventral mPFC (vmPFC) and dorsal medial prefrontal cortex were more active when individuals reflected on their current vs past selves and that there was no difference in medial prefrontal activity between the past-self and the 'other' condition, as if the past self were perceived, to some extent, as another individual, due to the perceived change, with time, in one's characteristics, activities and goals (Libby & Eibach, 2002; Pronin & Ross, 2006). Similarly, Ersner-Hershfield et al. (2009) found diminished vmPFC activity for the future vs

present self (see also D'Argembeau et al., 2010a,b), and a recent study confirmed that vmPFC activity while reflecting on our future self in 10 years is more similar to that observed while we think to another individual compared to our current self (Mitchell et al., 2011). None of these studies, however, has investigated whether or not the future self also has a privileged status in memory, and, in case it does, whether the future SRE would also be mediated by mPFC regions.

The aim of the present study is two-fold. First, we wish to confirm that the vmPFC is a crucial substrate of self-knowledge, showing that vmPFC damage is associated with a reduced SRE (as in Philippi et al., 2012a). There are several reasons to think that vmPFC is related to the SRE. This region is commonly activated during tasks requiring self-reflection (Jenkins et al., 2008; Wagner et al., 2012), and vmPFC patients are impaired in self-monitoring (Beer et al., 2006; Hiser & Koenigs, 2018) and reportedly unable to introspect and daydream (Ackerly & Benton, 1948; Wheeler et al., 1997; Bertossi & Ciaramelli, 2016). Additionally, vmPFC patients have been found to use fewer self-references than healthy and brain-damaged controls while narrating personal events, as if they failed to fill constructed experience with self-related content (Kurczek et al., 2015).

Second, we investigated whether vmPFC is a crucial underpinning of future self-knowledge, by additionally testing whether items related to the future self also give rise to an SRE in memory, and whether the future SRE, too, depends on vmPFC integrity. Previous neuropsychological work has shown that vmPFC damage impairs several components of future thinking, such as the ability to imagine specific future events (Bertossi et al., 2016a,b, 2017a; Verfaellie et al., 2019) and also to self-project into future time periods in more abstract terms (Fellows & Farah, 2005; Sellitto et al., 2010; Ciaramelli et al., 2021). However, vmPFC patients can normally report on semantic facts about their personal future (*e.g.* 'In my 70s I will be retired'; Bertossi et al., 2016a,b, 2017a; Verfaellie et al., 2019), suggesting that future

personal semantics, including knowledge about one's future self, may be retained in these patients. The fMRI evidence that vmPFC responds less to the future than to the present self (Ersner-Hershfield et al., 2009; D'Argembeau et al., 2010a) also leads to the prediction that vmPFC patients, compared to controls, would have an impaired representation of their present self, but not necessarily of their future self. fMRI evidence, however, is correlational in nature, and, therefore, lesion studies are necessary to clarify the functional interpretation of brain activity and its relation to behaviour. To this aim, we had vmPFC patients and brain-damaged and healthy controls judge whether each of a series of trait adjectives was descriptive of their present self, future self, another person and that person in the future and then to recognize them among distractors. If the representation of the future self is similar, to some extent, to that of another person (Parfit, 1971; Pronin & Ross, 2006), then the future SRE should have a smaller magnitude compared to the present SRE. Moreover, based on fMRI evidence (Ersner-Hershfield et al., 2009; D'Argembeau et al., 2010a), we predict that vmPFC patients would show a reduced present SRE (as in Philippi et al., 2012a) but a normal future SRE.

Finally, we sought to begin to shed light on the cognitive bases of the SRE and on possible reasons of SRE anomalies in vmPFC patients. Self-referenced (as opposed, for example, to phonetic) item processing is thought to lead to deep encoding. This is because incoming information is evaluated against the self-schema: participants compare trait adjectives with their self-view. This comparison can have variable epistemic and emotional consequences. Participants may be more or less certain that they possess (or not) a given trait (the 'epistemic investment' in the self-view, to say it with D'Argembeau et al., 2012), which depends on the amount and consistency of information one has about this aspect of the self in the self-schema (Pelham, 1991), and they may place more or less importance on having (or not) a trait (the 'emotional investment'), which reflects the extent to which the trait is related to one's personal goals and motives (Pelham, 1991). The vmPFC is implicated in schema-

related processing (Ghosh et al., 2014) and deemed to generate confidence signals based on the match between incoming information and the self-schema (Hebscher & Gilboa, 2016). vmPFC is also known for its role in emotion and valuation (Lieberman et al., 2019). D'Argembeau et al. (2012) indeed found that left BA 10 tracked the certainty of having a trait and right BA 10 tracked its perceived importance, suggesting that vmPFC represents the epistemic and emotional value of trait items. One possibility, therefore, is that the strength of epistemic and emotional responses to trait adjectives relates to the efficacy with which these items are encoded in memory and that the lack of SRE in vmPFC patients is associated with a reduction of these responses. To test this, we asked participants to judge, for each trait, how certain they were to possess or that they will possess that trait and the importance they attached to it.

Methods

Participants

Participants included 15 patients with brain damage and 23 healthy individuals. Patients were recruited at the Centre for Studies and Research in Cognitive Neuroscience, Cesena, on the basis of their lesion site, as documented by MRI or computerized tomography (CT) scans. Seven patients had lesions involving vmPFC (vmPFC patients; 7 males; mean age = 57 years, range = 43–74; mean education = 10 years, range = 5–13; see Table 1 for individual patients' demographic and neuropsychological data). vmPFC patients' lesions resulted, in all cases, from the rupture of an aneurysm of the anterior communicating artery. They were bilateral in six cases and right-lateralized in one case. The remaining eight patients had brain lesions that did not involve vmPFC (7 males; mean age = 61, range = 41– 74; mean education = 11 years, range = 5–18). Control patients' lesions were caused by ischemic or haemorrhagic stroke, traumatic brain injury or brain tumour and were in the left hemisphere in three cases and in the right hemisphere in five cases. Lesion sites mainly included the occipital cortex, extending into the

occipito-temporal area (six cases) and the fronto-parietal cortex (one case). For one of the eight control patients the lesion description was available but MRI scans were not, and therefore we could not reconstruct precisely the extension of the lesion. There was no significant difference in lesion volume between vmPFC patients and the remaining seven control patients (57 vs 33 cc., $p = 0.18$). Included patients were in the stable phase of recovery (at least 3 months post-morbid). The healthy control group comprised 23 participants without neurological or psychiatric history (21 males; mean age = 57, range = 47–74; mean education = 11 years, range = 5–18), which were matched to patients on age, education ($F_{2,35} < 0.84$; $p > 0.43$ in both cases) and gender balance ($\chi 2 < 0.94$, $p > 0.32$ in all cases). vmPFC patients' sample size was based on a previous study on the SRE in vmPFC patients (*i.e.* Philippi et al., 2012a: 6 vmPFC patients, 15 healthy controls and 8 control patients). A somewhat larger N was chosen for control participants (23 healthy controls and 8 control patients), based on the average effect size of the SRE ($d = 0.5$) in a meta-analysis of 129 studies (Symons & Johnson, 1997), which required a sample size of $N = 27$ to be replicated ($p = 0.05$) with a statistical power = 0.80. Participants gave written informed consent to participate in the experiment, which was performed in agreement with the 2008 World Medical Association Declaration of Helsinki, and approved by the Bioethical Committee of the University of Bologna and the Ethical Committee of Area Vasta (CEIIAV) of Emilia Romagna.

	vmPFC patients						
	p. 1	**p. 2**	**p. 3**	**p. 4**	**p. 5**	**p. 6**	**p. 7**
Sex	M	M	M	M	M	M	M
Age (years)	53	65	51	43	74	54	60
Education (years)	8	13	13	13	5	8	13
Raven Standard Matrices (cut-off = 15)	23.25	20	19	23.25	22	28.5	-
Attentional Matrices (cut-off = 31)	48.5	35	49.5	42.25	57	54.5	49
Phonemic fluency (cut-off = 17)	27	22	32	21	18	36	20
Semantic fluency (cut-off = 25)	37	36	35	40	34	61	34
Wisconsing Card Sorting Test perseverative errors (cut-off = 42)	41	64*	28	64*	-	87*	38
Short-term memory - Digit span (cut-off = 3,75)	5	5.75	5.75	6.5	5.5	5	2.75*
Short-term memory - Corsi tapping test (cut-off = 3,75)	4.75	4.75	3.5	5.5	4	5.75	2.75*
Long-term memory - Prose passage recall (cut-off = 4,75)	5	12.5	13.5	13	9.2	8.6	5.7
Rey Complex figure Copy (cut-off = 28.9)	32.5	36	36	36	-	35.5	30.25
Rey Complex figure Delay (cut-off = 9.5)	6.75	9.9	22	19.5	-	15.75	17.25

Table 1. Patients' demographic and clinical data.

Note: The table reports, for each patient (p), scores corrected for age, education and gender according to normative samples. For each test, we also report the cut-off score. Scores below the cut-off are considered indicative of impaired performance (corresponding to a percentile < 5), and signaled by an *. Dashes indicate missing data.

2.2 Lesion analysis

Patients' individual lesions, derived from the most recent MRI or CT scans were manually drawn by a trained neuroscientist directly on each slice of the normalized T1-weighted template MRI scan from the Montreal Neurological Institute distributed with MRIcro (Rorden & Brett, 2000). The MRIcro software was used to estimate lesion volumes (in cc) and generate lesion overlap images. Figure 1 shows the extent and overlap of brain lesions in vmPFC patients. Brodmann's areas (BA) mainly affected were BA 10, BA 11, BA 24, BA 25,

and BA 32, though one patient also had damage to lateral prefrontal regions involving BA 9, BA 46, and BA 47, which accounted for 4-9 % of his total lesion size. The region of maximal lesion overlap occurred in BA 11 (M = 21.51 cc, SD = 8.79), BA 10 (M = 12.93 cc, SD = 5.35), and BA 32 (M = 8.41 cc, SD = 4.33).

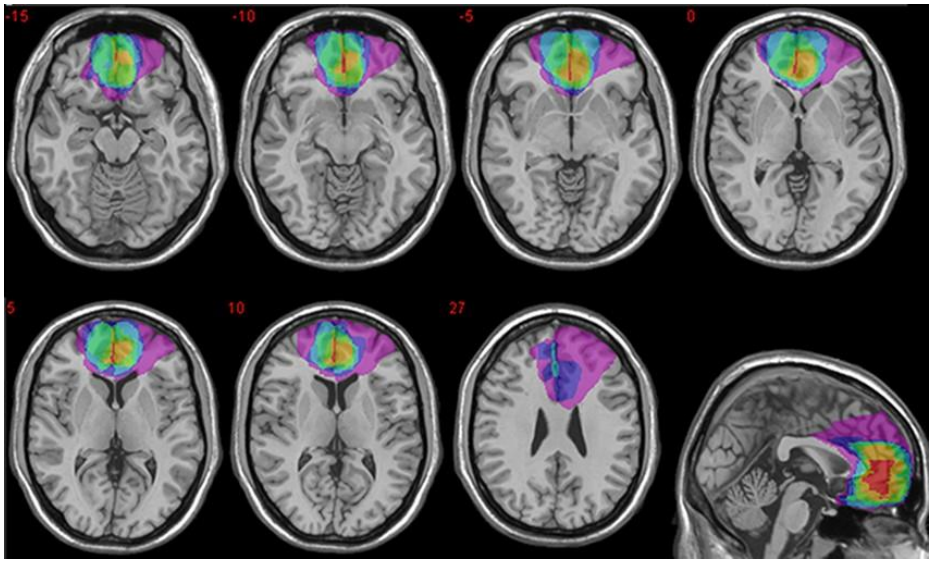

Figure 1. Location and overlap of brain lesions. The panel shows the lesions of the seven patients with vmPFC damage projected on the same seven axial slices and on the mesial view of the standard Montreal Neurological Institute brain. The level of the axial slices is indicated by white horizontal lines on the mesial view of the brain, and by *z*-coordinates. The color bar indicates the number of overlapping lesions. Maximal overlap occurs in BA 11, BA 10 and BA 32 of vmPFC. In axial slices, the left hemisphere is on the left side.

2.3. Neuropsychological assessment

Patients' general cognitive functioning was preserved, as indicated by the scores they obtained in the Raven Standard Matrices (Spinnler & Tognoni, 1987), which were on average within the normal range, and comparable across participant groups (vmPFC patients: M = 23, range: 19-29; control patients: M = 22, range: 11-25; healthy controls: M = 28, range: 13-39; $F_{1,31}$ = 2.98; p = 0.065; Capitani & Laiacona, 1988). vmPFC patients also received a more extensive neuropsychological evaluation, aimed at specifying their cognitive profile further. Table 2 portrays individual vmPFC patients' scores in standardized neuropsychological tests. vmPFC patients attained normal scores in tests assessing attentional skills (Attentional

Matrices; Spinnler & Tognoni, 1987), verbal and spatial short-term memory (Digit Span, Corsi test; Spinnler & Tognoni, 1987), and verbal long-term memory (Prose-passage recall test; Spinnler & Tognoni, 1987). As for executive functioning, both phonemic and semantic fluency were within the normal limits (Spinnler & Tognoni, 1987), but a few cases exhibited impaired cognitive flexibility, as apparent in an increased number of perseverative errors in the Wisconsin Card Sorting Test (Heaton et al., 2000).

Task Procedure

A set of 180 adjectives reflecting psychological traits (90 with a positive connotation and 90 with a negative connotation; *e.g.*, sincere, cynic) was selected from Anderson's (1968) list and translated to Italian. Ninety adjectives were used in the initial rating phase and served as studied items in the following recognition phase, whereas the remaining 90 adjectives served as distractors in the recognition phase. The assignment of trait adjectives to the different rating conditions or to the distractor status (in the recognition phase) was counterbalanced across participants.

In the rating phase, participants were presented with 90 adjectives (half positive and half negative), and were required to make different types of judgment depending on the experimental condition, namely, assess whether the adjective described their current psychological traits (Present-Self condition; 18 items), their anticipated psychological traits in 10 years (Future-Self condition; 18 items), the current psychological traits of Gerry Scotti, a famous Italian showman of approximately the same age of our participants (Present-Other condition; 18 items), and the anticipated psychological traits of Gerry Scotti in 10 years (Future-Other condition; 18 items). We also included a Standard condition (18 items), in which participants judged whether or not the adjective referred to a positive psychological trait, which involves semantic processing but not reflecting on the characteristics of a particular person

(self or other). Each trial started with a fixation cross shown for 500 ms. Then, a trait adjective appeared, along with the question pertaining to the relevant rating condition (*e.g.*, in the Present-self condition: how well does this trait describe YOU NOW?), which was written right above the adjective, and remained on the screen until the end of the trial. Across conditions, participants responded using a Likert scale from 1 (not at all) to 4 (totally), with no time limit for responding. Participants evaluated different adjectives in each rating condition (counterbalanced), and the order of trials pertaining to the different conditions was randomized for each participant.

About 15 minutes after the rating phase, which were filled with unrelated activities (the Raven task and demographic questionnaires), participants underwent an unanticipated recognition memory task (recognition phase), in which the 90 previously rated adjectives were presented again, this time intermixed with 90 new trait adjectives. Each trial started with a fixation cross shown for 500 ms. Then subjects were presented with an adjective and had to state whether they remembered it from the previous session or not (old/new judgment).

Finally, subjects were presented again with the trait adjectives they had previously evaluated with reference to the present and future self, and asked to report, for each trait, how certain they were that they possessed (or not) that trait (for items in the Present-Self condition; 18 items) or that they will possess (or not) that trait (for items in the Future-Self condition; 18 items) (epistemic response; D'Argembeau et al., 2012), and how important it was to them that they possessed (or not) that trait (Present-Self condition) or that they will possess (or not) that trait (Future-Self condition) (emotional response; D'Argembeau et al., 2012). In all cases, participants responded using a Likert scale from 1 (not at all) to 4 (totally).

Results

Rating (encoding) phase

We first investigated whether there were group differences in the time participants needed to evaluate trait adjectives across experimental conditions (Present-Self, Future-Self, Present-Other, Future-Other, Standard), and in the degree to which participants attributed psychological traits to the self (Present-Self and Future-Self conditions), to another person (Present-Other and Future-Other conditions), or felt that a personality trait was positive (Standard condition). An ANOVA on response times (RTs) with Group (vmPFC patients, control patients, healthy controls) and Condition (Present-Self, Future-Self, Present-Other, Future-Other, Standard) as factors revealed no significant effects or interactions ($p > 0.13$ in all cases), meaning that participant groups took a similar time to evaluate trait adjectives at encoding, which did not differ across encoding conditions. Because ratings were in some cases non normally distributed (Komolgorov-Smirnov $d > .20$, $p < 0.01$), the data were analyzed with non-parametric statistics. We found no significant group differences in mean ratings across conditions (Median test $\chi 2 < 4.84$, $p > 0.08$ in all comparisons). We obtain similar findings analysing positive and negative personality traits separately.

Recognition Phase

Table 2 shows mean accuracy (hit rates – false alarm rates) by participant group and rating condition (Present-Self, Future-Self, Present-Other, Future-Other, Standard), and Figure 2 shows the self-reference effect (SRE) relative to the present and the future by participant group. We obtained a similar pattern of results analysing recognition accuracy for positive and negative traits separately, and so, for clarity, we report on the collapsed results.

	Present-Self	*Present-Other*	*Future-Self*	*Future-Other*	*Standard*
vmPFC patients	0.11 (0.09)	0.12 (0.15)	0.06 (0.09)	0.05 (0.16)	0.10 (0.20)
Control patients	0.25 (0.26)	0.03 (0.15)	0.15 (0.24)	0.08 (0.19)	0.18 (0.16)
Healthy controls	0.48 (0.17)	0.27 (0.14)	0.43 (0.17)	0.25 (0.16)	0.43 (0.15)

Table 2. Mean recognition accuracy by participant group and encoding condition. The values in parenthesis are s.d. values.

Standard recognition accuracy. As a preliminary assessment of general recognition memory abilities across participant groups, we conducted a one-way ANOVA on recognition accuracy in the Standard condition with Group (vmPFC patients, control patients, healthy controls) as factor. The ANOVA revealed a significant effect of Group ($F_{2,35}$ = 13.70, p = 0.00004, η^2_p = 0.44). Post hoc comparisons, conducted with the Duncan test, showed that both vmPFC (0.10 vs. 0.43, p = 0.0002) and control patients (0.18 vs. 0.43, p = 0.002) had lower recognition accuracy compared to healthy controls, but there was no significant difference in recognition accuracy between vmPFC patients and control patients (p = 0.32; Table 2).

Recognition accuracy for self and other present and future traits. We next investigated the effect of self-reference and of time on recognition accuracy. We ran a three-way ANOVA on recognition accuracy with Group, Self-reference (Self, Other), and Time (present, future) as factors. The ANOVA revealed a significant effect of Time ($F_{2,35}$ = 5.89, p = 0.02, η^2_p = 0.14), indicating that trait adjectives evaluated with reference to the present were generally recognized better than those evaluated with respect to the future. Moreover, there were a significant effect of Group ($F_{2.35}$ = 16.66, p < 0.00001, η^2_p = 0.49) and a significant effect of Self-reference ($F_{2,35}$ = 16.85, p < 0.0001, η^2_p = 0.83), qualified by a Group x Self-reference interaction ($F_{2,35}$ = 4.73, p = 0.015, η^2_p = 0.21). Post hoc comparisons confirmed that patients'

recognition accuracy in the Other-conditions was significantly (in the case of control patients: 0.059 vs. 0.26; $p = 0.038$) or numerically lower (in the case of vmPFC patients: 0.085 vs. 0.26; $p = 0.06$) than that of healthy controls, while their recognition accuracy in the Self-conditions was significantly lower than the controls' (control patients: 0.20 vs. 0.46; $p = 0.006$; vmPFC patients: 0.085 vs. 0.46; $p = 0.0002$). There were no differences, however, in recognition accuracy between vmPFC patients and control patients in either the Other-conditions ($p = 0.76$) or the Self-conditions ($p = 0.17$). Crucially, whereas healthy controls (0.46 vs. 0.26; $p < 0.0003$) and control patients (0.20 vs. 0.06; $p = 0.008$) evinced higher recognition accuracy when evaluating adjectives with reference to the self than to the other, no such modulation was observed in vmPFC patients (0.085 vs. 0.085; $p = 1$), who attained a similar recognition accuracy in the Self- vs. Other-conditions, thus showing no SRE. There were no other significant effects ($p > 0.1$ in all cases) (see Table 2).

Present and future self-reference effect (SRE). To quantify the SRE (or lack of) directly, we computed an SRE index by subtracting accuracy in the Other-condition from that in the Self-condition, separately for the Present and Future time. An ANOVA performed on the SRE, with Group and Time as factors, showed a significant effect of Group ($F_{2,35} = 4.72$, $p = 0\ .015$, $\eta^2_p = 0.21$), indicating a reduced (virtually absent) SRE in vmPFC patients compared to healthy controls (0.00 vs. 0.20; $p = 0.008$) and control patients (0.00 vs. 0.14; $p = 0.04$), with no difference between the control groups ($p = 0.43$). There were no other significant effects ($p > 0.22$ in all cases). The effect of Group remained significant when we inserted (baseline) recognition accuracy in the Other-conditions (*i.e.*, collapsing across the Present-Other and Future-Other conditions) as a covariate ($F_{2,34} = 8.15$, $p = 0.001$, $\eta^2_p = 0.32$), indicating that the SRE was reduced in vmPFC patients compared to both healthy controls ($p = 0.005$) and control patients ($p = 0.03$), while there was no difference between the control groups ($p = 0.40$). The effect of the covariate was also significant ($F_{1,34} = 5.92$, $p = 0.02$, $\eta^2_p = 0.14$), such that

participants with the lowest performance in the Other-conditions were those that enjoyed the greatest SRE effect (β = -0.27). No other effect was significant (p > 0.27 in all cases).

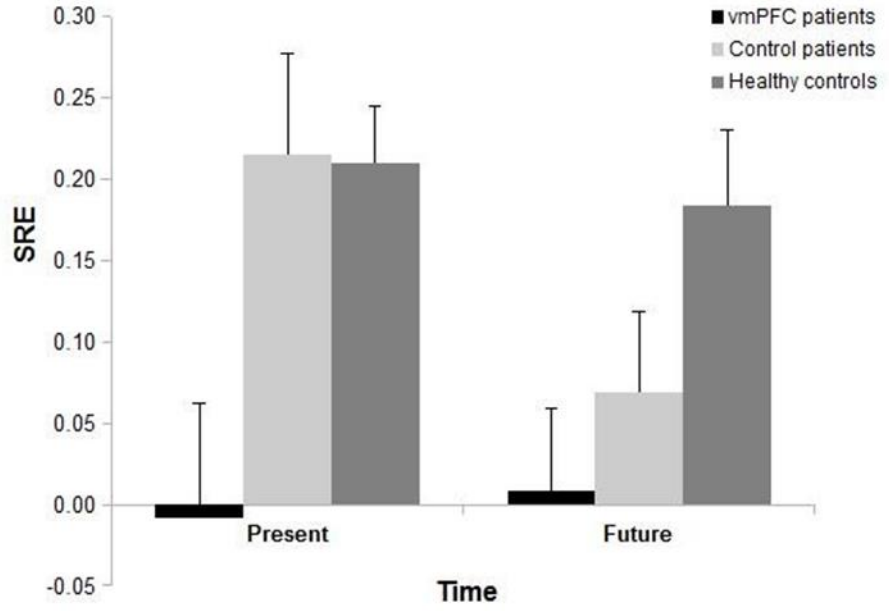

Figure 2. The self-reference effect (SRE) relative to the present and the future in vmPFC patients, and healthy and brain-damaged controls. Bars indicate standard errors of the mean.

Certainty and importance ratings of self traits

To begin investigating possible cognitive factors associated with the lack of SRE in vmPFC patients, we analyzed the certainty and importance ratings they gave to personality traits. An ANOVA on certainty ratings with Group and Time as factors showed a significant effect of Group ($F_{2,35} = 5.22$, $p = 0.01$, $\eta^2_p = 0.22$), indicating that vmPFC patients were less certain to possess (or not) given personality traits compared to both healthy controls (2.64 vs. 3.05; p = 0.01), and control patients (2.64 vs. 3.25; p = 0.001), with no difference between the control groups (p= 0.26). There was also a significant effect of Time ($F_{1,35} = 22.32$, $p = 0.00003$, $\eta^2_p = 0.39$), such that all participants reported they were less certain about the traits they anticipated they might possess (or not) in the future compared to those they thought they had (or not) now (2.90 vs. 3.12). The Group x Time interaction was not significant (p = 0.64). The same ANOVA on importance ratings evinced an effect of Group ($F_{2,35} = 3.77$, $p = 0.03$, $\eta^2_p =$

0.17), indicating that vmPFC patients attributed less importance than healthy controls to possessing (or not) given personality traits (2.73 vs. 3.13; p = 0.03), but their importance ratings were similar to those of control patients (2.73 vs. 2.87; p = 0.12). There was no difference between the control groups (p = 0.43). No other effect in the ANOVA was significant (p > 0.18 in all cases).

	Certainty ratings		**Importance ratings**	
	Present Self	***Future Self***	***Present Self***	***Future Self***
vmPFC patients	2.79 (0.46)	2.48 (0.36)	2.81 (0.43)	2.67 (0.48)
Control patients	3.34 (0.58)	3.15 (0.51)	2.81 (0.41)	2.94 (0.44)
Healthy controls	3.16 (0.27)	2.95 (0.38)	3.16 (0.34)	3.11 (0.41)

Table 3. Mean certainty and importance attributed to self traits by participant group and time condition. The values in parenthesis are s.d. values.

Relation between certainty and importance ratings and recognition accuracy: exploratory analyses

We investigated whether recognition accuracy in the Self-conditions was related to certainty and importance ratings to trait adjectives. We ran a linear mixed effect model on single trait adjective data (N = 1368) with recognition accuracy as the dependent variable (1 = hit and 0 = miss); Certainty ratings, Group (vmPFC patients, control patients and healthy controls) and Time (Present and Future) as fixed effects; and Subject as a random effect. There were a significant effect of Certainty ratings ($\chi2 = 11.4$, p = 0.001), such that trait adjectives associated with high certainty ratings were more likely to be correctly recognized, and a significant effect of Group ($\chi2 = 8.72$, p = 0.01), such that recognition accuracy in the Self-conditions was lower in vmPFC patients compared to healthy (p < 0.0001) and brain-damaged controls (p = 0.02). No other effect or interaction in the model was significant (p > 0.40 in all cases). The same

model considering Importance ratings, Group and Time as fixed effects and Subject as a random effect yielded a significant effect of Importance ratings ($\chi2 = 10.88$, $p = 0.001$) and a significant effect of Group ($\chi2 = 9.43$, $p = 0.01$), qualified by an Importance rating × Group interaction ($\chi2 = 7.78$, $p = 0.02$). The interaction indicated that importance ratings predicted recognition accuracy significantly in healthy controls ($\chi2 = 10.23$, $p = 0.02$) and in control patients ($\chi2 = 11.19$, $p = 0.01$), but not in vmPFC patients ($p = 0.7$).

These findings indicate that, in healthy controls and control patients, recognition accuracy for self-referenced items was related to the certainty and importance participants associated with possessing (or not) given personality traits. Certainty ratings predicted recognition accuracy in vmPFC patients also, whereas importance ratings appeared untied to recognition accuracy in this group. When we ran again the ANOVA on the SRE with Group and Time as factors, this time including certainty and importance ratings (collapsed across the Present-Self and Future-Self conditions) as covariates, the original effect of Group was no longer significant ($p = 0.16$), as were all other effects in the ANOVA ($p > 0.28$ in all cases), which suggests that the reduced SRE observed in vmPFC patients may be related, at least in part, to their reduced epistemic and emotional responses to adjective traits

Discussion

This study investigated the recognition memory advantage for items (trait adjectives) referenced to the self vs someone else (SRE) and relative to the present vs future time in vmPFC patients, control patients and healthy controls. First of all, we confirmed the presence of an SRE in healthy participants, which was abolished in vmPFC patients, in line with the findings obtained by Philippi et al. (2012a). Moreover, we showed that healthy controls and control patients also exhibit a future SRE, that is, better recognition accuracy in association with traits evaluated against their view of themselves (vs another individual) in the future. The future SRE

was, again, absent in vmPFC patients, as was the SRE for the present self. Contrary to our predictions, the present and future SREs had comparable magnitude. This was because, across groups, evaluating trait items from a future (as opposed to present) time perspective resulted in lower recognition accuracy, but this held for both self-referenced items and other-referenced items alike, and therefore did not affect the SRE (difference between Self- and Other-conditions).

Before discussing each of these three main findings in turn, we wish to emphasize that the lack of SRE observed in vmPFC patients is not a common consequence of brain damage, for example reflective of a weakened sense of self following illness and perceived vulnerability (Ciaramelli et al., 2019), as it was not observed in control patients (see also Philippi et al., 2012a). It is also unlikely to depend on generally poor recognition memory abilities or comprehension of task instructions on the vmPFC patients' part. Indeed, although vmPFC patients' recognition accuracy was worse than that of healthy controls across conditions, so was that of control patients, and yet they evinced a normal SRE. Moreover, it does not seem that vmPFC patients failed at distinguishing different task conditions (*e.g.* Self vs Other). Indeed, they showed better memory for items encoded with reference to the present vs future time perspective, as did the other groups, suggesting they were normally responsive to the encoding demands.

Our primary finding that the SRE is abolished in vmPFC patients confirms previous evidence that medial prefrontal regions (Philippi et al., 2012a), including vmPFC (this study), are crucially linked to the representation of the self. In addition, our study points to the persistence of the SRE when evaluating the future self in healthy controls and control patients and of its absence in vmPFC patients. The evidence of a future SRE suggests that, although fMRI evidence shows lower medial prefrontal activity for the future than for the present self (Ersner-Hershfield et al., 2009; D'Argembeau et al., 2010a), our future self is not an 'other':

what is encoded with reference to the self, whether past or future, is more frequently remembered than what is encoded with respect to others. The absence of a future SRE in vmPFC patients, therefore, reinforces the view of vmPFC as implicated in self-related processing (Northoff & Bermpohl, 2004; Northoff et al., 2006; Moran et al., 2006; Schmitz & Johnson, 2007).

We found that recognition accuracy for self-referenced traits is predicted by the certainty with which individuals think they possess or will possess those traits (or not) and by the importance they attribute to possessing (or not) those traits. This finding suggests that trait items that are more relevant to our self-schema, because they contribute to define ourselves (the way we definitely are and are not and the way we definitely think we will be or not be) and the value we attach to our (present and future) self-views, enjoy a privileged encoding in memory. Importantly, vmPFC patients were less confident about possessing or not possessing certain personality traits compared to healthy and brain-damaged controls, consistent with previous findings of activity in BA 10 in association with the expression of certainty in self-views (D'Argembeau et al., 2012). A possibility, therefore, is that vmPFC patients did not show an SRE because they are less certain about the traits they do or do not possess, or those they will or will not possess, due to a weakened self-schema or schema instantiation (Gilboa et al., 2006; Ghosh et al., 2014). Consistent with this proposal, vmPFC patients are particularly impaired at imagining self- vs other-related future events, as if they failed to activate schematic self-knowledge that drives the collection of individual details of events (D'Argembeau & Mathy, 2011; Verfaellie et al., 2019). Moreover, vmPFC is deemed to generate coherent confidence signals based on the evaluation of personal information against the self-schema (Hebscher & Gilboa, 2016). Indeed, vmPFC damage is often associated with confabulation, the production of false memories for (unhappened) events even blatantly inconsistent with the self-schema (Moscovitch, 1995; Gilboa et al., 2006), which are typically held with abnormal

conviction (Gilboa et al., 2006; Ciaramelli & Ghetti, 2007). vmPFC patients also attributed less importance to having given traits than controls, and, unlike the control groups' importance ratings, their importance ratings were not related to recognition accuracy for self-related items. However, vmPFC patients' importance ratings did not differ from those of control patients, who showed an SRE, and therefore are less likely to underlie vmPFC patients' lack of SRE. Although our findings indicate that recognition accuracy is related to certainty and importance ratings for self-related trait items, future studies involving more patients are needed to confirm whether the SRE reduction observed in vmPFC patients is critically linked to their reduced certainty (and importance) responses.

An interesting finding of our study is the mnemonic consequence of adopting a future time perspective. We observed, across groups, a decline in recognition accuracy when participants encoded (both self-referenced and other-referenced) information with respect to a future compared to a present time perspective. Why is information belonging to the future remembered less than that belonging to the present? All participant groups reported they were less certain about their traits in the future than in the present. This finding aligns with the 'failure of imagination theory', according to which people find it difficult to imagine how their future self will be (Frederick et al., 2009; Hershfield & Bartels, 2018), which appears to extend to others' future. We propose, therefore, that a less vivid representation of the future (vs present) led to relatively shallower trait encoding in both self- and other-referenced conditions, resulting in lower recognition accuracy. The fact that a time (future vs present)-dependent modulation of recognition accuracy was observed in vmPFC patients as well controls highlights areas of spared time processing in vmPFC patients. This finding, indeed, indicates that even though vmPFC patients are impaired in imagining specific future events (Bertossi et al., 2016a,b, 2017; Verfaellie et al., 2019), in self-projecting into future time periods (Ciaramelli et al., 2021; see also Sellitto et al., 2010) and also in representing future self-

knowledge (this study), they are capable of distinguishing between different time moments, suggesting that vmPFC integrity is not necessary to conceive time in abstract terms. D'Argembeau et al. (2010a) found that the inferior parietal cortex was more active when participants reflected on their past and future compared to current selves, a pattern of activity opposite to that displayed by vmPFC. One possibility, therefore, is that vmPFC supports self-related processing, but it is the inferior parietal cortex that mediates the representation of time and temporal distances (Bueti & Walsh, 2009; Nyberg et al., 2010).

To conclude, we have confirmed that self-related information is prioritized in memory and found that this mnemonic advantage extends to information that is relevant to the future self. The present and future SREs are crucially linked to vmPFC integrity, as we found them abolished in vmPFC patients, and this was not a common consequence of brain damage or poor recognition memory. Rather, vmPFC patients showed reduced certainty for self-relevant information (their own traits) compared to the control groups, which we interpret as a consequence of a weakened self-schema or schema instantiation. Interestingly, all participants evinced lower recognition accuracy for future-referenced compared to present-referenced items, suggesting that the present, in addition to the self, is prioritized in memory, which was linked, again, to increased certainty in association with present- vs future-referenced information. vmPFC patients, too, showed this present-related memory advantage, meaning they can represent different time moments, at least in these abstract, impersonal terms.

Chapter 2. Not sure who I am: instability of self-related judgments after vmPFC damage

In the first chapter, I explored the status of self-related information in memory in patients with a damage to vmPFC, concluding that their self-schema presents as degraded. However, if you know these patients, you might be not completely convinced yet. It is well known that vmPFC patients suffer from memory impairments, which could get in the way when asking them to encode and remember new information. This is the reason why in this chapter we try to understand whether the integrity of the self-schema is really compromised by a vmPFC lesion, or whether the loss of the *special* status of self-related items in memory found in study 1 could be attributable to patients' memory deficits. Moreover, this time we push our question even further, asking patients to judge much more "trivial" aspect of the self, such as habits and likings, something that does not require strong metacognitive evaluations.

Study 2. Who am I really? The ephemerality of the self-schema following vmPFC damage*

*This work has been published in Stendardi, D., Giordani, L. G., Gambino, S., Kaplan, R., & Ciaramelli, E. (2023). *Neuropsychologia*, *188*, 108651.

Introduction

Personal preferences, idiosyncrasies, and little quirks of personality are what make us uniquely ourselves. We know that we take our coffee black, are always late, like diving but hate when others take pictures of us. These instances of personal semantic knowledge are part of the self-schema, an articulated set of beliefs about oneself, generally deriving from the repeated categorization and subsequent evaluation of one's behaviour, which defines our identity and drives our behaviour (Markus, 1977). How do we know who we are?

Personal semantic knowledge is a memory system at the border between episodic memory, our ability to recollect personal experiences within their unique spatio-temporal context (*e.g.*, 'the first time I tried black coffee I liked it'), and semantic memory, our (culturally shared) knowledge of facts and concepts by now detached from the context of acquisition (*e.g.*, 'black coffee contains more caffeine'), as it is personal but relatively devoid of context (Renoult et al., 2012, 2016), although different domains of personal semantic memories differ in their relation to semantic and episodic memory (see also Grilli & Verfaellie, 2014). Self-knowledge is the most extensively studied instance of personal semantic memory; it has been classically operationalized as knowledge of one's own personality traits or, less frequently, personal preferences (*e.g.*, I am shy; I prefer dogs to cats; Renoult et al., 2012; Craik et al., 1999; Kelley et al., 2002; Martinelli et al., 2013; Wank et al., 2022). Personal semantic memory, however, also contains information that is less abstract and more directly related to (or inferable from) events (Grilli & Verfaellie, 2014, 2015), such as autobiographical facts (*e.g.*, I have a female dog) and repeated events (*e.g.*, I smoke everyday while walking my dog) that all contribute to shape the self-schema (Markus, 1977, 1983).

Past work has demonstrated that judging the self-relevance of a personality trait does not influence the time required to subsequently recollect an event in which one displayed that personality trait (episodic memory), or to define that trait (semantic memory; Klein & Loftus,

1993; Klein & Lax, 2010), suggesting some degree of functional independence between self-knowledge and semantic and episodic memory. Moreover, there is neuropsychological evidence that preserved self (trait) knowledge can withstand impairments in semantic and episodic memory (Klein & Lax, 2010; Renoult et al., 2012). Tulving (1993) first reported that the severely amnesic patient KC could describe his personality accurately and reliably. KC's self-reported personality traits were consistent between testing sessions (78% agreement) and in line with those provided by his mother (see also Klein et al., 1996; Klein et al., 2002a,c). Self-trait knowledge can remain intact even in the presence of profound semantic memory deficits (Westmacott et al., 2001; Klein et al., 2002c, Klein et al., 2003; Picard et al., 2013; Duval et al., 2013), such as in late-stage Alzheimer's disease (Klein et al., 2003) or semantic dementia (Duval et al., 2013). However, other results have challenged these findings (see Charlesworth et al., 2016; Tanguay et al., 2018; Wank et al., 2022), for example questioning the accuracy/completeness of self-knowledge in patients with episodic and semantic memory deficits (Klein et al., 2003). More in general, recent findings call for a qualification of the relation between the personal semantic and episodic memory systems reflecting the heterogeneity of the former (Renoult et al., 2012; Grilli & Verfaellie, 2014). For example, autobiographical facts that are 'experience-near' and not completely devoid of spatio-temporal information (*e.g.*, I was involved in sports in high school) have been found to depend on the integrity of the medial temporal lobe (Grilli & Verfaellie, 2015, 2016; see also Wank et al., 2022; Sawczak et al., 2022).

What are the neural bases of personal semantic memory? There is converging evidence that the medial prefrontal cortex (mPFC) is associated with self-knowledge. In functional neuroimaging (fMRI) studies, mPFC is more active for self (trait) judgments rather than general semantic evaluations (Johnson et al., 2002; Schmitz et al., 2004; D'Argembeau et al., 2010a,b). Zysset et al. (2002) reported a functional dissociation between mPFC and the inferior

precuneus, which proved more engaged by self-trait judgments and autobiographical memory retrieval, respectively (see also Sajonz et al., 2010). Also, mPFC is at the basis of the self-reference effect (SRE; Rogers et al., 1977; Symon & Johnson, 1997). This region is indeed more active when participants judge the self-relevance of personality traits (*e.g.*, are you an extrovert?) compared to their phonemic or semantic properties, or their descriptiveness of another individual (Kelley et al., 2002; Moran et al., 2006; D'Argembeau et al., 2007; Sui & Humphreys, 2015). Moreover, activity in mPFC predicts the level of enhanced memory for personality traits encoded with respect to the self than to another individual (Moran et al., 2006). Importantly, Philippi et al. (2012a) and Stendardi et al. (2021) reported a drastic reduction of the SRE in patients with lesion in mPFC, especially in its ventral sector (vmPFC; Stendardi et al., 2021). In these studies, vmPFC patients did not show a memory advantage for items encoded with respect to the self compared to another individual, suggesting that vmPFC is necessary to support the self-schema or to impart a mnemonic advantage to items relevant to the self-schema.

Although previous studies revealing a virtual absence of the SRE in vmPFC patients point to a crucial role of vmPFC in self-knowledge (Philippi et al., 2012a; Stendardi et al., 2021), in those studies the status of self-knowledge is inferred from performance in an episodic (anterograde) memory task. Because vmPFC patients may have anterograde memory impairments that go beyond their self-knowledge deficits (Della Sala et al., 1993; Kopelman et al., 1999; Ciaramelli et al., 2006; Ciaramelli et al., 2009; Ciaramelli & di Pellegrino, 2011; Ciaramelli et al., 2019; Bertossi et al., 2016b, 2017a; De Luca et al., 2018b), it is not clear the degree to which the reduction of the SRE following vmPFC damage is due to degraded self-knowledge or impaired self-referential encoding in these studies. Ideally, tests with low demands on anterograde episodic memory would be better suited to capture the status of the self-schema. Two single case studies adopted this approach. Marquine et al. (2016) required

J.S., a patient with a bilateral (mostly right-lateralized) mPFC damage, to provide (the same) self-related judgments (*e.g.*, "are you an introvert?") on two different testing sessions, under the assumption that a preserved self-schema should support highly consistent self-related judgments across sessions. J.S. was highly inconsistent in self-related judgments across sessions, despite a normal performance in other-related judgments, suggesting impaired self-knowledge. Philippi et al. (2012b) reported the case of a patient with bilateral (mostly right-lateralized) mPFC lesion, R., who was instead highly consistent between sessions. However, his judgments did not match his mother and sister's judgements, suggesting again an impairment of trait self-knowledge, though of a different kind. Although case studies are important to illuminate brain-behaviour relations, they have inherent limitations, and therefore it would be important to confirm these findings in a group study of patients with focal lesions to vmPFC.

Note, also, that most previous studies have focused on the role of mPFC in trait knowledge, and therefore it is not clear whether mPFC would also support different instances of personal semantic memory. The mPFC is consistently engaged by self-referential processing (Northoff et al., 2006; Jenkins & Mitchell, 2011). Renoult al. (2012), indeed, pointed out that mPFC regions are generally more engaged by self-knowledge, autobiographical facts, and repeated events than by general semantic knowledge. Paulus & Frank (2003) found that vmPFC activity was crucially linked to personal preferences, and Mitchell et al., (2011) showed that vmPFC was engaged while individuals predicted the probability with which they would enjoy a series of events.

The aim of this work is twofold. First, we aimed to confirm the role of vmPFC in personal semantic memory probing the domain or personal preferences and activities, instead of the most extensively studied self-trait knowledge (Philippi et al., 2012b; Marquine et al., 2016; Stendardi et al., 2021). Moreover, we aimed to use a test that does not make heavy

demands on anterograde memory, as was the case in previous studies (Philippi et al., 2012a; Stendardi et al., 2021). To this aim, we asked a sample of patients with focal lesions to the vmPFC (vmPFC patients), control patients with lesions outside vmPFC, and healthy controls to judge the likelihood with which they (or a close friend) engaged in a series of activities (*e.g.*, "going to work on foot", "eating a croissant"; "sleep more than 7 hours a night"). Preferences and activities tap a more concrete aspect of self-knowledge than trait-knowledge, and with greater commonality with other domains of personal semantic memory, such as autobiographical facts and repeated events. Participants rated the same stimuli on two separate occasions, a week apart. We predicted that vmPFC patients would show inconsistent self-related (but not necessarily other-related) judgments, indicative of an impaired self-schema. In addition, we investigated the confidence associated with self-related judgments. The vmPFC is thought to generate confidence signals resulting from the match between incoming information and the self-schema (Hebscher & Gilboa, 2016). If vmPFC patients have an impaired self-schema, as we predict, they should show a generally reduced confidence in self-related (but not necessarily other-related) judgments. Gathering confidence ratings also allowed us to explore whether the vmPFC patients were aware of the expected impairment in self-related knowledge.

Methods

Participants

Twenty-nine healthy participants (healthy controls; 9 females; mean age = 58.34 years, sd = 5.7, range = 47-74; mean education = 13.34 years, sd = 4.1, range = 5-22), 6 patients with lesions to vmPFC (vmPFC patients; 1 female; mean age = 55.7 years, sd = 5.04, range = 48-61; mean education = 11.33 years, sd = 2.6, range = 8-13) and 8 patients with lesions outside

vmPFC (control patients; 3 females; mean age = 52.6, sd = 18.2 years, range 28-78; mean education = 12.6 years, sd = 6.9, range = 5-22) participated in the study (see Table 1 for patients' demographic and clinical data). Patients were recruited at the Centre for Studies and Research in Cognitive Neuroscience, Cesena, based on their lesion site, as documented by MRI or computerized tomography (CT) scans. vmPFC patients' lesions resulted, in all cases, from the rupture of an aneurysm of the anterior communicating artery (ACoA). They were bilateral in all cases, although predominantly right lateralized for one patient.

The other 8 (control) patients had lesions caused by ischemic or haemorrhagic stroke, traumatic brain injury or brain tumour and were unilateral in seven cases (four right-lateralized, three left-lateralized), and bilateral in one case. Control patients' lesion sites involved the fronto-temporal area (three cases), the occipital cortex (one case), the occipito-parietal area (one case), the occipito-temporal cortex (one case), the temporo-parietal cortex (one case), and the thalamus (one case). There was no significant difference in mean lesion volume between vmPFC and control patients (53.7 cc vs 28.2 cc., $p = 0.052$). All patients were in the stable phase of recovery (at least 3 months post-morbid). vmPFC patients' general cognitive functioning was generally preserved, as indicated by scores within the normal range at the Raven Standard Matrices, phonemic and semantic fluency, the prose passage recall, and the digit span test, and their performance was comparable to the controls' (all ps > 0.09; see Table 1). vmPFC patients did not show clinical evidence of confabulation.

Healthy participants were matched to patients on age ($F_{2,40} = 1.28$, $p = 0.29$), education ($F_{2,40} = 0.5$, $p = 0.61$), and females/males ratio ($\chi^2 = 0.79$, $p = 0.67$). Participants gave written informed consent to participate in the experiment, which was performed in agreement with the Declaration of Helsinki, and approved by the Bioethical Committee of the University of Bologna and the Ethical Committee of Area Vasta (CEIIAV) of Emilia Romagna.

vmPFC patients							*Control patients*							
	p. 1	p. 2	p. 3	p. 4	p. 5	p. 6	cp.1	cp. 2	cp. 3	cp. 4	cp. 5	cp. 6	cp.7	cp.8
Sex	M	M	F	M	M	M	F	M	M	F	M	F	M	M
Age (years)	51	58	61	48	59	57	28	73	41	52	78	52	64	33
Education (years)	13	8	13	13	8	13	13	5	9	5	21	22	8	18
Raven Standard Matrices (cut-off = 15)	33	33.5	35.75	32.5	33.5	27	31.5	26.75	37.5	10.25	22	29.5	24	31.5
Phonemic Fluency (cut-off = 17)	34	27	25	21	36	32	-	15	36	20	31	56	41	48
Semantic Fluency (cut-off = 25)	31	37	42	40	61	35	36.5	25	54	28	33	54	57	75
Short term memory - Digit span (cut-off = 3.75)	6.75	5	5.75	6.5	5	5.75	5.44	4.5	6	3	4.5	5.25	5	8.25
Long-term memory - Prose passage recall (cut-off = 4.75)	2.5*	5	19	13	8.5	13.5	-	11.9	18	7.5	10.5	7	19.5	6.9
Chronicity (months)	14	198	25	73	118	91	10	5	3	15	84	72	4	84
Lesion size (cc)	31	55	74	40	52	69	16	16	33	64	6	14	7	69

Table 1. The table reports, for each vmPFC patient (p) and control patient (cp), scores corrected for age, education and sex according to normative samples (Spinnler & Tognoni, 1978). An impaired performance (percentile score < 5) is signalled by an *.

Lesion analysis

Patients' individual lesions derived from the most recent MRI or CT scans were manually drawn by a trained neuroscientist (not involved in the study) directly on each slice of the normalized T1-weighted template MRI scan from the Montreal Neurological Institute provided with the MRIcro software (Rorden and Brett, 2000). The standard template provides various anatomical landmarks to help experts plot the size and localization of the lesion using structural features such as sulci and gyri as guides. This manual procedure combines segmentation (identification of lesion boundaries) and registration (to a standard template) into a single step, with no additional transformation required (Kimberg et al., 2007). Manual

segmentation/registration procedures have the limit to rely greatly on anatomical expertise, and to be subjective in nature. On the other hand, they circumvent problems frequently encountered by automated normalization procedures, such as (1) warping scans from individuals with brain injury, which may be affected by structural distortions related to the lesion and not easily compensated for (*e.g.*, ventricular enlargement, large regions of atypical voxel intensity values, artifacts induced by the presence of metallic clips), and combining subjects scanned with different imaging modalities (*e.g.*, MRI vs. CT; see also Bertossi et al., 2016a; Kimberg et al., 2007).

The MRIcro software was used to estimate lesion volumes (in cc) and generate lesion overlap images. Figure 1 shows the extent and overlap of brain lesions in vmPFC patients. The Brodmann areas (BAs) mainly affected were BA 11, BA 10, BA 32, BA 25, and BA 24. The maximal lesion overlap occurred in BA 11 (M = 20.6 cc, s.d.= 9.01), BA 10 (M = 11.03 cc, s.d. = 7.26) and BA 32 (M = 8.29 cc, s.d. = 5.39).

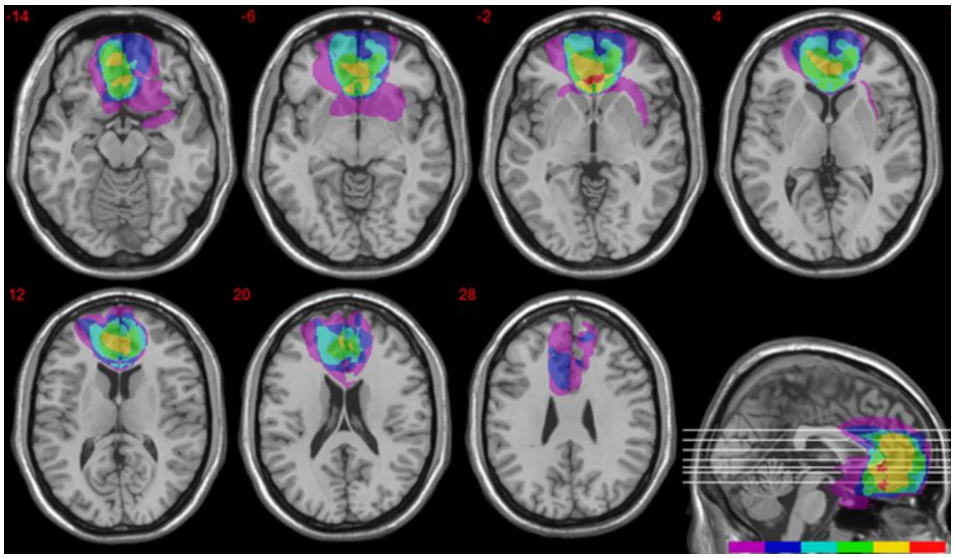

Figure 1. Extent and overlap of vmPFC patients' brain lesions. Lesions are projected on the same seven axial slices of the standard Montreal Neurological Institute brain. The white horizontal lines on the sagittal view are the positions of the axial slices. Numbers above the axial views represent the z-coordinates of each slice. The color bar indicates the number of overlapping lesions, from 1 (purple) to 6 (red). Maximal overlap occurs in BA 11, 10 and 32. The left hemisphere is on the left side.

Materials and procedure

The experimental procedure was articulated in two testing sessions. In the first testing session, participants were first asked to select one of their friends, someone they felt they knew very well, but with whom they had never lived. We required participants to select a friend they had never lived with to minimize the possibility that they engaged (or had engaged) in a series of everyday activities together, and therefore participants could answer about the other merely reiterating the answers about themselves. Participants were then administered a task requiring to answer questions about themselves and the friend they had selected. During the task, a list of 100 activities (*e.g.*, "read a novel", "play sudoku", "walk to work"; adapted from Kaplan & Friston, 2019), were presented, one at time, on the computer screen. In the Self condition, for each activity, participants had to rate on a Likert scale how likely they were to engage in that activity from 1 (= not likely at all) to 9 (= extremely likely). They then answered the same question about their friend (Other condition), rating how likely the friend was to engage in each of the same series of activities.

After each (self-related and other-related) judgment, participant also rated their confidence associated with the judgment on a Likert scale from 1 (= not sure at all) to 5 (= absolutely sure). During the second testing session, which was run about 1 week apart, participants were administered the same task, with the exception that confidence ratings were not collected.

Results

Judgment consistency across sessions (Δ)

For each participant, we computed a score change (Δ) as the difference between the ratings given to each activity in the first session and the second session (in absolute value), and then averaged it across activities, separately for the Self and Other conditions (see Figure 2). High Δ values represent low rating consistency between sessions. We then ran a mixed repeated

measure ANOVA, with Δ as the dependent variable, and Group (Healthy controls, vmPFC patients, Control Patients) and Condition (Self, Other) as predictors. Both the main effects of Group ($F_{2,40} = 24.6$, $p < 0.0000001$, $\eta_p^2 = 0.55$) and Condition ($F_{1,40} = 8.8$, $p = 0.005$, $\eta_p^2 = 0.18$) were significant. There was also a significant Group x Condition interaction ($F_{2,40} = 9.1$, $p = 0.0006$, $\eta_p^2 = 0.31$). Post-hoc Bonferroni tests revealed that in the Self condition vmPFC patients' Δ was higher than that of both healthy (vmPFC patients: 1.50 vs. Healthy controls: 0.64, $p < 0.00001$) and brain-damaged controls (vmPFC patients: 1.50 vs Control patients: 0.94, $p < 0.01$), with no difference between the control groups (Control patients: 0.94 vs Healthy controls: 0.64, $p = 0.08$). In the Other condition, vmPFC displayed a higher Δ score compared to healthy controls (vmPFC patients: 1.33 vs Healthy controls: 0.86, $p = 0.001$) but comparable to that of control patients (vmPFC patients: 1.33 vs Control patients: 1.28, $p = 1$). Crucially, whereas both control groups exhibited lower Δ scores in the Self condition vs. the Other condition (Healthy controls: 0.64 vs 0.86, $p < 0.001$; Control patients: 0.94 vs 1.28, $p < 0.01$), meaning that their self-related judgments were more robust (stable) than their other-related judgments, this self-advantage was not present in vmPFC patients (1.50 vs 1.33, $p = 1$), who showed a numerically higher Δ in Self compared to the Other condition, suggestive of more stable other- than self-related judgments.

Control analysis. To verify that the results were not driven by vmPFC patients p1, who had very low episodic memory scores (see Table 1), we ran again the ANOVA excluding p1's data. We confirmed our findings. The effect of Group ($F_{2,39} = 22.8$, $p < 0.000001$, $\eta p^2 = 0.54$), Condition ($F_{1,39} = 11.2$, $p < 0.01$, $\eta p^2 = 0.22$), and the Group x Condition interaction ($F_{2,39} = 6$, $p < 0.01$, $\eta p^2 = 0.24$) were significant: in the Self condition, vmPFC patients' Δ was higher than that of both control groups (vmPFC patients: 1.49 vs. Healthy controls: 0.64, $p < 0.00001$; vmPFC patients: 1.49 vs Control patients: 0.94, $p < 0.01$), with no difference between the control groups (Control patients: 0.94 vs Healthy controls: 0.64, $p = 0.09$). In the Other

condition, vmPFC patients showed Δ scores higher than healthy controls' (vmPFC patients: 1.39 vs Healthy controls: 0.86, p = 0.001) but comparable to control patients' (vmPFC Patients: 1.39 vs Control Patients: 1.28, p = 1). Both control groups had lower Δ scores in the Self compared to the Other condition (Healthy controls: 0.64 vs 0.86, p < 0.0001; Control Patients: 0.94 vs 1.28, p < 0.01), but this self-advantage in judgment stability was absent in vmPFC patients (1.49 vs 1.39, p = 1).

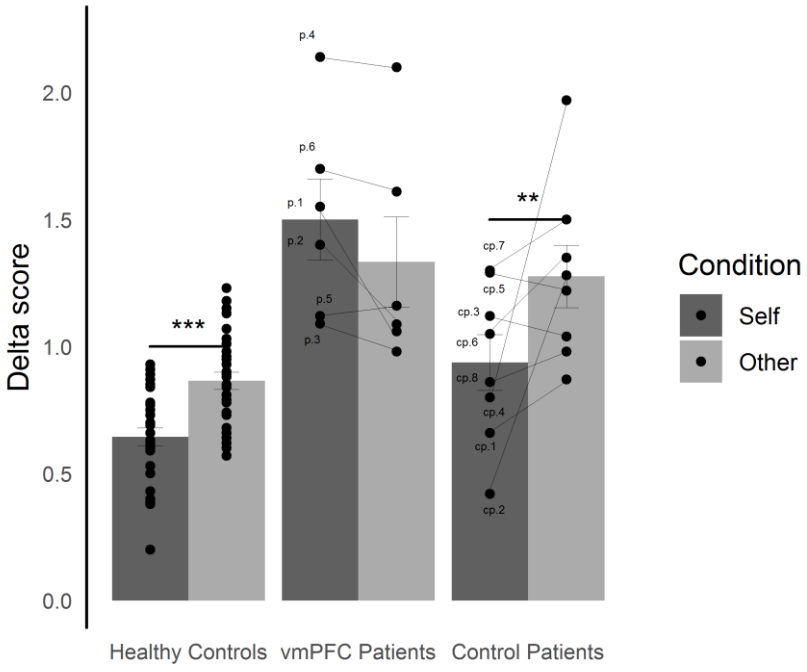

Figure 2. Mean score change between testing sessions (Δ score) by participant group and experimental condition. Bars represent standard errors. Labels denote individual vmPFC patients (p) and control patient (cp). ** = p < 0.01; *** = p < 0.001.

Intra-Class Correlation (ICC)

As an additional measure of rating consistency, we calculated the two-way mixed effect (absolute agreement) ICC for each group, separately for self- and other-related judgments (see Table 2; Koo & Li, 2016). According to the classification by Koo & Li (2016), ICC values lower than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 of moderate reliability, values between 0.75 and 0.9 of good reliability, and values higher than 0.90 of excellent reliability. According to this classification, vmPFC patients exhibited poor reliability

in the Self condition (ICC = 0.48; 95% CI [0.29; 0.77]) and higher (moderate) reliability in the Other condition (ICC = 0.52; 95% CI [0.32; 0.80]), whereas control patients showed good reliability in the Self condition (ICC = 0.75; 95% CI [0.57; 0.92]) and lower (moderate) reliability in the Other condition (ICC = 0.63; 95% CI [0.43; 0.86]). Healthy participants exhibited good reliability across conditions (ICC Self = 0.87; 95% CI [0.76 0.96], ICC Other = 0.79; 95% CI [0.63; 0.93]). These findings confirm that whereas the control groups have more reliable self-related than other-related judgments, vmPFC patients exhibit the opposite tendency.

Confidence

A repeated measure ANOVA, with confidence ratings as the dependent variable, and Group (Healthy controls, vmPFC patients, Control Patients) and Condition (Self, Other) as predictors revealed a significant main effect of Group ($F_{2,40} = 3.4$, $p = 0.04$, $\eta_p^2 = 0.15$) and a significant effect of Condition ($F_{1,40} = 47.1$, $p < 0.000001$, $\eta_p^2 = 0.54$), with no interaction ($F_{2,40} = 2.4$, $p = 0.1$). Post-hoc Bonferroni tests indicated that vmPFC were generally less confident in their judgments than healthy participants ($p = 0.049$), and, though only numerically, control patients ($p = 0.09$), with no difference between the control groups ($p = 1$). All groups showed more confident judgments in the Self compared to the Other condition ($p < 0.000001$; see Table 2).

	Condition	
Group	***Self***	***Other***
Healthy controls	4.69 (0.3)	4.14 (0.5)
vmPFC patients	4.07 (0.8)	3.85 (0.7)
Control patients	4.79 (0.3)	4.11 (0.4)

Table 2. Mean confidence ratings (and SD) by participant group and condition.

Relation between Δ and confidence

We investigated the relation between confidence ratings (in the first testing session) and score changes (Δ) from the first to the second session, under the assumption that more confident judgments would tend to remain stable from the first to the second session. To this aim, we ran a full factorial linear mixed effect model on Δ with repeated measures (here Δ represents the score change for each trial, leading to 100 data points per participant) with Confidence, Group and Condition as fixed effects, and Subject as a random effect. The model allowed estimating both a random intercept, and random slopes for the Confidence and Condition predictors, as specified by the lmer formula in R (Bates et al., 2014):

$$\Delta \sim \text{Group} * \text{Confidence} * \text{Condition} + (1 + \text{Confidence} + \text{Condition} \mid \text{Subject})$$

There were significant main effects of Group ($\chi^2 = 34.4$, $p < 0.001$) and Confidence ($\chi^2 = 130.2$, $p < 0.0001$), a significant Group x Confidence interaction ($\chi^2 = 7.6$, $p = 0.03$), a significant Condition x Confidence interaction ($\chi^2 = 3.9$, $p = 0.048$), and a significant Group x Confidence x Condition interaction ($\chi^2 = 7.8$, $p = 0.02$). The estimates of the regression coefficient β (and the 95% CI) for the variable Confidence are displayed in Table 3. As shown in Table 3, all β estimates were below 0, indicating a negative relation between the confidence associated with an answer and the score change for that answer in the second session. As expected, the more confident healthy and brain-damaged controls were in a judgment, the less that judgment changed in the second session (see Figure 3). This pattern of performance was also apparent in vmPFC patients, but only in the Other condition (β = -0.18, CI [-0.33, -0.03]). By contrast, in the Self condition there was no evidence for a significant relation between confidence and score change between sessions in vmPFC patients (β = -0.07, CI [-0.23, 0.08]; see Figure 3),

indicating that the change in self-related judgments across sessions did not depend on the level of confidence with which the judgment was endorsed in the first session.

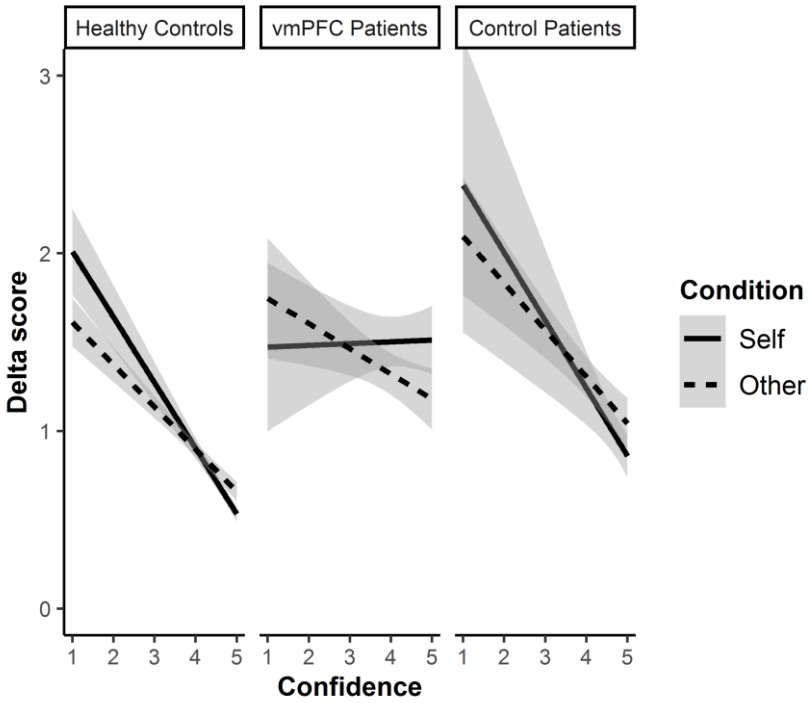

Figure 3. Relation between Δ (score change) and confidence ratings for each group in the two conditions (Self, Other).

Group	***Condition***	***Estimate (β)***	***SE***	***CI Lower***	***CI Upper***
Healthy controls	Self	-0.41	0.05	-0.51	-0.32
	Other	-0.29	0.03	-0.36	-0.22
vmPFC patients	Self	-0.07	0.08	-0.23	0.08
	Other	-0.18	0.08	-0.33	-0.03
Control patients	Self	-0.49	0.10	-0.69	-0.28
	Other	-0.29	0.06	-0.40	-0.17

Table 3. β coefficient and 95% confidence intervals for the variable Confidence, for each group and each condition. SE = Standard error; CI = 95% Confidence Interval.

Discussion

The self-schema maintains relatively stable information about one's personality and preferences that are at the core of one's identity. In the present study, we investigated the causal role of the vmPFC in supporting the self-schema by having patients with vmPFC lesions and controls provide judgments about the likelihood for them (vs. another person) to engage in a series of activities, on two separate occasions. To the extent that self-related judgments rely on a stable set of knowledge (self-schema), these judgments should prove relatively stable across testing sessions. This should not necessarily apply to other-related judgments, which supposedly rely on fewer or less strong memories or schemata.

As predicted, we found that healthy participants and control patients were consistent in their endorsement of self-related activities across sessions and were significantly more consistent for self- compared to other-related information. This consistency advantage for self-related judgments was not apparent in vmPFC patients, who displayed comparably unstable judgments about the self and the other. This finding confirms and extends previous single case observations of impaired (stability of) self-trait knowledge in patients with lesions to the mPFC (Philippi et al., 2012b; Marquine et al., 2016), pointing to the generalizability of these findings to different vmPFC patients and to other domains of self-knowledge (activities and personal preferences).

Before discussing this finding further, it is important to emphasize that the inconsistency in self-related judgments displayed by vmPFC patients cannot be ascribed to an unspecific effect of brain lesions in reducing cognitive functioning or the sense of self (Ciaramelli et al., 2019), as it was not observed in (control) patients with lesions not including vmPFC. Additionally, vmPFC patients' performance is unlikely to be reflective of erratic responding or poor compliance with the task because these patients were as consistent as control patients when judging other-related knowledge, suggesting a more prominent

impairment in self-related knowledge. There is another aspect of our results that underlines the selective impairment in self-related compared to other-related knowledge in vmPFC patients, which pertains to the confidence levels associated with vmPFC patients' judgments. First, vmPFC patients were generally less confident in their answers than were healthy controls (see also Barron et al., 2015; Lebreton et al., 2015; Hebscher & Gilboa, 2016; Gherman & Philastides, 2018), though they were not significantly less confident than control patients. Most importantly, in the brain-damaged and healthy control groups confidence levels associated with both self-related and other-related judgments (at the first testing session) predicted the consistency of these judgments from the first to the second session. That is, unsurprisingly, judgments associated with high confidence tended to change less between sessions. In vmPFC patients, the expected relation between confidence and consistency was only present for other-related judgments. By contrast, vmPFC patients' confidence levels and score change for self-related information were completely unrelated: their judgments about themselves and the typical activities they engaged in fluctuated over time, regardless of whether the judgments were associated with high or low confidence. Thus, while not all judgments are unreliable in vmPFC patients, self-related judgments are. It is self-related- and not other-related- knowledge that proves inconsistent and disconnected from confidence in vmPFC patients, whereas both self- and other-related knowledge are tied to confidence in the control groups.

Together, these findings point to a selective degradation of the self-schema following vmPFC damage, which is consistent with fMRI studies and meta-analyses showing an involvement of the medial prefrontal cortex in self- vs. other-related processing (D'Argembeau et al., 2007, 2014; Jenkins et al., 2008; Wagner et al., 2012), with a ventral-to-dorsal gradient, with vmPFC mostly associated to self-related processing, and dorso-medial prefrontal cortex (dmPFC) with other-related processing (Denny et al., 2012; see also Lieberman et al., 2019). Neuropsychological evidence from patients with medial prefrontal damage is also consistent

with this. We have shown (Stendardi et al., 2021), as have others (Philippi et al., 2012a), that vmPFC damage abolishes the self-reference effect (SRE), that is, the increase in memory for information encoded with (as opposed to without) reference to the self. These tasks, however, inherently tap anterograde memory abilities, and therefore the absence of the SRE could be (at least in part) explained by vmPFC patients' encoding deficits. Here, we confirm an impairment in the self-schema in vmPFC patients in a task with no anterograde memory demands, in line with previous evidence (Philippi et al., 2012a; Marquine et al., 2016). Moreover, we extend previous evidence on impaired self (trait) knowledge in vmPFC patients probing knowledge about one's preference and common activities, which is a form of personal semantic less abstract than self-trait summaries, and more likely to be tied to autobiographical facts and events (*e.g.*, did I ever go bungee jumping? Grilli & Verfaellie, 2014). Together, our current results and those of previous studies probing self (trait) knowledge (Philippi et al., 2012a,b; Marquine et al., 2016; Stendardi et al., 2021) reinforce the view of an impairment of the self-schema following vmPFC damage. One important follow up of this study would be to manipulate the closeness of the other, hence the strength of other-related schemata, to verify whether vmPFC patients' impairment is selective for self-related knowledge or rather extends to other types of schema-related knowledge (see Aron et al., 1991; Kim & Johnson, 2014).

Although the vmPFC patients involved in the present study do not show evidence of confabulation, confabulation is a common consequence of vmPFC damage, and therefore our findings speak to current theories on the role of vmPFC in confabulation. According to Gilboa and his colleagues, confabulation arises as a failure of the "feeling of rightness" (FOR), a pre-conscious monitoring process mediated by vmPFC at the basis of the confident endorsement (or rejection) of information based on the automatic intuition of its veracity (Gilboa, 2004, 2010; Gilboa et al., 2006; Hebscher & Gilboa, 2016). The intensity of FOR is deemed to depend on the match between incoming information and schematic knowledge, with strong schemata,

such as the self-schema, giving rise to the strongest confidence signals (Gilboa, 2004, 2010; Hebscher & Gilboa, 2016). On this view, a damage to vmPFC should lead to an inability to filter our self-relevant yet false information, and to high confident false memories (see also Gilboa & Verfaellie, 2010; Kopelman, 2019; Ciaramelli & Spaniol, 2009). To test this hypothesis, Gilboa et al. (2006) tested confabulating and non-confabulating vmPFC patients in an autobiographical recognition memory task involving true statements about their past, plausible lures, and implausible lures that were blatantly inconsistent with vmPFC patients' life history. Confabulating compared to non-confabulating vmPFC patients showed significantly more false recognitions of implausible lures and were highly confident in their (false) memories. This finding is consistent with an impairment of the FOR and the self-schema in (confabulating) vmPFC patients, and makes contact with our current finding of impaired self-knowledge and untied confidence and consistency of self-related knowledge following vmPFC damage.

To conclude, we have shown that vmPFC patients have an impairment in self-related knowledge, which proved highly unreliable across testing sessions, as if retrieved from, or through, a degraded self-schema. In addition, we found that the confidence levels accompanying self-related judgments were not reflective of their consistency, a finding reminiscent of confabulatory behaviour, and that applied selectively to self-related but not other-related judgments. These findings indicate that the vmPFC is crucial to maintain the self-schema and support the reliable retrieval of self-related information.

Chapter 3. Disentangling schema reinstatement and instantiation following vmPFC damage

Since, at this point, I hope that *I* (or better, the literature) have convincingly established a robust link between schematic cognition and vmPFC activity, now we turn the question onto the precise role vmPFC plays in it. Hence, here we ask: what specific schema-mediated processes depend on vmPFC integrity? Thus, in this chapter, I will focus my investigation on vmPFC involvement in activating (*reinstating*) and using (*instantiating*) event schemata (scripts).

Study 3. Follow the script: the role of vmPFC in reinstatement and instantiation of event schemata

Debora Stendardi, Nicola Ciavatti, Eloisa Bianchi Rossi, Erida Meminaj, Davide Braghittoni & Elisa Ciaramelli

Introduction

What will happen at your next conference? And how do you have an answer even if it has not happened yet? Having (supposedly) attended many conferences, what you (we) have by now is a schematic representation of what a conference entails. We know registration is the first step, and lectures, talks, posters, and hopefully some good catering will follow. This representation has been formed by extracting the regularities from multiple similar experiences,

i.e., all the conference we have been to (Ghosh & Gilboa, 2014; Robin & Moscovitch, 2017). Schemata play a crucial role in shaping the structural organization of learning, memory, and behaviour (Ghosh et al., 2014; Spalding et al., 2015). They influence memory formation and retrieval, facilitating or biasing perception, encoding and recollection of information and events (Bartlett, 1932; Wagner et al., 1998; Wang et al., 2012). Despite their significance, the neural mechanisms underlying intricate knowledge structures such as schemata are not fully understood. Recent evidence point to medial prefrontal regions as having a prominent role in schema-mediated cognition (for a review, see Gilboa & Marlatte, 2017). fMRI studies have consistently implicated the medial prefrontal cortex (mPFC) in schematic processing. For example, mPFC has been observed to exhibit patterns of activity that track the structure of specific schemata during event perception (Baldassano et al., 2018). Moreover, the congruency effect in memory (*i.e.* the mnemonic superiority of information consistent with prior knowledge), has been convincingly linked to mPFC (van Kesteren et al., 2013). mPFC is also heavily involved in scene- and event-construction, supposedly concerned with the activation of schema-related knowledge in neocortex (Ciaramelli et al., 2019; Barry et al., 2019; Monk et al., 2021). Even one of the most important schemata we have, the self-schema, has been shown to be strongly supported by medial prefrontal (and more specifically, ventro-medial prefrontal, vmPFC) regions (for a review, see Wagner & Haxby, 2012).

Schemata are also necessary structures to encode and coherently organize information in memory (Spalding et al., 2015; Bahk & Choi, 2018). Indeed, mPFC has been shown to be consistently activated during autobiographical memory recollection (Gilboa, 2004; McCormick et al., 2020), which is well known to share many of its neural bases with episodic future thinking (EFT) because of their (re)constructive nature, heavily recruiting the Default Mode Network (Atance & O'Neill, 2001; Addis et al., 2007; Schacter et al., 2007; Buckner et al., 2008). It has been suggested that what enables scene and event construction, which are the

bases of EFT, is an interplay between vmPFC and the hippocampus (Ciaramelli et al., 2019). Specifically, vmPFC might be responsible for the activation of schematic knowledge in neocortex, which is then used by the hippocampus to construct a sketch of a rudimentary scene; then, vmPFC would engage in feedback loops with the hippocampus to create and monitor the dynamic unfolding of an event (McCormick et al., 2021), a hypothesis supported by MEG studies (Barry et al., 2019; Monk et al., 2021). One way to test predictions of the model is to study patients with brain lesions. Indeed, if the role of the hippocampus is primarily to build a spatial representation, then hippocampal patients should present with severe deficits in all tasks concerning spatial cognition, findings that are by now classics of the neuropsychological literature (Abrahams et al., 1997; Bohbot et al., 1998; Kessels et al., 2001). Event generation and EFT are also severely hindered by a hippocampal damage, leaving amnesic patients unable to imagine new experiences (Hassabis et al., 2007b; see also Maguire & Hassabis, 2011). Interestingly, a hippocampal damage results in deficits in script generation tasks only for scene- (vs object-) based scripts (Lynch et al., 2020). According to the model, vmPFC lesions, on the other hand, should degrade the activation of schematic knowledge and the dynamic unfolding of events.

Indeed, event generation deficits of vmPFC patients are well established by the previous literature (Bertossi et al., 2017a; McCormick et al., 2018a). When asked to imagine specific events, patients with vmPFC damage typically produce less detailed accounts than healthy and brain-damaged controls (Bertossi et al., 2016a,b), an impairment that cannot simply be ascribed to poor working memory capacities (see Bertossi et al., 2017a). vmPFC patients also tend to focus their descriptions to momentary snippets (Kurczek et al., 2015); what they seem to lack is precisely the dynamic unfolding of events observed in the narratives of healthy controls. Some studies also noted how their event construction deficits tend to present more severely for future- (vs. past-) oriented cognition (see Fellows & Farah, 2005; Ciaramelli et al., 2021).

Recently, Lieberman et al. (2019) used a multi-method and multi-domain approach to review evidence in support of functions linked to sub-divisions of mPFC; as per vmPFC, they highlighted its involvement in "situational processing", which they define as the way a situation is represented, integrating spatial, temporal, causal, evaluative, and social aspect, a finding that fits perfectly with lesion studies testing vmPFC patients' event construction abilities.

Another striking impairment typically observed in patients with vmPFC lesions is a degrading of the self-schema. Indeed, a damage to vmPFC causes the disappearance of the self-reference effect, *i.e.* the mnemonic advantage for items related to the self-schema (Rogers et al., 1977; Stendardi et al., 2021), a deficit not attributable to patients' memory impairments (Stendardi et al., 2023). Furthermore, the literature suggests that vmPFC patients seem to struggle in incorporating the self-schema into imagined events (Verfaellie et al., 2019). As mentioned above, these bodies of literature can be bridged together through the schema theory (Ghosh & Gilboa, 2014; Gilboa & Marlatte, 2017; Ciaramelli et al., 2019). What fundamentally underlies event construction, future thinking, self-related cognition, and situational processing, is schematic knowledge (van Kesteren et al., 2010a,b; Ghosh et al., 2014; Robin & Moscovitch, 2017). Consequently, the question of whether patients with vmPFC lesions can explicitly represent schematic knowledge becomes of central importance (see Ghosh et al., 2014; Ghosh & Gilboa, 2014; Spalding et al., 2015; Hebscher & Gilboa, 2016).

Because of the relatively recent advancements in lesion analyses and lesion mapping techniques, the firsts studies investigating the matter retained a broad focus on patients that suffered frontal (and later, prefrontal) lesions (see Sirigu et al., 1995; Godbout & Doyon, 1995; Allain et al., 1999, 2001; Godbout et al., 2004; Wood et al., 2005). However, albeit with mixed findings on the exact nature of their impairment (probably due to differences in the tasks and scoring procedures employed), all studies reported some form of degradation of event-script knowledge in frontal patients, be it sequencing or boundary errors, choice of incoherent

headlines or difficulties in the classification of an action's belonginess to its event schema (script). Recent studies have started to focus more strictly on event script knowledge in patients with ventromedial prefrontal lesions. Ghosh et al. (2014) tested whether vmPFC patients were able to classify actions' belonginess to their schema, and found a striking difference in performance between non-confabulating and (prior or present) confabulating vmPFC patients, the former performing comparably to healthy adults, and the latter failing at rejecting lures and correctly pair actions and schemas. However, in a subsequent study, Giuliano et al. (2021) reported the same impairment even in non-confabulating vmPFC patients. Here, we note that (to our knowledge) no study has ever tested vmPFC patients in a simple script generation task.

Since the literature has convincingly established a strong connection between schematic processing and vmPFC, now the question turns to the precise role of the region in schema-mediated cognition. Specifically, vmPFC could be concerned with the activation (reinstatement) or the usage (instantiation) of schemata, or both. Reinstating a schema involves both initiating and maintaining a representative, abstracted prototypical model, comprised of both its elements and their interconnections, *i.e.*, the schema (Gilboa & Marlatte, 2017). Instantiation, on the other hand, refers to the process of matching input from the environment with the already reinstated schemata, and can in turn facilitate (or even bias) ongoing perception, memory, and imagination (see also Baldassano et al., 2018; Giuliano et al., 2021; Masis-Obando et al., 2022; Guo et al., 2023a,b). Thus, reinstatement is the first process to take place, and then, instantiation follows through. However, we note that, since the two processes occur together in naturalistic settings and in day-to-day life, one must disentangle the two in order to study the precise role of vmPFC in one or the other.

Here, we aim to study script reinstatement in patients with vmPFC lesions by using a script generation task, that requires reinstating a script which is then not used (instantiated) to perform any task. Moreover, we aim to test whether vmPFC patients' event generation deficits

are more severe when they imagine an event that does not adhere to a precise script, which supposedly requires the selection, reinstatement and instantiation of many different scripts. To this aim, we test participants in an event generation task, in three conditions: when the event to imagine does not obey any script (non-scripted), when it does obey a script (scripted) and when it obeys a script that is externally provided during event imagination (cued scripted). We hypothesize that vmPFC patients' poor schematic knowledge might underlie (or contribute to) their event generation deficits. To specifically tackle script instantiation, we test whether vmPFC patients would be able to instantiate schematic knowledge when externally provided with the relevant script (cued scripted), thereby bypassing endogenous reinstatement.

Methods

Preliminary phase: selection of scripted events and script extraction. A group of 34 healthy young adults (15 M; mean age = 22.5 years; mean education = 16 years), not involved in the main experiment, were presented with 28 short headlines indexing events that would normally last for a day or less, through the platform Qualtrics (http://www.qualtrics.com). Activities were preselected to either obey a script (N = 14, *e.g.*, going to a restaurant, taking a shower) or not obey a script (N = 14; *e.g.*, looking for a lost watch, getting to know new neighbours). To pre-select events not obeying a script we adapted some open-ended activities mentioned in the Means and Ends Problem Solving task (MEPS; N = 5; Platt & Spivack, 1975; *e.g.*, looking for your lost watch, getting to know new neighbours) or created new ones (N = 9; *e.g.*, entertain kids while babysitting, retrieve a ball from a tree). The preliminary rating experiment served to confirm whether the preselected events did in fact obey a script (or not).

After reading the main headline relative to each of the 28 events (*e.g.*, taking your dog for a stroll), participants had to list the 10 actions most likely to constitute that event, that is,

which most people would perform while engaging in it, while avoiding idiosyncratic actions referring specifically to their own personal behaviour. The initial and final action comprising each event were specified in parenthesis beside the headline.

For each action, we calculated the frequency with which participants mentioned it, regardless of the order it appeared in the list. Actions that were described with different words but referred to the same activity (*e.g.* "write down the shopping list" and "take note of what to buy at the grocery shop") were considered as the same action. We defined major actions those that were mentioned by at least 65% of participants, minor actions those that were mentioned by 45-64% of participants, and trivial actions those mentioned by 25-44% of participants (see also Godbout & Doyon, 1995, 2000). We then selected 12 events that contained $\geq$ 5 major actions (mean and SD: 6.6 ± 1.4), which served as the scripted events sample, and 4 events that contained $\leq$ 2 major actions (mean and SD: 1 ± 1.2), which served as the non-scripted event sample. The complete list of non-scripted and scripted stories, along with their major, minor and trivial actions, is reported in Appendix A.

Experimental phase

Participants

The study involved 8 patients with lesions to the vmPFC (6 M, mean age = 58 ± 5.9 years, range = 50-65; mean education = 10.6 ± 3.2 years, range = 5-14), 11 patients with lesions not including vmPFC (9 M, mean age = 55.8 ± 12.4 years, range = 34-68; mean education = 13.3 ± 3.1 years, range = 8-19), and 46 healthy participants (32 M, mean age = 58.2 ± 6.5 years, range = 40-68; mean education = 11.5 ± 3.3 years, range = 5-18). Patients were recruited at the Centre for Studies and Research in Cognitive Neuroscience of the University of Bologna, Cesena Campus. The three groups were matched for gender balance ($\chi^2_2 = 0.7$, $p = 0.7$), age ($F_{2,62} = 0.4$ $p = 0.7$), and education ($F_{2,62} = 1.8$, $p = 0.2$). All patients were in the stable phase of

recovery (at least 3 months postmorbid). vmPFC patients' lesions resulted from a ruptured aneurysm of the anterior communicating artery (AcoA) in all cases. In 6 cases, vmPFC lesions were bilateral, while the other two vmPFC patients had unilateral lesions (right in one case, left in the other). Control patients' lesions were caused by ischemic or haemorrhagic stroke, or tumour resection. Eight patients presented with right brain lesions and two with left brain lesions. Lesions affected the right temporal lobe in four cases, the right fronto-temporal cortex in two cases (in one of these cases, the patient had an additional left temporal lesion), the right temporo-parietal cortex in one case, the right occipito-temporal cortex in one case, the left temporal lobe in one case, and the left occipital cortex in one case. The remaining patient presented with an entirely sub-cortical lesion**.** The average volume lesion size did not differ significantly between the two groups of patients (vmPFC Patients: 48.7 cc vs Control Patients: 52.7 cc, $p = 0.42$). Patients underwent neuropsychological testing (except for one control patient), and their performance was comparable in terms of attentional capacity (attentional matrices: 49.1 vs 45.4, $t_{16} = 1.1$, $p = 0.3$), digit and visuo-spatial memory span (digit span: 5.25 vs 5.8, $t_{16} = 0.8$, $p = 0.4$; Corsi tapping test: 4.6 vs 4.5, $t_{16} = 0.3$, $p = 0.7$), and phonemic and semantic fluency (phonemic fluency: 30.1 vs 36.8, $t_{16} = 1.5$, $p = 0.2$; semantic fluency: 40.6 vs 53.1, $t_{16} = 1.6$, $p = 0.1$). vmPFC patients received a more extensive neuropsychological evaluation, which we report in table 1. All participants provided informed consent in compliance with the 2008 World Medical Association Declaration of Helsinki, and approved by the Bioethical Committee of the University of Bologna and the Ethical Committee of Area Vasta (CEIIAV) of Emilia Romagna.

	vmPFC patients							
	p.1	**p.2**	**p.3**	**p.4**	**p.5**	**p.6**	**p.7**	**p.8**
Sex	M	M	F	M	F	M	M	M
Age (years)	55	63	65	53	50	61	64	53
Education (years)	13	8	13	10	5	13	9	14
Progressive Raven Matrices (cut-off 15)	34.5	33.5	29	29	15.25	27	34	24.25
Digit Span (cut-off 3.75)	6.75	5	4.75	3.75°	4.25	5.75	5.25	5.5
Corsi Block Tapping (cut-off 3.75)	4.75	5	4.87	3.5*	3.75°	3.5*	6	5.5
Attentional Matrices (cut-off 31)	49.5	48.5	53.75	49.5	49	49.5	52.75	40.25
Prose Passage Recall (ES)*	4.35 (0)*	2.55 (0)*	23 (4)	15 (4)	8 (1)	13.5 (3)	10.5 (2)	12.25 (3)
Rey Complex Figure - Copy (cut-off 28.9)	31.75	32.5	36	36	23.75*	36	34	36
Rey Complex Figure – Recall (cut-off 9.5)	14.25	6.75*	34.7	12.4	11.5	22	17	20
Phonemic Fluency (cut-off 17)	45	27	24	17°	19	32	23	41
Semantic Fluency (cut-off 25)	45	37	62	31	28	35	52	35
Stroop Test – Errors (cut-off 7.5)	0	0	0.5	0	0	0	0.25	6.5
Stroop Test – Interference time (cut-off 27.5)	7.75	17	16	24.75	28*	16	40.75*	40*
Tower of London - Total Move Score (cut-off 60)	78	124	104	62°	<60*	82	122	96
Tower of London - Rule Violation Score (cut-off 60)	<60*	104	110	104	<60*	104	114	104
WCST - Tot Number of Errors (cut-off 62)	81	62*	92	119	<62*	81	62*	62*
WCST - Number of Perseverative Errors (cut-off 62)	81	81	92	138	<62*	92	62*	62*

Table 1. **vmPFC p**atients' demographic and clinical data

Note. Unless specified, we report corrected scores, ∗in this case, we report both corrected scores and equivalent scores in parenthesis (ES), because different versions of the test were used across patients, and therefore corrected scores are not comparable. An impaired performance is signalled by a *, a borderline performance by a °.

Lesion analysis

Individual patients' lesions, extracted from the most recent MRI or computed tomography scans, were manually drawn by two trained psychologists on each slice of the normalized T1-weighted template MRI scan provided by the Montreal Neurological Institute with MRIcro software (Rorden & Brett, 2000). This template aligns approximately with Talairach space (Talairach & Tournoux, 1988). Overlaying the lesion of each patient onto the standard brain allowed us to calculate the overall brain lesion volume in cubic centimetres (cc). Lesion overlap images for all vmPFC patients were generated using the MRIcro software and are displayed in Fig. 1. Brodmann's areas (BA) mainly affected were areas BA 10, BA 11, BA 24, BA 25, BA 32, BA 46, BA 47, with the region of maximal overlap occurring in BA 11 (M = 19.16 cc, SD = 10.27), BA 10 (M = 10.08 cc, SD = 7.04), and BA 32 (M = 6.58 cc, SD = 5.61). On average, BA 11 accounted for 37.4% of each individual patient's lesion (range: 17% - 53%), whereas BA 10 accounted for an average of 18.4% (range 0% - 26%). For all vmPFC patients, the Brodmann area mainly affected was either BA11 (7 patients) or BA10 (1 patient).

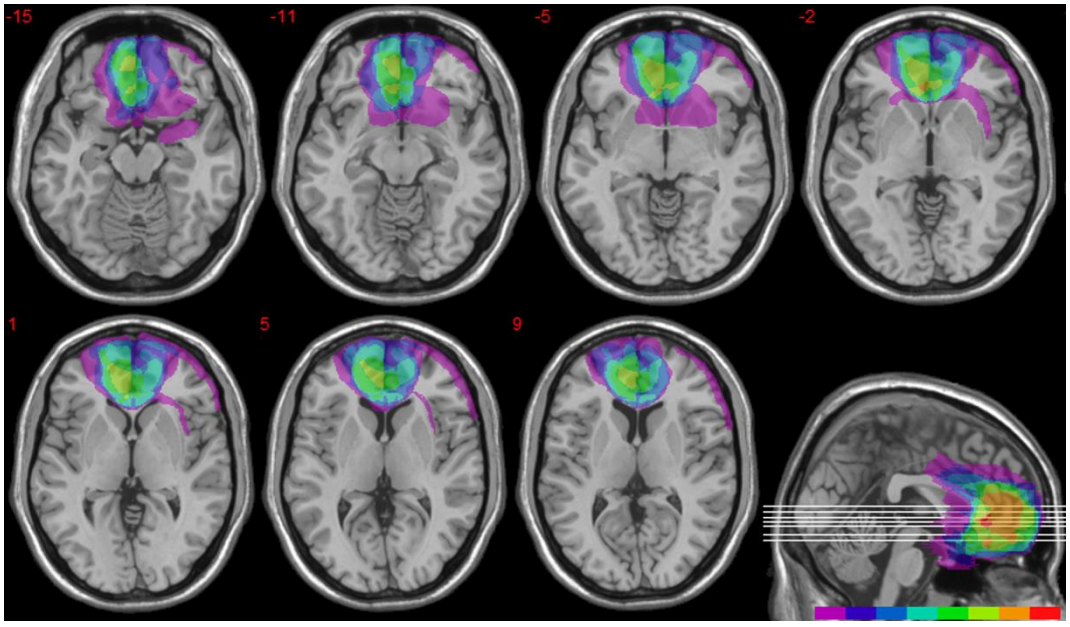

Figure 1. Location and overlap of vmPFC patients' brain lesions. Lesions are projected on the same axial slices and on the mesial view of the standard Montreal Neurological Institute brain. The level of the axial slices is indicated by white horizontal lines on the mesial view of the brain, and by *z*-coordinates. The color bar represents the number of overlapping lesions.

Procedure

Before data collection, we divided the 12 headlines of scripted events into three groups of four events each. This was to ensure balance in the average number of the scripts' major actions across tasks and participants. The three groups of stories had an average number of major actions of 6.75 (group A), and 6.5 (groups B and C). All groups of stories were used for all participants, but in different conditions: a group of stories could be assigned to one of three tasks (scripted event generation, cued-scripted event generation, script generation) in a pseudo-random order. The same group of stories was never repeated within subjects in another task.

Participants were tested individually in two separate sessions, approximately a week apart. Whenever possible, participants were tested in presence, at the lab, otherwise, they were tested online, in video-call with the experimenter (healthy participants: N = 14/46; vmPFC patients: N = 4/8; control patients: N = 7/11). Sessions were audio-recorded. All subjects completed the three event generation conditions (non-scripted, scripted, cued-scripted), plus a script generation task. In the first session, participants were always administered the non-scripted and scripted event generation tasks. Non-scripted stories were the same across participants, whereas scripted stories were the ones from group A, B or C (for a total of 8 stories in the first session). Their order was random, even across conditions (non-scripted, scripted). For each headline, participants were asked to imagine and describe a fictitious event that could plausibly happen to them in the future (*e.g.*, imagine attending a wedding). They were asked to be as rich and detailed as possible in their description. At the end of each story, we asked participants to answer a series of questions about their imagination experience (perceived difficulty, perceived vividness and detailedness, sense of presence, and similarity to a memory) on a Likert scale from 1 to 5. Then, to calculate the Spatial Coherence Index (SCI), we followed the procedure of Hassabis et al. (2007b), and asked participants to evaluate the veracity of 12 individual statements describing the scene and event they imagined. Eight of the statements

related to a rich and vivid imagination experience (*e.g.* "I could see the scene in color"), whereas the other four described a fragmented and more "distant" scene (*e.g.* "It was more a collection of separate images rather than a scene"). Each flagged positive statement received a +1, and each negative one received a -1. We then subtracted two points from the SCI score, so that its range of values would be centred on 0 (from -6 to +6).

In the second session, participants were first administered a script generation task. Here, we presented four headlines of scripted events (from group A, B or C), one at a time, and we asked participants to list the series of actions they thought was most likely to constitute the relevant event; this time, the starting and ending point of each script were specified in the headline. Participants were explicitly instructed to list the relevant actions that the event would usually entail for most people in our culture rather than the actions that they would normally do for that activity. Then, participants completed the last event generation task, the cued-event generation. Here, instructions were identical to session 1 (scripted and non-scripted event generations), but this time all the major actions (isolated in the preliminary phase) of the relevant scripts were displayed. Participants were thus shown the actions that normally constitute the event they were asked to imagine; we made clear that using actions from the list was not mandatory. The list of actions remained visible all throughout participants' accounts. At the end of each story, we asked the same questions about their imagination experience of session 1 (perceived difficulty, perceived vividness and detailedness, sense of presence, similarity to memory, and Spatial Coherence Index). The tasks' order of the second session was set as such to avoid a possible learning effect from the cued event construction phase to script generation.

Scoring

Event generation (non-scripted, scripted, cued-scripted). All stories were transcribed and scored according to the Autobiographical Interview (AI) protocol developed by Levine et al. (2002). First, the text was segmented into details, which denote unique bits of information: then, details were categorized as either internal or external. Internal details are defined as pertaining to the main event, which, in our case, was the event given as cue for the trial. Internal details were further classified based on the type of information conveyed, and could be scored as event, place, time, perceptual, or thought/emotion. External details were categorized as event, semantic, repetition, or other (see Levine et al., 2002). For the cued event generation task, for each major action participants used in their account, one detail was subtracted from the story (since it was externally provided).

Script generation. Actions mentioned by participants were assessed through a scoring procedure that allowed us to investigate the semantical aspect of a script (core), along with different possible types of errors (see Bower et al., 1979; Roman et al., 1987; Godbout & Doyon, 1995; 2000; Godbout et al., 2004; St-Laurent et al., 2009; Lynch et al., 2020). First, the total number of actions reported by participants was noted. All actions of a given script were then compared with the ones isolated in the preliminary phase; all matches with major, minor and trivial actions were classified as such[4]. The other actions could be assigned to one of three error categories: idiosyncratic errors, relevant intrusions and irrelevant intrusions. Boundary errors and sequencing errors were counted separately, as they were not part of the total number of actions. An error was considered idiosyncratic when the described action was specific for the participant, not generalizable to most people within the culture (*e.g.* "order steak" instead of "order food" at the restaurant script; see also Lynch et al., 2020). Actions that

[4] When a major action was hyper-segmented, *i.e.* broken down into two or more actions instead of one (*e.g.* "take products" and "put them in the shopping trolley" instead of "shop"), it was counted as one major action: the others were not scored as anything else, and would only affect the total number of actions uttered by the participant.

were mentioned by less than 25% of participants of the preliminary phase were scored as intrusions; whenever an action was appropriate in the script's context (*e.g.* "starting the car to go home" in the shopping for groceries script) it was scored as a relevant intrusion; otherwise, it was counted as an irrelevant intrusion (*e.g.* "going for a pizza after the movies" in the cinema script; see also Godbout & Doyon, 2000; Scott et al., 2011). Boundary errors referred to an incorrect starting and/or ending of the script (which were given in the headline, but not always respected by participants), hence, for each script, boundary errors could range from 0 to 2. Sequencing errors described the occurrence of an impossible or not natural sequence of events (*e.g.* "take your keys" after "leave your house"; see also Scott et al., 2011; Lynch et al., 2020).

Results

Script generation

The composition of scripts for each group (excluding hyper-segmented actions) is reported in Fig 2.

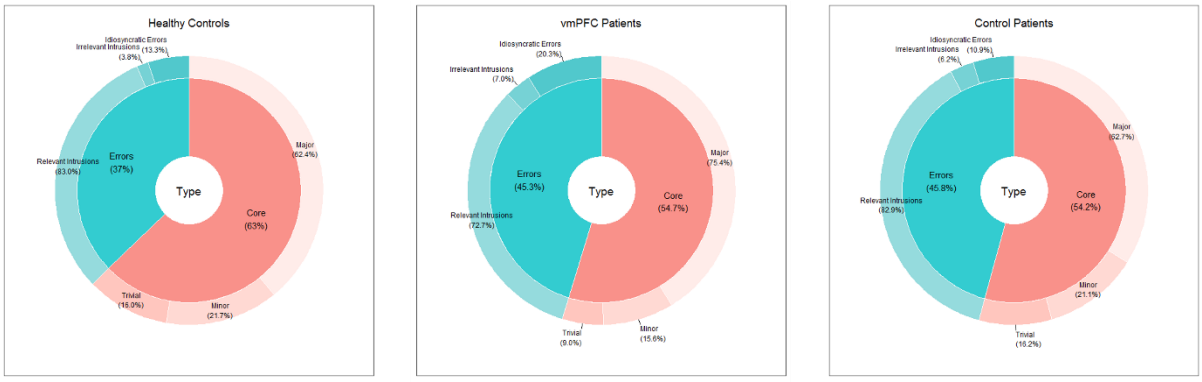

Figure 2. Composition of scripts in the three groups. Note that hyper-segmented actions are not considered.

To ensure that any potential difference among groups was not driven by a gross difference in the total number of actions reported, we first conducted a one-way ANOVA on the total number of actions with Group as predictor, and confirmed that the three groups did indeed nominate a comparable total number of actions ($F_{2,62} = 2.4$, $p = 0.1$). Then, we examined

the core of scripts. For each category (major, minor, trivial), and each participant, we computed the percentage of actions correctly stated, by dividing the total number of actions of that category uttered by the participant by the number of actions of that same category that were present in the four scripts administered (*i.e*. mentioned by 25% or more of the sample in the preliminary phase). Then, we conducted three separate ANOVAs with Group as predictor: the ANOVA on the percentage of major actions did not unveil any significant effect of Group ($F_{2,62}$ = 0.7, p = 0.5), meaning the three groups correctly stated a comparable number of major actions. Conversely, the ANOVA on the percentage of minor actions revealed a significant effect of group ($F_{2,62}$ = 4.2, p = 0.02, η_p^2 = 0.12), with post-hoc Fisher's tests showing significant differences when comparing vmPFC patients to both healthy controls (0.22 vs 0.39, p = 0.005) and control patients (0.22 vs 0.36, p = 0.04), and no difference between the control groups (0.39 vs 0.36, p = 0.7). Similarly, a significant effect of Group was present in the ANOVA on the percentage of trivial actions ($F_{2,62}$ = 3.2, p = 0.048, η_p^2 = 0.09). Post-hoc Fisher's tests highlighted significant differences between vmPFC patients and both healthy controls (0.13 vs 0.27, p = 0.02) and control patients (0.13 vs 0.29, p = 0.03), and no differences between control groups (0.27 vs 0.29, p = 0.7; see Fig. 3). None of the ANOVAs on the total number of errors committed by participants (sequencing errors, idiosyncratic errors, relevant and irrelevant intrusions, boundary errors) revealed any effects (all ps > 0.2), indicating that vmPFC patients and healthy and brain-damaged controls committed a similar number of mistakes.

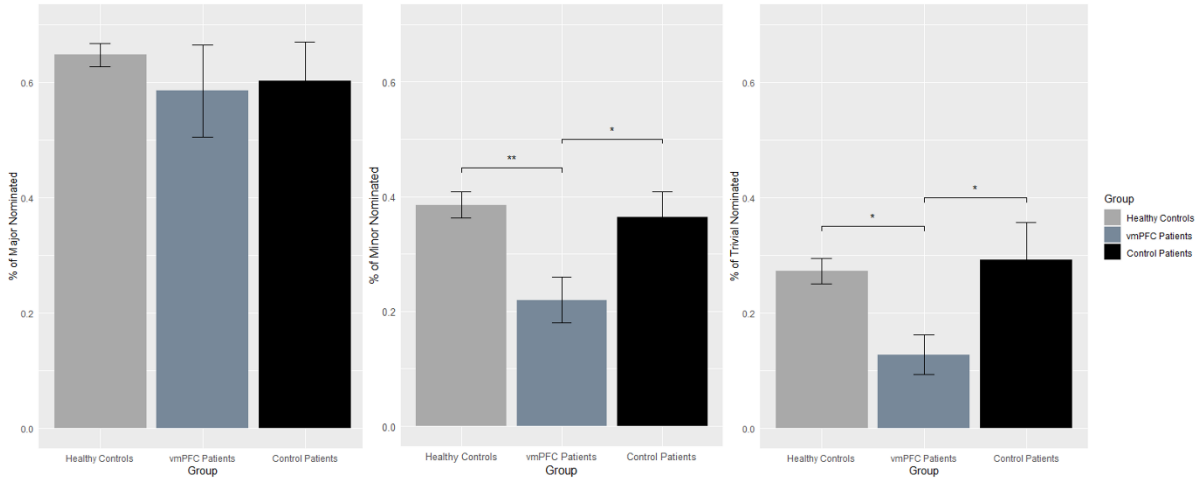

Figure 3. Percentages of major, minor and trivial actions correctly stated by the three groups.

Event Generation

Self-reported ratings. We ran a series of repeated measures ANOVAs on self-reported ratings, with Group (healthy controls, vmPFC patients, control patients) and Condition (non-scripted, scripted, cued-scripted) as factors. First, the analysis on the Spatial Coherence Index highlighted a significant effect of Condition ($F_{2,124} = 13.7$, $p < 0.0001$) and a Group*Condition interaction ($F_{4,124} = 2.6$, $p = 0.04$, $\eta_p^2 = 0.09$), which was driven by control patients displaying a different SCI trend than the other groups. Specifically, healthy controls and vmPFC patients both had lower SCIs in the non-scripted condition as compared to the scripted and cued-scripted conditions (all ps < 0.01), but this was not the case for control patients (all ps > 0.4). There were no differences between healthy controls and vmPFC patients in SCI when comparing the same conditions across the two groups (all ps > 0.3). Similarly, the ANOVA on perceived difficulty revealed an effect of Condition ($F_{2,124} = 13.2$, $p < 0.0001$) and a Group*Condition interaction ($F_{4,124} = 3.3$, $p = 0.01$, $\eta_p^2 = 0.1$), with post-hoc comparisons showing that control patients did not display the typical trend of perceived difficulty increasing from the non-scripted to the other conditions. Indeed, healthy controls and

vmPFC patients rated the non-scripted condition as more difficult than the other two (all ps < 0.01), while control patients did not (all ps > 0.6). Again, healthy controls and vmPFC patients gave similar ratings of difficulty across the same conditions (all ps > 0.5). The ANOVAs on the other self-reported ratings (perceived detailedness, perceived vividness, similarity to memory, sense of presence) all revealed Condition as the only significant effect (all ps < 0.0001). In all cases, the non-scripted condition was statistically different than the scripted and cued-scripted conditions, that were instead comparable to each other. Specifically, participants rated the imagination experience for the non-scripted condition as less detailed, less vivid, less similar to memories, end evoking a weaker sense of presence (all ps < 0.0001), with the scripted and cued-scripted conditions being judged similarly across the same self-reported ratings (all ps > 0.09).

Internal and External details. We ran a repeated measures ANOVA on the number of details with Group as between factor, and Condition (non-scripted, scripted, cued-scripted) and Type of detail (internal, external) as within factors (see Fig. 4). There were significant main effects of Condition ($F_{2,124} = 3.8$, $p = 0.02$), and of Type ($F_{1,62} = 44.6$, $p < 0.0001$), a significant Group*Type interaction ($F_{2,62} = 5$, $p < 0.01$), and a three-way Group*Condition*Type interaction ($F_{4,124} = 3.2$, $p = 0.02$, $\eta_p^2 = 0.09$). Post-hoc Fisher's tests revealed different patterns of performance in terms of internal details across groups and conditions. In particular, healthy controls' internal details remained stable when comparing non-scripted and scripted conditions (30.2 vs 27.9, $p = 0.2$), but increased from the scripted to the cued-scripted condition (27.9 vs 33.2, $p = 0.004$), with no difference between non-scripted and cued-scripted conditions (30.2 vs 33.2, $p = 0.09$). Control patients, on the other hand, displayed an increase of internal details from the non-scripted to the scripted condition (28.5 vs 37.4, $p = 0.02$), which then remained stable in the cued condition (37.4 vs 35.0, $p = 0.5$), again with no difference between non-scripted and cued-scripted conditions (28.5 vs 35.0, $p = 0.07$). vmPFC patients' performance

was instead characterized by a “step by step” improvement in the number of internal details, which did not reach significance when comparing non-scripted and scripted conditions (11.6 vs 18.3, $p = 0.1$), or scripted with cued-scripted (18.3 vs 24.4, $p = 0.2$), but revealed itself when contrasting non-scripted and cued scripted conditions (11.6 vs 24.4, $p = 0.003$). Indeed, vmPFC patients reported less internal details in the non-scripted condition than both healthy controls (11.6 vs 30.2, $p = 0.002$) and control patients (11.6 vs 28.5, $p = 0.02$). However, the magnitude of this difference decayed in the scripted condition, reaching significance only when comparing vmPFC patients to control patients (18.3 vs 37.4, $p = 0.007$), but not to healthy participants (18.3 vs 27.9, $p = 0.1$). In the cued-scripted condition, there was no evidence that vmPFC patients reported fewer internal details than the control groups (vmPFC patients vs healthy controls: 24.4 vs 33.2, $p = 0.1$; vmPFC patients vs control patient: 24.4 vs 35, $p = 0.1$).

External details did not vary across conditions for neither vmPFC patients, nor control patients (all ps > 0.4), although healthy controls reported a higher number of external details in the scripted condition as compared to the other conditions (all ps < 0.02). Across all three conditions, the three groups reported a comparable number of external details (all ps > 0.2). Interestingly, control groups’ stories consistently had a higher number of internal rather than external details in all conditions (all ps < 0.0001), whereas for vmPFC patients, this was the case only for the cued-scripted condition (non-scripted: 11.6 vs 12.5, $p = 0.8$; scripted: 18.3 vs 14.5, $p = 0.4$; cued-scripted: 24.4 vs 14.8, $p = 0.03$).

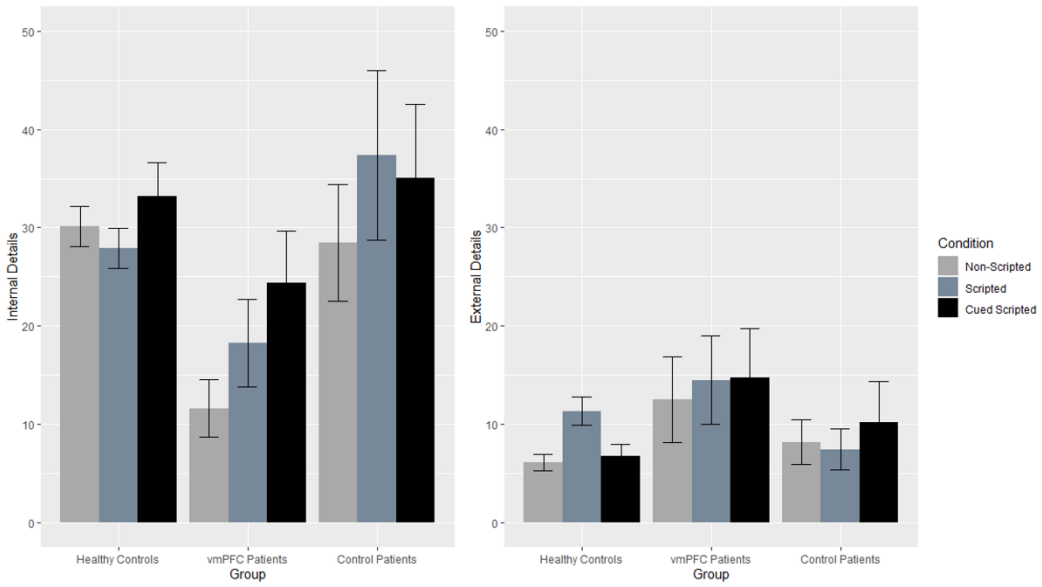

Figure 4. Average number of internal (left panel) and external (right panel) details in participants' stories by group and condition.

Correlation between script- and event-generation. To see whether vmPFC patients' impoverished knowledge of finer aspect of event scripts (*i.e.* minor and trivial actions) could be an underlying cause for their poor event generation performance, we ran a multiple linear regression on the average number of internal details (collapsed across conditions) with the percentage of major, minor and trivial actions mentioned at the script generation as predictors. The only significant effect was that of trivial actions ($\beta = 0.5$, 95% CI [0.24, 0.77], $p < 0.001$), indicating that participants who reported more trivial actions also generated richer accounts at the event generation task. To test whether this correlation held for all event generation conditions, we ran a repeated measures ANOVA on the number of internal details, with Condition (non-scripted, scripted, cued scripted), and the percentage of major, minor and trivial actions as predictors, including all possible two-way interaction terms. Again, the only significant effect was that of the percentage of trivial actions ($F_{1,61} = 14.3$, $p < 0.001$, $\eta_p^2 = 0.19$). All interactions involving the variable condition were not significant, indicating that the correlation between the percentage of trivial actions reported at the script generation and the number of internal details in participant's stories was present for all event generation

conditions. To specifically test whether the knowledge of a script's core could also support event generation when there is no script to refer to (non-scripted event generation), we computed correlations between the percentage of major, minor and trivial actions mentioned, and the number of internal details at the non-scripted event generation. The percentage of major was only a marginally significant predictor of the number of internal details (r_{63} = 0.24, p = 0.059), whereas minor and trivial actions were strongly correlated to them (minor: r_{63} = 0.44, p < 0.0001; trivial: r_{63} = 0.52, p < 0.0001; see fig 5).

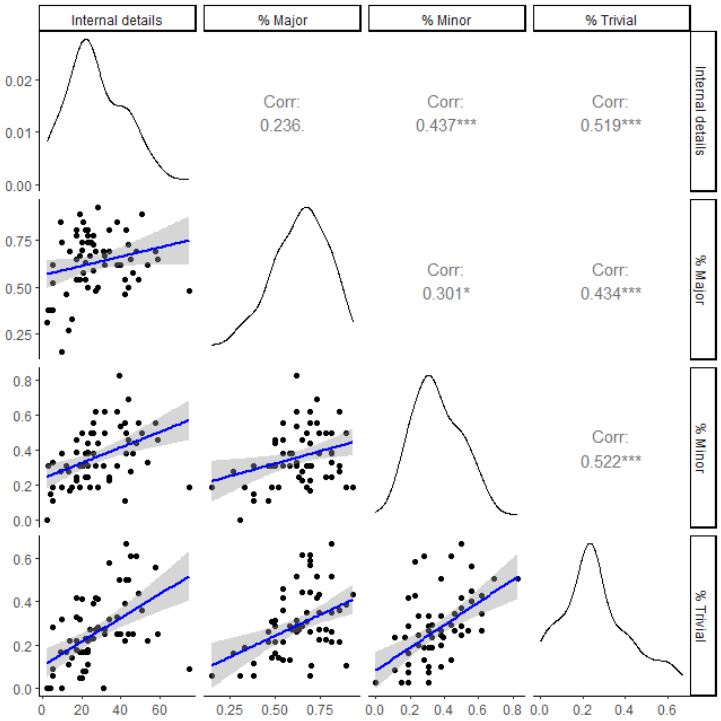

Figure 5. Correlations between the number of internal details at the non-scripted event generation and the percentage of major, minor and trivial actions correctly stated at the script generation task**.**

Discussion

In the present study, we investigated the role of vmPFC in reinstatement and instantiation of event schemata (scripts). Reinstatement was assessed through a script generation task, in which vmPFC patients and healthy and brain damaged controls were asked to generate scripts of everyday activities. Script instantiation, on the other hand, was evaluated by means of an event-generation task, presented in three distinct versions: non-scripted events, scripted events, and cued scripted events, the latter involving externally provided scripts during event imagination. The strategic inclusion of externally provided scripts allowed us to circumvent endogenous reinstatement, a process supposedly impaired in individuals with vmPFC lesions, thereby enabling a focused exploration of their script instantiation abilities.

At the script generation, vmPFC patients demonstrated an adequate knowledge of a script's "backbone", namely, of the actions that most characterize a specific script (*i.e.* major actions). Nevertheless, they failed to articulate a sufficient number of actions that make up the finer details of a script (*i.e.* minor and trivial actions). Interestingly, vmPFC patients did not commit more mistakes than healthy and brain-damaged controls. During event generation, participants perceived non-scripted events as less vivid, evoking a diminished sense of presence, less resemblance to memories, and eliciting less detailed accounts. Also, although control patients did not perceive non-scripted events as more difficult and less spatially coherent than scripted and cued-scripted events, healthy controls and vmPFC patients did. Overall, these results strongly suggest that the experimental manipulation influenced participants' subjective perception of their imaginative experience. In terms of internal details, vmPFC patients performed poorly as compared to the control groups when imagining events than did not adhere to a script, gradually improving when events obeyed a script and when scripts were externally provided to them during event imagination. This enhancement was

apparent in comparison to the control groups, to themselves when cued with the relevant script, and in the fact that only in the cued condition did they produce stories with more internal than external details, something that healthy and brain-damaged controls did in all conditions. Furthermore, correlation analyses revealed how an accurate knowledge of finer aspects of scripts was predictive of the richness of stories at the event generation task, even when those stories did not adhere to any script.

Previous studies have described different kinds of impairments regarding event script knowledge in frontal patients, reporting failure to close scripts, mistakes in ordering actions and choice of irrelevant actions as part of a script (Sirigu et al., 1995; Godbout & Doyon, 1995; Allain et al., 1999; Godbout et al., 2004; Wood et al., 2005). In our study, however, we found that vmPFC patients committed a similar number of errors as compared to the control groups. The sole distinction in their performance at the script generation task was a difficulty in including as many minor and trivial actions as the control groups, aligning with a finding previously reported by Godbout & Doyon (1995). Nonetheless, we note that comparing our results with the ones from the studies mentioned above would be a loose analogy, given their broad focus on patients with lesions spanning the whole frontal lobe. More recent studies, narrowing the investigation to patients with vmPFC lesions, would be more suitable. Ghosh et al. (2014) examined the ability of vmPFC patients to categorize actions within their schema, interpreting their poor performance as evidence of a degraded process of reinstatement, which is consistent with the protective role of vmPFC damage against intrusion errors at the Deese-Roediger-McDermott paradigm (Melo et al., 1999; Warren et al., 2014). vmPFC patients might be unable to activate the relevant schematic knowledge that would later influence decision processes (in Ghosh et al. paradigm) and memory recall (at the Deese-Roediger-McDermott task). Gilboa & Moscovitch (2017) tested vmPFC patients in a face familiarity task, whilst recording evoked-response potentials (ERP). The authors found that, prior to face presentation

(during schema reinstatement), there was a theta band desynchronization between vmPFC and lateral temporal regions associated with semantic knowledge, a desynchronization that predicted N170 modulation from face familiarity. Patients, however, exhibited diminished vmPFC-posterior cortical desynchrony and lacked the N170 modulation, a pattern that aligns with the notion of a degraded schema reinstatement in vmPFC lesioned patients. Giuliano et al. (2021) investigated the temporal dynamics of both reinstatement and instantiation in healthy controls and vmPFC patients. They reported prolonged, sustained theta desynchronization resulting from the interaction between vmPFC and posterior cortical regions, which are known to be relevant to schema reinstatement. vmPFC patients showed the least amount of interregional desynchronization (see also Gilboa & Marlatte, 2017). Instantiation was instead associated with post-stimulus desynchronization in the alpha and beta frequency ranges, reflecting interactions between vmPFC and the lateral temporal cortex (LTC), a finding that seems to suggest an involvement of vmPFC in both reinstatement and instantiation. However, here we note that, if a vmPFC damage compromises schema reinstatement, then schema instantiation, which crucially depends on a successful reinstatement, would be (at least partially) degraded. Consequently, one cannot precisely pinpoint the nature of such degradation, since vmPFC could fail to instantiate a schema because it rests on an already unsuccessful reinstatement.

Consistent with the previous literature, our findings reveal a partially degraded reinstatement ability after vmPFC damage. Indeed, when asked to generate scripts of everyday activities, patients with vmPFC lesions were able to correctly pinpoint the actions that make up a script "skeleton", but they were unable to mention an adequate number of actions that constitute a script's finer details, as if vmPFC patients reinstated an impoverished or incomplete script. Our finding aligns with the hypothesis that vmPFC patients do indeed reinstate a schema, but one that is too broad or "nebulous" (Giuliano et al., 2021; see also

Shallice & Cooper, 2012; Ghosh et al., 2014). This interpretation also fits with vmPFC patients' performance at the event generation task. Given that, for event construction to be successful, the process must rely on an already reinstated schema (Irish, 2020; see also D'Argembeau & Mathy, 2011; Ciaramelli et al., 2019), when there is an ambiguity on which schema is appropriate to reinstate (non-scripted event generation), the vmPFC has to perform an additional step by selecting, activating and re-arranging numerous different scripts, which would then pose the bases for event construction (see Benoit et al., 2014; Peters et al., 2017). Consequently, vmPFC patients' non-scripted event generation stands on many rearranged nebulous scripts, something that leaves their stories deprived of contextual elements (*i.e.* internal details), and more inclined towards semantic information (*i.e.* external details), which are mostly preserved after a vmPFC damage (Ciaramelli & Spaniol, 2009; De Luca et al., 2019; Giuliano et al., 2021). This is also supported by the fact that an accurate knowledge of a scripts' core, and more specifically of its finer aspects, is strongly correlated to the richness of stories at the event generation task, even when stories do not obey a script. The lack of a definite script to base a story on does not impoverish healthy controls' accounts because they have no difficulties in reinstating and re-arranging scripts; however, much like vmPFC patients, they perceive the heavier demands of the task, something that is evident in their self-reported ratings. Conversely, when there is a precise script to refer to (scripted event generation), the ambiguity on which script to reinstate is resolved, which we propose is the reason for the (albeit moderate) improvement displayed by vmPFC patients. However, their stories still rest on a nebulous schema, and present as impoverished and mainly composed of semantic information. vmPFC patients improved their performance when provided with the relevant script of the event, which we speculate happened because vmPFC is relieved from schema reinstatement. Hence, we propose that the process of instantiation might in fact be intact after a vmPFC

damage. However, when vmPFC patients are tasked with both reinstatement and instantiation, the latter appears deficient because it rests on the reinstatement of a broad, nebulous schema.

Further studies are needed to determine whether vmPFC patients suffer from the same loss of fine details of schemata even in other kinds of schematic knowledge (*i.e.*, scene schemata, self-schema). Moreover, it is not clear in what manner are vmPFC patients able to retain some form of schematic knowledge (even if broad and incomplete). It is possible that what we call here "script backbone" (*i.e.* the collection of major actions of a script) is something that is represented in the brain at a completely semanticized level, and hence, not dependent on vmPFC anymore. Alternatively, it might be possible that the degraded schematic knowledge we observed in vmPFC patients comes from the loss of specific characteristics of a schema (*e.g.* the flexibility), perhaps imposed by vmPFC in a healthy brain (see Ghosh & Gilboa, 2014).

To conclude, we have demonstrated that a vmPFC damage causes an impairment of schema reinstatement, and that impairment manifest itself in the form of an incomplete, broad schema activation, in which vmPFC patients lose finer details of event schemata. We propose that vmPFC patients display event generation deficits because their schema instantiation poses on the reinstatement of a broad, nebulous schema. When externally provided with the relevant script during event imagination, vmPFC patients perform comparably to healthy and brain damaged controls, a result that speaks in favour of an intact process of schema instantiation after vmPFC damage.

Chapter 4. Recreating frontal functions in a hybrid neural network

For this section, I feel the need to explain a bit more of the what, why, and how we will try to accomplish what the chapter title aims to achieve. Our goal here is to represent posterior and frontal functions in a neural network. Every neuroscientist knows that, anatomically speaking, brain regions have different characteristics: after all, this was all that Brodmann needed to base his parcellation of the brain into the Brodmann areas we know today. These anatomical differences are even more evident at the macro-scale, if we consider a crude distinction between frontal and posterior parts of the brain. In a neural network, we can model these macroscopic differences by assigning different *cortical parameters*. In our case, the network is "hybrid" because it tries to represent two distinct (but communicating) networks (frontal and posterior). We shall then observe what the two networks do, specifically, how they *latch*. The latching of a network is essentially the jumping between successive *patterns*, which, in a neural network, represent specific configurations of the states of the neural ensembles modelled. Since these kinds of neural architectures are often used to represent memory functions, the patterns are taken to represent memories which have been previously stored in the network, and are then retrieved in the *latching sequence*.

The reason why we embark in such a quest is to investigate whether we can, in fact, create a representation of frontal functions that is accurate enough to reproduce the deficits that we observed in patients with frontal damage (see Chapters 1, 2, 3). We will see that it is indeed possible to artificially simulate lesions to parts of the network.

The "how" we will achieve this, is by using a particular type of neural network: the Potts network. In a Potts network, each node (the Potts unit) represents a patch of cortex composed of many neurons and can exist in one of several possible *states*, generalizing the Hopefield binary network. Essentially, we do not attempt to model each and every neuron of the brain, which given they are roughly 86 billion, would be impractical. Instead, what we do model are *ensembles* of neurons, the Potts unit, which represent patches of cortex. If one is modelling an individual neuron, biology imposes that it should only exist as either "on" or "off". However, if we are modelling a group of neurons, said group can exist in many different *states*, which, as we will see, is one of the parameters that allows us to differentiate the sub-networks we represent.

Study 4. Taking time to compose thoughts with prefrontal schemata*

*This work has been published in Ryom, K. I., Basu, A., Stendardi, D., Ciaramelli, E., & Treves, A. (2024). *Experimental Brain Research*, 2024 Mar 14. doi: 10.1007/s00221-024-06785-z. Epub ahead of print.

Constructive associative memories

Recent explorations of the mechanisms underlying creative forms of human cognition (Mekern et al., 2019; Benedek et al., 2023), ranging from musical improvisation (Beaty, 2015) through visual creativity (Aziz-Zadeh et al., 2013) up to poetry (Stockwell, 2019), or mere mind wandering (Ciaramelli & Treves, 2019), have again questioned the validity of reducing the cortex to a machine operating a complex transformation of the input it currently receives.

On the one hand, sophisticated and massive artificial intelligence systems like ChatGPT or midJourney, with their impressive performance, have adhered to the standard operational paradigm of producing a response to a query. On the other, a simple observation of cortical circuitry, with its extensive recurrence and quantitatively limited external inputs, have long ago led to the proposal that the cortex is (largely) a machine talking to itself (Braitenberg & Schuz, 1991). Likewise, when confronted with an artistic or literary creation we sometimes ask: what was the query? Was there a query?

If it is the cortex itself that takes the initiative, so to speak, is it the entire cortex?

Understanding the mechanisms of cortico-cortical dialogue that generate spontaneous behaviour cannot eschew their statistical character, that of a system with very many imprecisely interacting elements. Valentino Braitenberg suggested a framework for such a statistical analysis, which to a first approximation considers the cortex as a homogeneous structure, not differentiated among its areas (nor, other than quantitatively, among mammalian species; Braitenberg, 1978): the only distinction is between long-range connections and local ones - those which reach in the immediate surround of the projecting neuron and do not travel through the white matter. Importantly, by asking whether there is any computational principle other than just associative memory operating at both long-range and local synapses (Braitenberg, 1991), Braitenberg pushes the age-old debate of whether cortical activity is more like a classic orchestra led by a conductor or more like a jazz jam session, beyond the limits of abstract information-processing models. In traditional box-and-arrows models of that kind, a box, whether it represents a specific part of the brain or not, can operate any arbitrary transformation of its input, which makes it difficult to relate it to physiological measures, and tends to leave the debate ill-defined. If at the core one is dealing solely with associative memory, instead, the issue can be approached with well-defined formal models, generating statistical insights that can be later augmented with cognitive qualifications.

Given the canonical cortical circuit (Douglas et al., 1989) as a basic wiring plan for the generic cortical plaquette, or patch, getting at the gist of how it contributes to the exchanges mediated by long-range cortico-cortical connectivity among different patches requires considering the fundamental aspects that vary, at least quantitatively, among the areas. A number of reviews (Finlay & Uchiyama, 2015; Hilgetag et al., 2022) have pointed out that several prominent features align their gradients of variation, across mammals and in particular in the human brain, along a natural cortical axis, roughly from the back to the front of the cortex. Actual observations and measurements may be incomplete or even at variance with such a sweeping generalization, but here we take it as a convenient starting point. Anatomical measures point at more spines on the basal dendrites of pyramidal cells, indicating more local synaptic contacts in temporal and especially frontal, compared to occipital cortex (Elston et al., 2001). This may support a capacity for more and/or stronger local attractor states. More linear and prompt responses to afferent inputs in posterior cortices, *e.g.* visual ones (Miller et al., 1996; Rotshtein et al., 2005), also suggest reduced local feedback relative to more anterior areas.

The rapidity of the population response to an incoming input has been related to the notion of an intrinsic *timescale* that might characterize each cortical area, and that may produce highly non-trivial effects, for example when inhibiting a particular area with TMS (Cocchi et al., 2016). The timescales measured with similar methods have been shown to differ considerably, even within individual areas (Cavanagh et al., 2020), and to define distinct cortical hierarchies, when extracted in different behavioural states, *e.g.* in response to visual white noise stimuli (Chaudhuri et al., 2015) or during free foraging (Manea et al., 2023). Thus it remains unclear whether the ambition to define a unique hierarchy of timescales can really be pursued (Gao et al., 2020), and whether they can be related to patterns of cortical lamination (Barbas & Rempel-Clower, 1997) and to biophysical parameters, including the I_h current and

others underlying firing rates and firing frequency adaptation (Chang et al., 2005). Still, in broad terms multiple timescale hierarchies do roughly align with the natural axis, from faster in the back to slower in the front of the brain, and ignoring a factor of, say, four (Gao et al., 2020) would appear to grossly overlook a basic principle of cortical organization.

Here, we ask what are the implications of major differences in *cortical parameters* for how basic associative memory mechanisms may express cortically-initiated activity. We focus on a simple differentiation between a posterior and a frontal half of the cortex, and neglect finer distinctions, *e.g.*, rostrocaudal hierarchies within prefrontal cortex (Koechlin et al., 2003; Badre, 2008) or the undoubtedly major differences within posterior cortices.

A simply differentiated Potts model

The mathematically defined model we use is based on the abstraction of a network of $\sqrt{N}$ patches of cortex (where N are all its pyramidal cells), interacting through long-range, associatively modified synapses, an abstraction close to that informing connectome research (Roe, 2019). Each patch would be a densely interconnected network of $\sqrt{N}$ pyramidal cells interacting through local synapses, also associatively modifiable according to some form of Hebbian plasticity. Such a local cortical network may operate as an autoassociative memory once it has acquired through learning a number S of attractor states. In the simplified Potts formulation adopted here, the local network realized in each patch is replaced by a Potts unit with S states, and the analysis can focus on the network of long-range effective interactions between Potts units, which are no more mediated by simple synaptic connections, rather the connections are mathematically expressed as tensors (Naim et al., 2018).

We refer to previous studies (Ryom et al., 2021) and to Appendix B for a description of the standard model and of its key parameters. Suffice here to note that while the number S of local attractor states measures the range of options available for the dynamics of a patch of cortex, the feedback coefficient w quantifies how deep those options are, *i.e.*, how strongly the patch is driven to choose one of them, and the adaptation time constant τ_2 parametrizes the time it takes for it to be eventually eased out of its current attractor.

A network of Potts units can express spontaneous behaviour when it latches, *i.e.*, it hops from a quasi-stationary pattern of activity to the next, in the absence of external input - of a query (Treves, 2005). Latching dynamics are a form of iterated associative memory retrieval; each extended activity pattern acts briefly as a global cortical attractor and, when destabilized by the rising thresholds which model firing rate adaptation, serves as a cue for the retrieval of the next pattern. Studies with brain-lesioned patients indicate, however, that there is structure in such spontaneous behaviour. In studies of mind-wandering, for example, patients with lesions to ventromedial prefrontal cortex (vmPFC) show reduced mind-wandering, and their spontaneous thoughts tend to be restricted, focused on the present and on the self, suggestive of a limited ability to project coherently into the future (Bertossi & Ciaramelli, 2016).

We then take our standard, homogeneous Potts network, differentiate it in two halves, and ask whether a structure of this type may reflect a basic differentiation between frontal and posterior cortices in the number or in the strength of their local attractor states, or in the time scale over which they operate, as expressed in differences, in the model, in the three relevant parameters, ΔS, Δw and $\Delta \tau_2$.

We assume that the two sub-networks store the same number p of memory patterns (with the same sparsity a), and that all the connections already encode these p patterns, as a result of a learning phase which is not modelled. We have seen in a previous study (Ryom &

Treves, 2023) that a differentiation ΔS has important dynamical implications during learning itself, but here we imagine learning to have already occurred. For a statistical study, we take the activity patterns to have been randomly generated with the same statistics, therefore any correlation between pattern μ and ν is random, and randomly different if calculated over each sub-network. These restrictive and implausible assumptions - they discard for example the possibility of structured associations between frontal and posterior patterns of different numerosity, statistics and internal non random correlations - are needed to derive solid quantitative conclusions at the level of network operation, and might be relaxed later in more qualitative studies.

Connectivity in the differentiated network

For the statistical analysis, carried out through computer simulations, to be informative, the structure of the network model and in particular its connectivity have to be chosen appropriately. First, each sub-network should have the same number of units (half the total) and each unit the same number of inputs, for the comparisons between different conditions to be unbiased by trivial factors. Second, each sub-network should be allowed to determine, to some extent, its own recurrent dynamics, which requires the inputs onto each unit from the two halves not to be equal in strength, which would lead to washing away any difference, effectively, at each recurrent reverberation.

We then set the connection between units i and j, in their tensorial states k and l, as

$$J_{ij}^{kl,\text{intra,inter}} = \frac{c_{ij}}{c_m a \sqrt{(1-\frac{a}{S_i})(1-\frac{a}{S_j})}} \sum_{\mu=1}^{p} \left(\delta_{\eta_i^\mu k} - \frac{a}{S_i}\right)\left(\delta_{\xi_j^\mu l} - \frac{a}{S_j}\right)(1-\delta_{k0})(1-\delta_{l0}), \qquad (1)$$

where $\{c_{ij}\}$ is a sparsity $\{0, 1\}$ matrix that ensures that Potts unit receives c_m *intra* inputs from other units in the same sub-network and also receives c_m *inter* inputs from units of the other sub-network. Note that the number of Potts states of each unit, S, may depend on which sub-network the unit belongs to.

The partially differential dynamics is obtained by setting the strength coefficients as

$$J_{ij}^{kl} = \frac{(1+\lambda)}{2} J_{ij}^{kl,\text{intra}} + \frac{(1-\lambda)}{2} J_{ij}^{kl,\text{inter}}, \tag{2}$$

where the parameter $\lambda \in [-1, 1]$ controls the relative strength of two terms. For $\lambda = 0.0$, the connectivity matrix becomes homogeneous and we cannot distinguish the two sub-networks from connectivity alone. If $\lambda = 1.0$, each sub-network is isolated from the other. For values of λ between 0 and 1, the recurrent connections within a sub-network prevail over those from the other sub-network, generating partially independent dynamics. We set $\lambda = 0.5$ as our reference value.

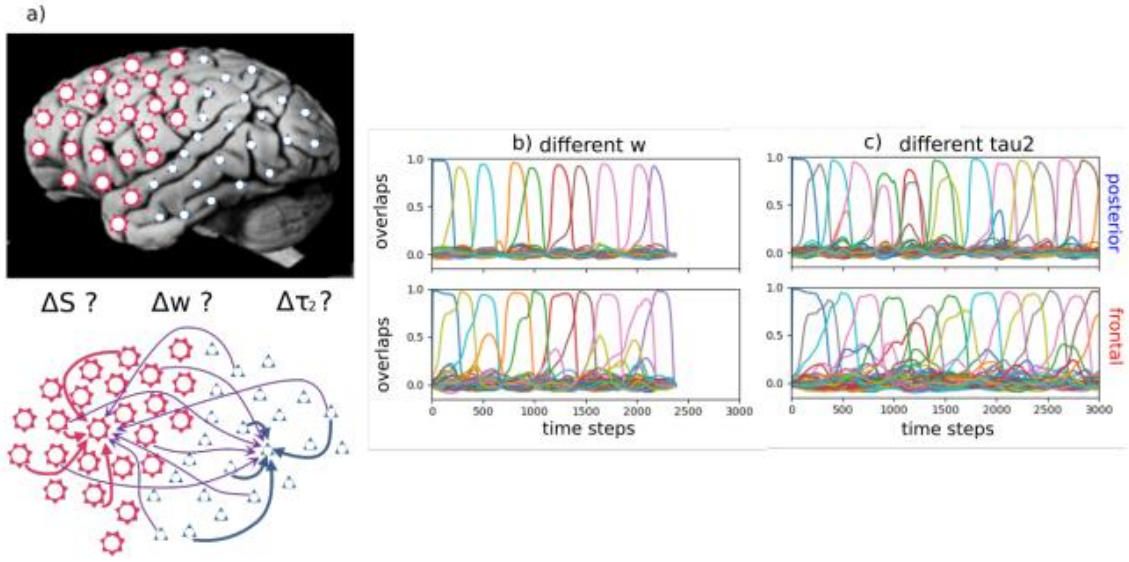

Figure 1. The differentiated network and examples of latching sequences. (a): The differentiated network is comprised of frontal and posterior halves, in each of which units receive the same number of inputs from both halves, but not of the same average strength. (b) and (c): The latching sequences – visualized by the overlaps, *i.e.*, by how close the state of the network at time t is to each memory pattern (assigned an arbitrary color) – are very similar if extracted from the posterior (upper panels) or the frontal sub-network (bottom panels). In (b), parameters are set as in Fig. 2e. In (c), parameters are set as in Fig. 3c. Close inspection reveals that in (b) the transitions in the frontal network appear to anticipate those in the posterior one, while in (c) the trend is not clear, consistent with the results described below.

Results

We assume that the attractors of the frontal network have been associated one-to-one with those of the posterior network, via Hebbian plasticity, during a learning phase, which we do not model. When there is no external stimulus, *e.g.* when modelling creative thinking and future imaging, the network can sustain latching dynamics, *i.e.* it can hop from state to state, as in Fig. 1, provided its activity is appropriately regulated by suitable thresholds, as we have reported elsewhere (Treves, 2005). Such spontaneous dynamics of the entire network might be led to a different extent by its frontal and posterior halves, depending on their characteristic parameters.

In order to quantify the relative influence of the two sub-networks on the latching sequences produced by the hybrid Potts model, we look at whether the actual occurrence of each possible transition depends on the correlations, computed separately in the frontal and posterior parts, between the two patterns before and after the transition.

For the randomly correlated patterns used here, the correlations are relatively minor, but they can be anyway quantified by two quantities, C_{as} and C_{ad} (Russo & Treves, 2012; Boboeva et al., 2018), that is, the fraction of active units in one pattern that are co-active in the other and in the same, C_{as}, or in a different state, C_{ad}. In terms of these quantities, two memory patterns are highly correlated if C_{as} is larger than average and C_{ad} is smaller than average, and we can take the difference $C_{ad} - C_{as}$ as a simple compact indicator (actually, a proxy) of the "distance" between the two patterns.

How strongly are transitions in a latching sequence driven by pattern correlations in each subnetwork? To measure this, we take the weighted average of C_{as} and C_{ad} with the weights given by latching sequences; that is, we compute

$$\langle C_{as} \rangle_T \equiv \sum_{(\mu,\nu)} t_{\mu\nu} C_{as}^{\mu\nu}, \tag{3}$$

(and analogously for $\langle C_{ad} \rangle_T$) where the sum $\sum_{(\mu,\nu)}$ runs over all possible pairs of memories and $t_{\mu\nu}$ is the normalized frequency of latching transitions for the pair μ, ν: $\sum_{(\mu,\nu)} t_{\mu\nu} = 1$. This average is compared with the "baseline" average, *e.g.*,

$$\langle C_{as} \rangle_B \equiv \frac{2}{p(p-1)} \sum_{(\mu,\nu)} C_{as}^{\mu\nu}, \tag{4}$$

independent of the transitions, where p is the number of stored memories in the network. The comparison between the two averages, $\langle C_{as(d)} \rangle_T$ and $\langle C_{as(d)} \rangle_B$, is one index of how strongly latching sequences are related to correlations between patterns in one of the two sub-networks. Second, based on the hypothesis that the frequency of transitions tends to decrease exponentially with the distance between the two patterns, as defined above, we look for the linear regression between the logarithm of the normalized transition frequency, $\log(t)$, and the proxy of the distance, $C_{ad} - C_{as}$. We first consider a case when all the macroscopic parameters are equal between the two sub-networks, while the connection parameter is set as $\lambda = 0.5$. In this case, the intra-connections (within each sub-network) are 3 times, on average, as strong as the inter-connections (between the two sub-networks), but the two halves are fully equivalent, or Not Differentiated (ND). With the appropriate parameters, in particular the feedback w, we find that the network as a whole shows robust latching and that latching sequences in each sub-network are well synchronized with each other: the two sub-networks essentially latch as one. Comparing latching dynamics in two sub-networks, we find that latching is largely driven by correlations between patterns, in either half or in both, as found previously (Russo & Treves, 2012). This can be seen, leftmost bars of Fig. 2a and Fig. 2b, by the higher value of $\langle C_{as} \rangle_T$ relative to $\langle C_{as} \rangle_B$, and vice versa for C_{ad}, in the ND case. Correlations in the two sub-networks

appear to contribute equally to determine latching sequences, as expected. This is confirmed by the similar negative slopes in the two scatterplots of Fig. 2c.

Different *S*. We now examine a case in which the two networks share the same values of all but one parameter: the number of Potts states, *S*. When the posterior network has fewer states ($S = 3$ instead of the reference value, 7), the baselines for both C_{as} and C_{ad} are shifted, above and below, respectively, but their transition-weighted values are similarly positioned, above and below the respective baselines, as in the frontal network. Also in terms of the second indicator, the scatterplot of Fig. 2d shows rather similar slopes, with only a modest quantitative "advantage" for the frontal network (in red), which can be said to lead the latching sequence somewhat more than the posterior one. One should note that, with these parameters, both sub-networks would latch if isolated.

Different *w*. In contrast to the two cases above, ND and ΔS, we see a major difference between the two sub-networks if it is the *w* parameter which is lower for the posterior network (the rightmost bars of Figs. 2a,b). In this case, it is obviously the correlation structure of the frontal patterns, not of the posterior ones, that dominates in determining latching sequences. This is also evident from the very different slopes, *k*, in the scatterplot of Fig. 2e. With the lower value $w = 0.6$ chosen for the posterior sub-network, this time it would not latch, if isolated. Note that to preserve its latching, and for it to be a clear single sequence, we would have to set *w* at almost the same value as for the frontal sub-network, unlike the case with the *S* parameter.

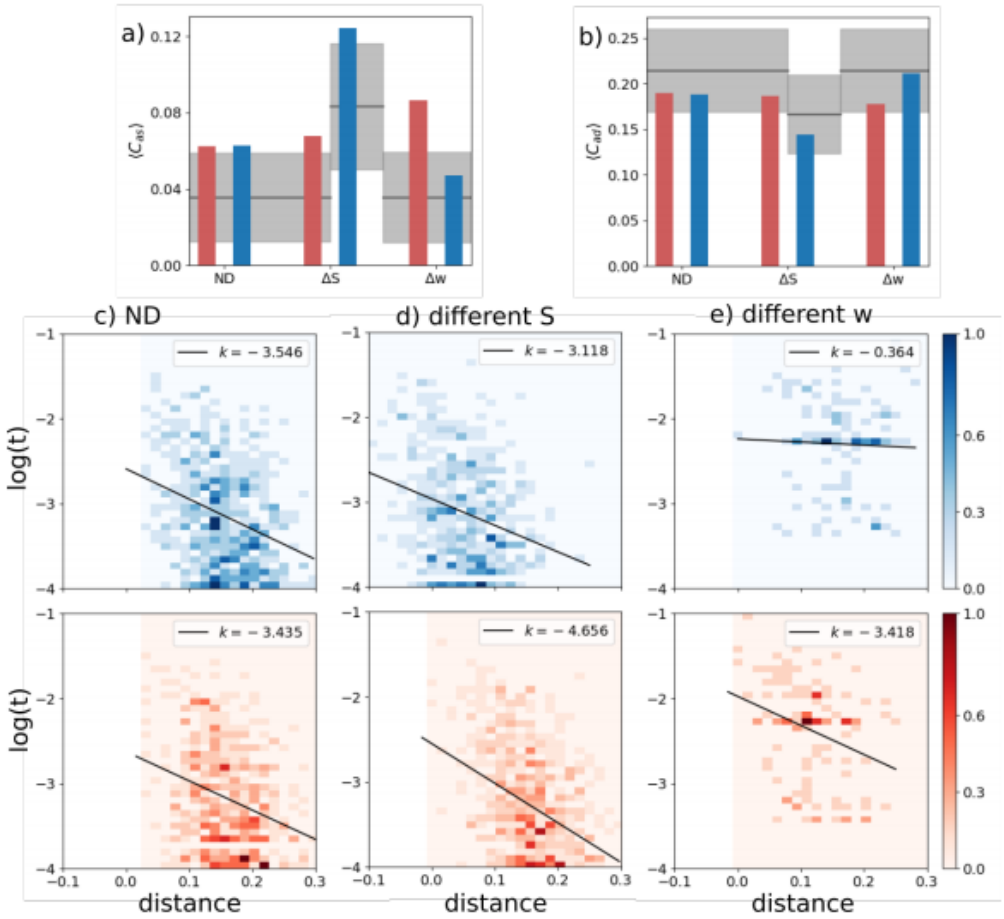

Figure 2. A latching frontal network leads a non-latching posterior network. Red indicates the frontal and blue the posterior network in this and other figures. (a) and (b). The transition-weighted averages of C_{as} and C_{ad} are compared to their baseline values for three cases: no difference between the two networks (ND, leftmost bars), a difference in S (ΔS, middle bars) and a difference in w (Δw, rightmost bars). The gray horizontal line and shaded area indicate the baseline average and its standard deviation. (c), (d) and (e) Scatterplots of (log) transition frequencies between individual patterns pairs versus their "distance", for the three conditions. The darkness of color indicates the number of pairs at each combination of abscissa and ordinate. For the ND condition, parameters are set as $w_p = w_f = 1.1$, $S_p = S_f = 7$. For the other conditions, the parameters of the frontal network are kept the same as in the ND condition, while the parameters of the posterior sub-network are set as $S_p = 3$ and $w_p = 0.6$, respectively, in (d) and (e). Note the negative values on the x-axis, particularly in panel (d)upper, due to using just a proxy of a proper distance measure, a proxy which reaches in the negative range when $S = 3$.

And/or different τ_2. We now allow the adaptation timescale, τ_2, to differ between two sub-networks. We first note that latching sequences between the two networks are remarkably well synchronized despite their different adaptation timescales (Fig. 1c). If isolated, the two sub-networks would each latch at a pace set by its own τ_2. Their synchronization thus shows that, even with this relativity weaker connectivity coupling (inter-connections 1/3 of the average strength of the intra-connections) the two halves are willing to compromise, and latch at some intermediate pace, close to the one they sustained when τ_2 was not differentiated.

Furthermore, latching sequences are affected predominantly by frontal correlations rather than posterior ones. In Fig. 3, we show two cases: the two sub-networks have two different adaptation timescales; and in the second case also different w. We see a moderate effect if τ_2 is the only parameter that differs between the two. Note that in this case the posterior sub-network, if isolated, would latch.

The effect is most pronounced if w is also lowered to w = 0.6 for the posterior sub-network, as is evident from the weak positive slope k it shows, see Fig. 3d. In this case it would not latch if isolated.

We have also inverted the τ_2 difference, making the posterior sub-network, still with a lower w, slower in terms of firing rate adaptation. In this case (not shown) latching is virtually abolished, showing that the parameter manipulations do not simply add up linearly.

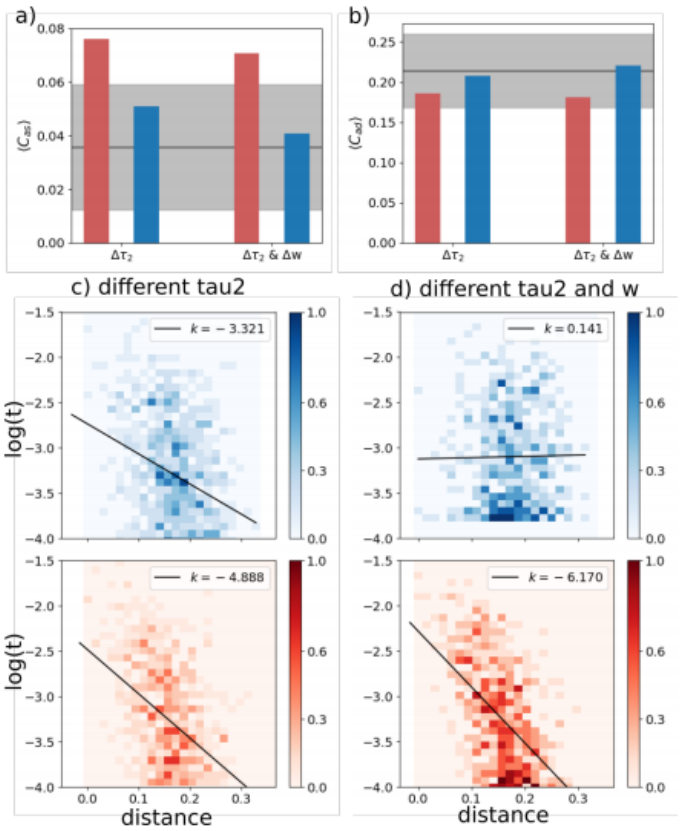

Figure 3. The frontal sub-network is even more dominant with slower adaptation. Color code and meaning are the same as in Fig. 2. (a) and (b) Transition-weighted averages of C_{as} and C_{ad} versus their baselines are shown for two conditions: only τ_2 is different and both w and τ_2 are different. In both conditions, τ_2 is 100 for the posterior network and 400 for the frontal network. In the Δw condition, w is 0.6 for the posterior network and 1.1 for the frontal network. (c) and (d) Log-transformed transition frequencies between individual patterns pairs versus their distance.

Lesioning the network

To model lesions in either sub-network, we define a procedure that still allows us to compare quantities based on the same number of inputs per unit, etc. The procedure acts only on the relative weights of the connections (through λ), which are modulated while keeping their average for each receiving unit always to 1/2. Other parameters of the network are set in such a way that the frontal sub-network leads the latching sequences and that lesions do not push the network into a no-latching phase: the self-reinforcement parameter is set as $w = 0.7$ for the posterior sub-network and $w = 1.2$ for the frontal one, while S and τ_2 are set as specified in Table 1 and thus take the same value for both sub-networks. For "healthy" networks, we use $\lambda = 0.5$ in Eq. (2), meaning the intra-connections (within the frontal and within the posterior half) are 3 times, on average, as strong as the inter-connections (between frontal and posterior halves). For lesioned networks, we use smaller values of λ than 0.5 for their input connections: the smaller the value is, the stronger the lesion is. So, for example, a frontal lesion with $\lambda = 0.2$ implies that its recurrent weights are weighted by a factor 0.6 (instead of 0.75) and the weights from the posterior sub-network by a factor 0.4 (rather than 0.25), *i.e.* the internal weights are only 1.5 times those of the interconnections. The posterior sub-network in this case has the same weights as the control case.

We then quantify the effect of the lesions with the slopes in the scatterplots as before, but also with an entropy measure. The entropy at position z in a latching sequence measures the variability of transitions encountered at that position, across all sequences with the same starting point. It is computed as

$$S(z) = \left\langle -\sum_{\mu \neq \nu} P_\gamma^{\mu\nu}(z) \log_2 P_\gamma^{\mu\nu}(z) \right\rangle_\gamma, \tag{5}$$

where $P^{\mu\nu}{}_{\gamma}(z)$ is the joint probability of having two patterns μ and ν at two consecutive positions z and $z + 1$ relative to the cued pattern γ in a latching sequence, and $\langle\cdot\rangle_{\gamma}$ means that we average the entropy across all the p patterns that are used as a cue. Note that if all transitions were incurred equally, asymptotically for large z, the entropy would reach its maximum value $S_{\infty} = \log_2[p(p-1)]$ (with p patterns stored in memory and available for latching). Therefore $\exp\{[S(z) - S_{\infty}]\ln(2)\}$ is an effective measure of the fraction of all possible transitions that the network has explored at position z, on average.

In terms of the slopes in the scatterplots, we see that posterior lesions do not have a major effect, while frontal lesions reduce the relation between the probability of individual transitions and the correlation between the two patterns, particularly in the frontal sub-network where it was strong in the "healthy" case (Fig. 4).

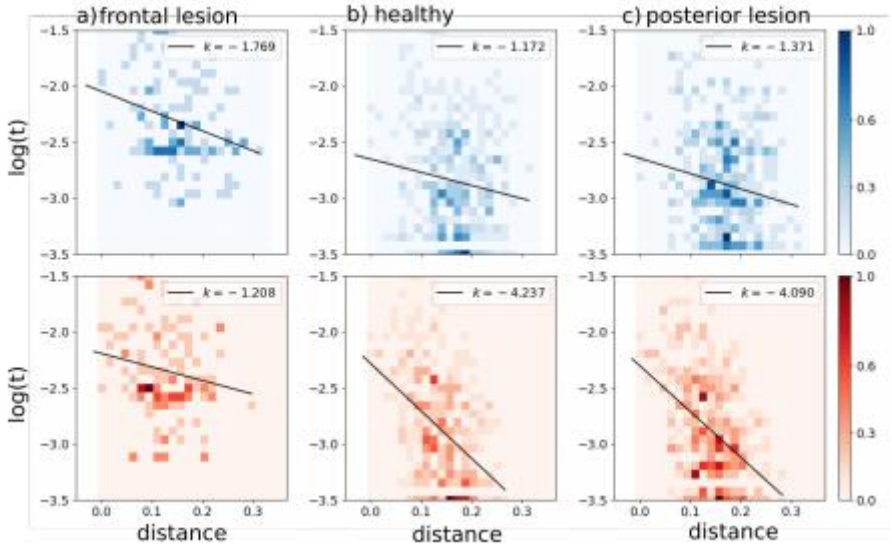

Figure 4. Correlations between transition frequency and pattern distance are shown for a network with frontal lesions (a), for a healthy network (b) and for a network with posterior lesions (c). Lesions are modelled by setting $\lambda = 0.2$ (see main text). The self-reinforcement parameter is set as $w = 1.2$ for the frontal sub-network and $w = 0.7$ for the posterior one.

In terms of entropy, we see that lesions in the posterior sub-network do not affect the entropy curve, relative to that for the healthy network (Fig. 5). Lesions in the frontal sub-

network, however, tend to restrict the sequences to a limited set of transitions, leading to a marked reduction in the fraction of possibilities explored by the lesioned network.

Simulated frontal lesions, therefore, produce in our model two effects that, while not opposite, are not fully congruent either. The first, manifested in the reduced slope of Fig. 4a, is suggestive of a loss of coherence in individual transitions between brain states; the second, seen in the limited entropy of Fig. 5, indicates a restriction in the space spanned by the trajectories of spontaneous thought. To reconcile the two outcomes, we have to conclude that while less dependent on the similarity between the two patterns, or states, individual transitions are not really random, and some become in the lesioned network much more frequent than others, gradually veering from creative towards obsessive (or perseverative) thought.

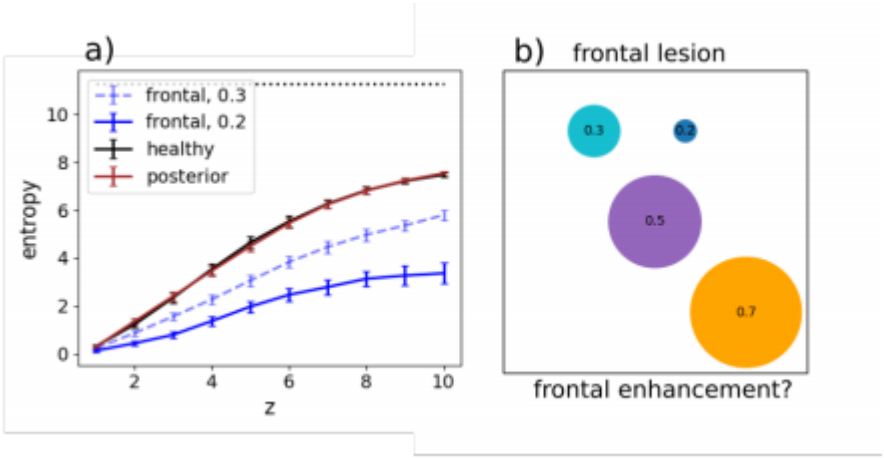

Figure 5. (a) The entropy $S(z)$ and its standard error of the mean are shown for healthy (black), frontal-lesioned (blue) and posterior-lesioned (red) networks. Lesions are implemented by setting $\lambda = 0.2$ for solid curves, whereas the dashed blue curve is for a milder lesion in the frontal network ($\lambda = 0.3$). The black horizontal line indicates the asymptotic entropy value for a completely random sequence generated from a set of $p = 50$ patterns. The self-reinforcement parameter is set as $w = 1.2$ for the frontal network and $w = 0.7$ for the posterior network. (b) A schematic view of the diversity of transitions expressed by latching sequences. Circles are centered around an arbitrary position, while their areas extend over a fraction $2^{S(10)-S_\infty}$ of the area of the square (which would correspond to an even exploration of all possible transitions, asymptotically). The large orange circle is obtained by setting $\lambda = 0.7$, thus modelling a sort of cognitive frontal enhancement, perhaps obtained with psychoactive substances.

Discussion

Simulating our model provides some insight about the conditions that may enable frontal cortices to determine the sequence of states in spontaneous thought dynamics. It is important, in assessing the computational findings, to distinguish what has gone into defining the model from what the model gives out in return. For example, much cognitive neuroscience research has been devoted to understanding the process of segmenting our ongoing experience into separate sub-events, or event segmentation (Kurby & Zacks, 2008). Baldassano and colleagues (2017) have recently demonstrated how brain activity within sub-events resembles temporarily stable activity patterns, dubbed "neural states" (Geerligs et al., 2022), which may be identified with those long posited to occur in the cortex of primates (Abeles et al., 1995) and other species (Jones et al., 2007), from analyses of single-unit activity. This notion is conceptually similar to the Potts states in a latching sequence, but finding evidence that a continuous input flow is segmented into discrete or quasi-discrete states in the brain is a major achievement, whereas in the Potts network it is a straightforward outcome of the ingredients used to define the model in the first place. Interestingly, these neural states were found to occur on different timescales across regions, with more but short-lasting transitions in low-level (posterior) sensory cortices and fewer but longer-lasting transitions in higher-level (frontal/parietal) regions. Strikingly, for some of the higher order brain regions, neural state transitions appeared to overlap with behavioural measures of event boundary perception (Baldassano et al., 2018).

In our study, the central question is which portion of the differentiated model network controls the sequence of discrete event states. We have seen that three types of differentiation, each capturing some aspect of caudo-rostral cortical variation, bias sequence control towards the "frontal" half of the network, albeit with different effectiveness. A comparison across the

three types of differentiation is inherently ill-defined and somewhat arbitrary, because ΔS, Δw and $\Delta \tau_2$ are all measured on different scales, but it is apparent that the first type has a much milder effect than the second, and the third is somewhere in between. The major effect seen with Δw is likely due to the posterior network being unable to latch on its own, with the lower w value we have used. The lower S and τ_2 values do not have much of an effect on latching per se. The three types of differentiation are of course not mutually exclusive, and it is plausible that in the real brain, if the model makes sense, their effect would be cumulative. They do not appear to add up linearly, though: we have mentioned that inverting the τ_2 difference with respect to the w difference (*i.e.*, making firing rate adaptation faster in the frontal sub-network) tends to abolish latching altogether, rather than reduce the frontal advantage in leading it.

A limitation of our study is that to compare the sub-networks on an even footing we have considered an artificial scenario in which activity patterns are only randomly correlated, and also there are p in each half network and they have been paired one-to-one during learning. Obviously in this scenario there is no benefit whatsoever if the network follows a frontally- rather than a posteriorly-generated sequence: they are equivalent, and both devoid of content. It will be therefore important, in future work, to understand whether the insights derived under these assumptions are applicable also to more plausible conditions, in which the frontal and posterior patterns are not paired one-to-one, and can take distinct roles, for example along the lines of the classic operator/filler (also denoted as role/filler) distinction (Do & Hasselmo, 2021). In this more complex scenario, the frontal patterns, if they have to serve as operators, would "take" or be paired in certain cases to a single filler and in others to multiple fillers (and possibly to other operators, in a hierarchical scheme); but even if just to one, it would be one among several options, so the pairing scheme in long-term-memory would be considerably more complex than the one considered here.

A relevant cognitive construct we mention, only partially overlapping with that of operator, is that of a temporally-oriented schema. A schema is a regularity extracted from multiple experience, in which B follows A and is then followed by C, although the particular instantiation of A, B and C will be different every time (Gilboa & Marlatte, 2017). Note that to be implemented in our network, the skeleton of the ABC representation would have to stay activated while the specific filling items A, B and C are specified, in succession, in the posterior cortex. Alternatively, ABC could be conceptualized as a short tight latching sequence. Clearly, more attention has to be paid to the possibility of formalizing these constructs in a future well-defined network model.

Mind wandering and creativity

Within its present limitations, still our approach may offer insights relevant to the dynamics of state transitions in spontaneous cognition, such as those underlying mind wandering. Mind wandering occurs when attention drifts away from ongoing activities and towards our inner world, focusing for example on memories, thoughts, plans, which typically follow one another in a rapid, unconstrained fashion (Smallwood & Schooler, 2015; Christoff et al., 2016). The dynamics governing the flow of thoughts can indeed be described as latching (see also Ciaramelli & Treves, 2019).

Mind wandering is known to engage the Default Mode Network (DMN), a set of interconnected brain regions, spanning from posterior, temporal, and frontal cortices (Buckner et al., 2008; Stawarczyk et al., 2011; Smallwood, 2013; Andrews-Hanna et al., 2014; Raichle, 2015; Christoff et al., 2016), underlying introspection and spontaneous (endogenously triggered) cognition. Ciaramelli and Treves (2019) and McCormick et al. (2018a) have proposed that the prefrontal cortex, especially in its ventral-medial sectors (vmPFC) might

support the initiation (internal triggering) of mind-wandering events. Indeed, recent MEG findings show that activity in the vmPFC precedes (presumably drives) hippocampal activity during (voluntary) scene construction and autobiographical memory retrieval (Barry et al., 2019; see also Monk et al., 2020; 2021), and this region may play a similar role during spontaneous cognition. Indeed, damage (Bertossi & Ciaramelli, 2016; Philippi et al., 2021) or inhibition (Bertossi et al., 2017b; Giacometti Giordani et al., 2023) of the vmPFC (but not the hippocampus; McCormick et al., 2018a) reduce the frequency of mind-wandering.

On one view, vmPFC initiates event construction by activating schemata (about the self, or common events) that help collect relevant details that the hippocampus then binds in coherent, envisioned scenes (Ciaramelli et al., 2019; see also Benoit et al., 2014; Moscovitch et al., 2016; Rolls, 2022). Consistent with the schema hypothesis, vmPFC (but not hippocampal) patients are particularly impaired in event construction when the task benefits from the activation of the self schema (Verfaellie et al., 2019; Stendardi et al., 2021), and are not impaired when the need for self-initiation is minimized (De Luca et al., 2019). vmPFC may also govern schema-congruent transitions between successive scenes of constructed events based on event schemata (scripts) (Stawarczyk et al., 2011; Lieberman et al., 2019), which may explain why vmPFC patients are particularly poor at simulating extended events as opposed to single moments selected from events (Kurczek et al., 2015; Bertossi & Ciaramelli, 2016). The results from our computational simulations accord with and complement this view. Lesioning the frontal (but not the posterior) sector of the network led to more random state transitions, less dependent on the correlation between patterns, and also led to shorter-lasting sequences, that fade out after fewer state transitions. This pattern of findings is expected if transitions in thought states were not guided by schematic knowledge, making them less coherent in content and self-exhausting.

A second effect we observed is a reduced entropy following lesions in the frontal (but not posterior) half of the network, which indicates that the trajectories of state transitions were confined in a limited space, as if mind wandering lost its 'wandering' nature to become more constrained, with recurring thoughts characteristic of the perseverative responses long observed in prefrontal patients; suggesting that vmPFC patients, in addition to an impaired activation of relevant schemata, also fail in flexibly deactivating current but no longer relevant ones (Gilboa & Marlatte, 2017).

The most characteristic memory deficit following vmPFC damage is confabulation, the spontaneous production of false memories. Confabulations often involve an inability to inhibit previously reinforced memory traces (Schnider, 2003). For example, confabulators can falsely endorse personal events as true because these were true in the past (*e.g.*, that they just played football while in fact they used to play football during childhood). If presented with modified versions of famous fairy tales to study, confabulators tend to revert to the original versions of the stories in a later recall phase (Attali et al., 2009). Similarly, during navigation, confabulators may get lost because they head to locations they have attended frequently in the past, instead of the currently specified goal destination (Ciaramelli, 2008).

The inability to flexibly switch between relevant time schemata and memory traces has been linked to reduced future thinking and reduced generation of novel scenarios in prefrontal patients (de Vito et al., 2012; see also Bertossi & Ciaramelli, 2016), who admitted they found themselves bound to recast past memories while trying to imagine future events. More in general, prefrontal lesions impair creativity. There is interaction between the DMN and the fronto-parietal control network while generating (DMN) and revising (fronto-parietal network) creative ideas (Beaty et al., 2014; Bendetowicz et al., 2017). Bendetowicz et al. (2017) found that damage to the right medial prefrontal regions of the DMN affected the ability to generate remote ideas, whereas damage to left rostrolateral prefrontal region of the fronto-parietal

control network spared the ability to generate remote ideas but impaired the ability to appropriately combine them.

Note, however, that the originality associated with creative ideas can be conceived as disrupting the automatic progression from a thought to the one most correlated to it. Fan et al. (2023) had participants perform a creative writing task, and indeed found the semantic distance between adjacent sentences to be positively correlated with the story originality. Also, semantic distance was predicted by connectivity features of the salience network (*e.g.*, the insula and anterior cingulate cortex) and the DMN. Green et al. (2006) have also reported a putative role of mPFC (BA 9/10) in connecting semantically distant concepts during abstract relational integration. In a following study (Green et al., 2010), mPFC activity was found to vary monotonically with increasing semantic distance between abstract concepts, even when controlling for task difficulty. Indeed, preliminary evidence from patients with vmPFC lesions is indicative of a greater global semantic coherence in speech compared to healthy participants (Stendardi et al., in preparation). These results align with our finding that a lesion of the frontal component of the network produces a reduction in entropy, making latching dynamics "less creative"; but not, *prima facie*, with the reduced slope in Fig.4a, which indicates that the lesion would produce more random transitions, frequent also among distant patterns. The apparent contradiction can be reconciled by noting that, as seen above, *individual* random transitions can still result in reduced entropy, if they tend to recur perseveratively within a sequence; and also that semantic coherence may reflect pattern correlation in posterior rather than frontal cortices, whereas it is logical/syntactic consequentiality that is expected to be impaired by random frontal transitions. In fact, in our model lesion, the decreased slope in the frontal sub-network seen in Fig.4a (more random transitions) is accompanied by a slightly increased slope, suggestive of more semantic coherence, posteriorly.

Clearly, a major refinement of our approach is required, before these suggestions can be taken seriously, and articulated in a more nuanced and anatomy-informed view of how operating along the time dimension may be coordinated across cortical areas.

Conclusion

We opened this thesis by discussing the complexity and intriguing obscureness of the human brain, and particularly of one of the largest regions of the frontal cortex: the ventromedial prefrontal cortex, vmPFC. We have considered how the diversity of vmPFC anatomical components and the intricacies of its connections reflect in its involvement in a plethora of cognitive functions. The vmPFC stands in interaction between a myriad of cortical and subcortical regions, and is one of the central hubs of the most fascinating, yet still puzzling, systems of the human brain: the Default Mode Network (DMN). The heterogeneity of vmPFC functions has been perfectly captured by Gage & Baars (2018) when stating that "*vmPFC is an integrative hub for emotional, sensory, social, memory, and self-related information processing*" (see also Roy et al., 2012). Here, we might note, it seems an integrative hub for basically everything the brain does. The quest to solve the conundrum of what, essentially, is the basic function of vmPFC has endured for more than a century, since the moment an iron rod shot through the skull of an impressively (un)lucky stagecoach driver (Harlow, 1848, 1993). Such a query has likely been moved by an idea shared by countless neuroscientists: functional specialization. While it is undeniable that brain regions are connected and entangled in about 100 trillion different paths of communication (Zimmer, 2011), it is also true that specific neurons, specific neural populations, specific *regions*, do something that other regions simply don't. Neurons in the frontal cortex are unambiguously different from neurons in the occipital cortex, and whilst we may think that the concept of functional specialization only applies to sensory areas, recent evidence has started to prove its suitability even for high-level, abstract cognitive functions (Pestrides, 1991; Poldrack et al., 1999; Gilbert et al., 2006, 2010; Beckmann et al., 2009; Kanwisher, 2010; Zysset et al., 2013). So, what essential

function can possibly bridge together all the cognitive domains that vmPFC seems to be implicated in? What theory can explain the process to which vmPFC activity boils down to?

The schema theory has been one of the attempts to answer such a question, and here, our results shall be interpreted within its framework (Gilboa, 2004, 2010; Gilboa & Marlatte, 2017; Gilboa & Moscovitch, 2017).

The schema theory posits that vmPFC plays the critical role of activating schematic knowledge in the neocortex, whilst also suppressing schemata that are irrelevant in the current environment (Gilboa & Moscovitch, 2002; Gilboa et al., 2006, 2009; Hebscher & Gilboa, 2016). Our finding of a degraded self-schema in patients with vmPFC lesions strongly aligns with this proposal. A damage to vmPFC might thus hinder the capacity to correctly reinstate the self-schema, which, in normal condition, would give rise to a mnemonic advantage for self-related items (see Chapter 1), and to a high stability of self-related judgments over time (see Chapter 2). Also, the pattern of performance displayed by vmPFC patients in Study 2 seems to suggest a behaviour reminiscent of confabulation, but only within the self-schema. However, we note that the patients tested here were not spontaneous confabulators. Hence, we speculate that the self-schema could be one of the first and preferential schematic structures to be affected by vmPFC lesions, likely because it is overrepresented and hyper-connected with other kinds of memory and neural representations (Wagner et al., 2012, 2019; D'Argembeau, 2013; Lieberman et al., 2019). Therefore, it constitutes a preferential substrate for confabulatory content, something that aligns with the observation that the contents of confabulation often relate to a patient's personal life history (Gilboa & Moscovitch, 2002; Gilboa, 2004, 2010; Fotopoulou , 2010; Gilboa & Verfaellie , 2010; Hebscher & Gilboa, 2016). This proposal is also consistent with the study of Gilboa et al. (2006), in which vmPFC patients falsely endorsed statements that were blatantly inconsistent with their self-schema.

In Study 3 we extended the investigation onto the reinstatement and instantiation of schemata, specifically event-scripts, finding that vmPFC patients do indeed display difficulties in schema reinstatement, since they reinstate a broad and nebulous script. In turn, a faulty script reinstatement gives rise to a schema instantiation that rests on a fragile and fragmentary schematic structure, something that leaves vmPFC patients incapable of constructing rich and detailed events (Bertossi et al., 2016a,b, 2017a; McCormick et al., 2018a). According to the model proposed by Ciaramelli et al. (2019), in such a situation, schematic structures would only be partially active, and in a degraded form, which would force the hippocampus to create a rudimentary sketch, but devoid of schematic elements and unconstrained by neocortical monitoring. This monitoring, achieved through neocortical-hippocampal feedback loops, would allow for the dynamic unfolding of an event, or of spontaneous, internally generated cognition. Indeed, vmPFC patients tend to construct events that mostly resemble a momentary snapshot, rather than extended events that unfold over time (Kurczek et al., 2015; McCormick et al., 2018a; Ciaramelli & Treves, 2019; Ciaramelli et al., 2019).

This same lack of monitoring by frontal cortices onto posterior parts of the brain is clearly reproduced in Study 4 by simulating a frontal lesion in a hybrid Potts neural network. Here, we sought to determine whether modelling frontal and posterior cortices in a biologically plausible fashion (Miller et al., 1996; Elston et al., 2001; Rotshtein et al., 2005; Cocchi et al., 2016; Cavanagh et al., 2020) could mimic frontal functions (and *dysfunctions*) observed in experimental practice. We observed the frontal sub-network, when functional, to drive and bias the latching dynamics of the posterior sub-network. We propose that, in the real brain, this is achieved by means of its *schematic influence*. Indeed, in an auto-associative network, the latching dynamics are governed by pattern correlations, namely, what causes the jumping from a particular pattern to the next is essentially the correlations amongst them. This, in fact, is precisely what happens in the undifferentiated networks (*i.e.* without any differences in cortical

parameters). However, when we model a frontal and a posterior sub-division according to biological parameters, we observe that the frontal cortex drives not only its own latching, but also the posterior one. In our view, this is achieved by vmPFC reinstating a schema that would then bias posterior activity, which strongly resonates with MEG and EEG findings exploring the temporal dynamics of frontal and posterior interactions during autobiographical memory recollection, schema reinstatement and instantiation, and scene construction (Gilboa & Moscovitch, 2017; Barry et al., 2019; Monk et al., 2020, 2021; McCormick et al., 2020; Giuliano et al., 2021). This schematic influence on posterior cortices is abolished when we simulate a frontal lesion in the network, something that resembles the deficits we observed in vmPFC patients in Study 3, and their apparent lack of schematic influences in tasks evaluating a multitude of cognitive domains (Koscik & Tranel; 2012; Warren et al., 2014; Spalding et al., 2015, 2018; Gilboa & Moscovitch, 2017; Cameron et al., 2018; Giuliano et al., 2021).

Future research is required to establish whether vmPFC might also be responsible for supporting fine details of other types of schematic knowledge, such as scene schemata and the self-schema. Also, the way vmPFC patients are capable to preserve certain aspects of schematic knowledge remains unclear. Additionally, the exact differences between schematic and semantic processing are not always well defined: future studies are needed to determine whether semantic and schematic cognition are to be defined as separable constructs, or could be placed along a continuum of increasing complexity (see Giuliano et al., 2021). Moreover, if we wish to shed light onto the intricate multitude of cortical interactions, future investigations shall need to attempt to model more precise and nuanced anatomical sub-divisions in hybrid neural networks.

In summary, our results speak in favour of a pivotal role of vmPFC in schema-mediated cognition, aligning with the proposal that the primal role of vmPFC is the activation of schemata in neocortex. vmPFC lesions provoke a degradation of the self-schema and of finer

aspects of event schemata, that are reinstated in an incomplete and nebulous form, something that compromises their capacity to bias incoming information and support self-generated cognition. The same dynamics can be observed in a hybrid neural network, in which frontal and posterior cortices are modelled in a biologically plausible fashion.

References

Abeles, M., Bergman, H., Gat, I., et al. (1995). Cortical activity flips among quasi-stationary states. *Proceedings of the National Academy of Sciences*, **92**, 8616–20

Abrahams, S., Morris, R.G., Polkey, C.E., et al. (1999). Hippocampal Involvement in Spatial and Working Memory: A Structural MRI Analysis of Patients with Unilateral Mesial Temporal Lobe Sclerosis. *Brain and Cognition*, **41**, 39–65

Ackerly, S.S., Benton, A.L. (1948). Report of case of bilateral frontal lobe defect. *Research Publications-Association for Research in Nervous and Mental Disease*, **27**, 479–504

Addis, D.R. (2020). Mental Time Travel? A Neurocognitive Model of Event Simulation. *Review of Philosophy and Psychology*, **11**, 233–59

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**, 1363–77

Adolphs, R. (1999). Social cognition and the human brain. *Trends in cognitive sciences*, **3**, 469–79

Alexander, L., Wood, C.M., Roberts, A.C. (2023). The ventromedial prefrontal cortex and emotion regulation: lost in translation? *The Journal of Physiology*, **601**, 37–50

Alicke, M.D., Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European review of social psychology*, **20**, 1–48

Allain, P., Gall, D.L., Etcharry-Bouyx, F., et al. (2001). Influence of Centrality and Distinctiveness of Actions on Script Sorting and Ordering in Patients With Frontal Lobe Lesions. *Journal of Clinical and Experimental Neuropsychology*, **23**, 465–83

Allain, P., Le Gall, D., Etcharry-Bouyx, F., et al. (1999). Mental representation of knowledge following frontal-lobe lesion: Dissociations on tasks using scripts. *Journal of Clinical and Experimental Neuropsychology*, **21**, 643–65

Aminoff, E.M., Kveraga, K., Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in cognitive sciences*, **17**, 379–90

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews neuroscience*, **7**, 268–77

Anderson, N.H. (1968). Likableness ratings of 555 personality-trait words. *Journal of personality and social psychology*, **9**, 272

Anderson, R.C. (2018). Role of the reader's schema in comprehension, learning, and memory. In: *Theoretical models and processes of literacy*. Routledge, p. 136–45.

Anderson, R.C. (1978). Schema-directed processes in language comprehension. In: *Cognitive psychology and instruction*. Springer, p. 67–82.

Anderson, R.C., Pearson, P.D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of reading research*, **1**, 255–91

Anderson, R.C., Pichert, J.W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of verbal learning and verbal behavior*, **17**, 1–12

Anderson, S.W., Barrash, J., Bechara, A., et al. (2006). Impairments of emotion and real-world complex behavior following childhood-or adult-onset damage to ventromedial prefrontal cortex. *Journal of the International Neuropsychological Society*, **12**, 224–35

Andreasen, N.C., O'Leary, D.S., Cizadlo, T., et al. (1995). Remembering the past: two facets of episodic memory explored with positron emission tomography. *The American journal of psychiatry*, **152**, 1576–85

Andrewes, D.G., Jenkins, L.M. (2019). The Role of the Amygdala and the Ventromedial Prefrontal Cortex in Emotional Regulation: Implications for Post-traumatic Stress Disorder. *Neuropsychology Review*, **29**, 220–43

Andrews-Hanna, J.R. (2012). The brain's default network and its adaptive role in internal mentation. *The Neuroscientist*, **18**, 251–70

Andrews-Hanna, J.R., Kaiser, R.H., Turner, A.E., et al. (2013). A penny for your thoughts: dimensions of self-generated thought content and relationships with individual differences in emotional wellbeing. *Frontiers in psychology*, **4**, 900

Andrews-Hanna, J.R., Reidler, J.S., Huang, C., et al. (2010a). Evidence for the Default Network's Role in Spontaneous Cognition. *Journal of Neurophysiology*, **104**, 322–35

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., et al. (2010b). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, **65**, 550–62

Andrews-Hanna, J.R., Smallwood, J., Spreng, R.N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, **1316**, 29–52

Antony, J.W., Ferreira, C.S., Norman, K.A., et al. (2017). Retrieval as a Fast Route to Memory Consolidation. *Trends in Cognitive Sciences*, **21**, 573–76

Arkes, H.R., Freedman, M.R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition*, **12**, 84–89

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of personality and social psychology*, **60**(2), 241.

Arzy, S., Bick, A., Blanke, O. (2009). Mental time in amnesia: evidence from bilateral medial temporal damage before and after recovery. *Cognitive neuropsychology*, **26**, 503–10

Atance, C.M., O'Neill, D.K. (2001). Episodic future thinking. *Trends in cognitive sciences*, **5**, 533–39

Atance, C.M., O'Neill, D.K. (2005). The emergence of episodic future thinking in humans. *Learning and motivation*, **36**, 126–44

Attali, E., De Anna, F., Dubois, B., et al. (2009). Confabulation in Alzheimer's disease: poor encoding and retrieval of over-learned information. *Brain*, **132**, 204–12

Audrain, S., McAndrews, M.P. (2022). Schemas provide a scaffold for neocortical integration of new memories over time. *Nature Communications*, **13**, 5795

Ayala, F.J. (1987). Biological Reductionism. In: F. E. Yates, A. Garfinkel, D. O. Walter, et al. (eds). *Self-Organizing Systems*. Boston, MA: Springer US, p. 315–24.

Aziz-Zadeh, L., Liew, S.-L., Dandekar, F. (2013). Exploring the neural correlates of visual creativity. *Social cognitive and affective neuroscience*, **8**, 475–80

Baddeley, A., Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and cognition*, **7**, 212–30

Baddeley, A.D., Kopelman, M.D., Wilson, B.A. (2003). *The handbook of memory disorders*. John Wiley & Sons.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences*, **12**, 193–200

Bahk, Y.-C., Choi, K.-H. (2018). The relationship between autobiographical memory, cognition, and emotion in older adults: A review. *Aging, Neuropsychology, and Cognition*, **25**, 874–92

Baird, B., Smallwood, J., Mrazek, M.D., et al. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological science*, **23**, 1117–22

Baird, B., Smallwood, J., Schooler, J.W. (2011). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and cognition*, **20**, 1604–11

Bakker, I., Takashima, A., Van Hell, J.G., et al. (2015). Changes in Theta and Beta Oscillations as Signatures of Novel Word Consolidation. *Journal of Cognitive Neuroscience*, **27**, 1286–97

Baldassano, C., Chen, J., Zadbood, A., et al. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, **95**, 709-721.e5

Baldassano, C., Hasson, U., Norman, K.A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *The Journal of Neuroscience*, **38**, 9689–99

Bar, M., Aminoff, E., Mason, M., et al. (2007). The units of thought. *Hippocampus*, **17**, 420–28

Baraduc, P., Duhamel, J.-R., Wirth, S. (2019). Schema cells in the macaque hippocampus. *Science*, **363**, 635–39

Barbas, H. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cerebral Cortex*, **7**, 635–46

Barbas, H., Pandya, D.N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, **286**, 353–75

Bargh, J.A., Bond, R.N., Lombardi, W.J., et al. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality and Social Psychology*, **50**, 869

Baron, S.G., Gobbini, M.I., Engell, A.D., et al. (2011). Amygdala and dorsomedial prefrontal cortex responses to appearance-based and behavior-based person impressions. *Social Cognitive and Affective Neuroscience*, **6**, 572–81

Baron-Cohen, S., Ring, H., Moriarty, J., et al. (1994). Recognition of mental state terms. *British Journal of Psychiatry*, **165**, 640–49

Barrash, J., Asp, E., Markon, K., et al. (2011). Dimensions of personality disturbance after focal brain damage: Investigation with the Iowa Scales of Personality Change. *Journal of clinical and experimental neuropsychology*, **33**, 833–52

Barrash, J., Bruss, J., Anderson, S.W., et al. (2022). Lesions in different prefrontal sectors are associated with different types of acquired personality disturbances. *Cortex*, **147**, 169–84

Barrash, J., Tranel, D., Anderson, S.W. (2000). Acquired Personality Disturbances Associated With Bilateral Damage to the Ventromedial Prefrontal Region. *Developmental Neuropsychology*, **18**, 355–81

Barrios, V., Kwan, V.S.Y., Ganis, G., et al. (2008). Elucidating the neural correlates of egoistic and moralistic self-enhancement. *Consciousness and Cognition*, **17**, 451–56

Barron, H.C., Garvert, M.M., Behrens, T.E. (2015). Reassessing VMPFC: full of confidence? *Nature neuroscience*, **18**, 1064–66

Barry, D.N., Barnes, G.R., Clark, I.A., et al. (2019). The Neural Dynamics of Novel Scene Imagery. *The Journal of Neuroscience*, **39**, 4375–86

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge university press.

Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage*, **76**, 412–27

Bateman, J.R., Ferguson, M.A., Anderson, C.A., et al. (2023). Network localization of spontaneous confabulation. *The Journal of Neuropsychiatry and Clinical Neurosciences*, appi. neuropsych. 20220160

Bates, D., Mächler, M., Bolker, B., et al. (2014). Fitting Linear Mixed-Effects Models using lme4

Bathelt, J., Geurts, H.M. (2021). Difference in default mode network subsystems in autism across childhood and adolescence. *Autism*, **25**, 556–65

Battaglia, S., Garofalo, S., di Pellegrino, G., et al. (2020). Revaluing the role of vmPFC in the acquisition of Pavlovian threat conditioning in humans. *Journal of Neuroscience*, **40**, 8491–8500

Beaty, R.E. (2015). The neuroscience of musical improvisation. *Neuroscience & Biobehavioral Reviews*, **51**, 108–17

Beaty, R.E., Benedek, M., Wilkins, R.W., et al. (2014). Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia*, **64**, 92–98

Bechara, A., Damasio, H., Damasio, A.R., et al. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *Journal of neuroscience*, **19**, 5473–81

Beckmann, M., Johansen-Berg, H., Rushworth, M.F.S. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. *The Journal of Neuroscience*, **29**, 1175–90

Beer, J.S., Heerey, E.A., Keltner, D., et al. (2003). The regulatory function of self-conscious emotion: insights from patients with orbitofrontal damage. *Journal of personality and social psychology*, **85**, 594

Beer, J.S., John, O.P., Scabini, D., et al. (2006). Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. *Journal of cognitive neuroscience*, **18**, 871–79

Bein, O., Reggev, N., Maril, A. (2014). Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia*, **64**, 320–30

Bellana, B., Liu, Z. -X., Diamond, N.B., et al. (2017). Similarities and differences in the default mode network across rest, retrieval, and future imagining. *Human Brain Mapping*, **38**, 1155–71

Bem, D.J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological review*, **74**, 183

Bem, D.J. (1972). Self-perception theory. In: *Advances in experimental social psychology*. Elsevier, p. 1–62.

Bendetowicz, D., Urbanski, M., Aichelburg, C., et al. (2017). Brain morphometry predicts individual creative potential and the ability to combine remote ideas. *Cortex*, **86**, 216–29

Benedek, M., Beaty, R.E., Schacter, D.L., et al. (2023). The role of memory in creative ideation. *Nature Reviews Psychology*, **2**, 246–57

Benoit, R.G., Szpunar, K.K., Schacter, D.L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences*, **111**, 16550–55

Benson, D.F., Djenderedjian, A., Miller, B.L., et al. (1996). Neural basis of confabulation. *Neurology*, **46**, 1239–1239

Berboth, S., Morawetz, C. (2021). Amygdala-prefrontal connectivity during emotion regulation: A meta-analysis of psychophysiological interactions. *Neuropsychologia*, **153**, 107767

Berkers, R.M.W.J., Van Der Linden, M., De Almeida, R.F., et al. (2017). Transient medial prefrontal perturbation reduces false memory formation. *Cortex*, **88**, 42–52

Berlucchi, G., Aglioti, S.M. (2010). The body in the brain revisited. *Experimental brain research*, **200**, 25–35

Berlyne, N. (1972). Confabulation. *The British Journal of Psychiatry*, **120**, 31–39

Bertossi, E., Aleo, F., Braghittoni, D., et al. (2016a). Stuck in the here and now: Construction of fictitious and future experiences following ventromedial prefrontal damage. *Neuropsychologia*, **81**, 107–16

Bertossi, E., Candela, V., De Luca, F., et al. (2017a). Episodic future thinking following vmPFC damage: impaired event construction, maintenance, or narration? *Neuropsychology*, **31**, 337

Bertossi, E., Ciaramelli, E. (2016). Ventromedial prefrontal damage reduces mind-wandering and biases its temporal focus. *Social Cognitive and Affective Neuroscience*, **11**, 1783–91

Bertossi, E., Peccenini, L., Solmi, A., et al. (2017b). Transcranial direct current stimulation of the medial prefrontal cortex dampens mind-wandering in men. *Scientific reports*, **7**, 16962

Bertossi, E., Tesini, C., Cappelli, A., et al. (2016b). Ventromedial prefrontal damage causes a pervasive impairment of episodic memory and future thinking. *Neuropsychologia*, **90**, 12–24

Bhanji, J., Smith, D.V., Delgado, M. (2019). *A Brief Anatomical Sketch of Human Ventromedial Prefrontal Cortex*. PsyArXiv.

Bicks, L.K., Koike, H., Akbarian, S., et al. (2015). Prefrontal cortex and social cognition in mouse and man. *Frontiers in psychology*, **6**, 1805

Binder, J.R., Frost, J.A., Hammeke, T.A., et al. (1999). Conceptual processing during the conscious resting state: a functional MRI study. *Journal of cognitive neuroscience*, **11**, 80–93

Bird, C.M., Capponi, C., King, J.A., et al. (2010). Establishing the boundaries: the hippocampal contribution to imagining scenes. *Journal of Neuroscience*, **30**, 11688–95

Blair, R.J. (2016). The neurobiology of impulsive aggression. *Journal of child and adolescent psychopharmacology*, **26**, 4–9

Blumer, D., Benson, DF (1975). Personality changes with frontal and temporal lobe lesions. *Psychiatric aspects of neurologic disease*, **1**, 151–70

Boboeva, V., Brasselet, R., Treves, A. (2018). The Capacity for Correlated Semantic Memories in the Cortex. *Entropy*, **20**, 824

Bohbot, V.D., Kalina, M., Stepankova, K., et al. (1998). Spatial memory deficits in patients with lesions to the right hippocampus and to the right parahippocampal cortex. *Neuropsychologia*, **36**, 1217–38

Bonnici, H.M., Chadwick, M.J., Lutti, A., et al. (2012). Detecting Representations of Recent and Remote Autobiographical Memories in vmPFC and Hippocampus. *The Journal of Neuroscience*, **32**, 16982–91

Bonnici, H.M., Maguire, E.A. (2018). Two years later – Revisiting autobiographical memory representations in vmPFC and hippocampus. *Neuropsychologia*, **110**, 159–69

Borsutzky, S., Fujiwara, E., Brand, M., et al. (2008). Confabulations in alcoholic Korsakoff patients. *Neuropsychologia*, **46**, 3133–43

Bower, G.H., Black, J.B., Turner, T.J. (1979). Scripts in memory for text. *Cognitive Psychology*, **11**, 177–220

Bowman, A.D., Griffis, J.C., Visscher, K.M., et al. (2017). Relationship Between Alpha Rhythm and the Default Mode Network: An EEG-fMRI Study. *Journal of Clinical Neurophysiology*, **34**, 527–33

Bowman, C.R., Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience*, **38**, 2605–14

Braitenberg, V. (1978). Cortical architectonics: general and areal. In: *Architectonics of the cerebral cortex*. Raven Press, p. 443–65

Braitenberg, V., Schüz, A. (2013). *Anatomy of the cortex: statistics and geometry*. Springer Science & Business Media

Bransford, J.D., Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of verbal learning and verbal behavior*, **11**, 717–26

Branzi, F.M., Pobric, G., Jung, J., et al. (2021). The left angular gyrus is causally involved in context-dependent integration and associative encoding during narrative reading. *Journal of cognitive neuroscience*, **33**, 1082–95

Braunstein, L.M., Gross, J.J., Ochsner, K.N. (2017). Explicit and implicit emotion regulation: a multi-level framework. *Social cognitive and affective neuroscience*, **12**, 1545–57

Brod, G., Lindenberger, U., Wagner, A.D., et al. (2016). Knowledge acquisition during exam preparation improves memory and modulates memory formation. *Journal of Neuroscience*, **36**, 8103–11

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L. (2008). *The Brain's Default Network*: *Anatomy, Function, and Relevance to Disease*. *Annals of the New York Academy of Sciences*, **1124**, 1–38

Buckner, R.L., Carroll, D.C. (2007). Self-projection and the brain. *Trends in cognitive sciences*, **11**, 49–57

Buckner, R.L., Snyder, A.Z., Shannon, B.J., et al. (2005). Molecular, Structural, and Functional Characterization of Alzheimer's Disease: Evidence for a Relationship between Default Activity, Amyloid, and Memory. *The Journal of Neuroscience*, **25**, 7709–17

Bueti, D., Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 1831–40

Buhle, J.T., Silvers, J.A., Wager, T.D., et al. (2014). Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies. *Cerebral cortex*, **24**, 2981–90

Burgess, P., Mcneil, J. (1999). Content-Specific Confabulation. *Cortex*, **35**, 163–82

Burgess, P.W., Shallice, T. (1996). Confabulation and the Control of Recollection. *Memory*, **4**, 359–412

van Buuren, M., Kroes, M.C., Wagner, I.C., et al. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *Journal of Neuroscience*, **34**, 16662–70

Cabeza, R., St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, **11**, 219–27

Cai, D.J., Mednick, S.A., Harrison, E.M., et al. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences*, **106**, 10130–34

Callard, F., Smallwood, J., Golchert, J., et al. (2013). The era of the wandering mind? Twenty-first century research on self-generated mental activity. *Frontiers in Psychology*, **4**

Cameron, C.D., Reber, J., Spring, V.L., et al. (2018). Damage to the ventromedial prefrontal cortex is associated with impairments in both spontaneous and deliberative moral judgments. *Neuropsychologia*, **111**, 261–68

Campbell, K.L., Madore, K.P., Benoit, R.G., et al. (2018). Increased hippocampus to ventromedial prefrontal connectivity during the construction of episodic future events. *Hippocampus*, **28**, 76–80

Cann, D.R., McRae, K., Katz, A.N. (2011). False recall in the Deese–Roediger–McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology*, **64**, 1515–42

Capitani, E., Laiacona, M. (1988). Aging and psychometric diagnosis of intellectual impairment: some considerations on test scores and their use

Carmichael, L., Hogan, H.P., Walter, A.A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, **15**, 73–86

de Caso, I., Karapanagiotidis, T., Aggius-Vella, E., et al. (2017). Knowing me, knowing you: Resting-state functional connectivity of ventromedial prefrontal cortex dissociates memory related to self from a familiar other. *Brain and Cognition*, **113**, 65–75

Catani, M., Dell'Acqua, F., De Schotten, M.T. (2013). A revised limbic system model for memory, emotion and behaviour. *Neuroscience & Biobehavioral Reviews*, **37**, 1724–37

Catrambone, R., Beike, D., Niedenthal, P. (1996). Is the self-concept a habitual referent in judgments of similarity? *Psychological Science*, **7**, 158–63

Cavanagh, S.E., Hunt, L.T., Kennerley, S.W. (2020). A diversity of intrinsic timescales underlie neural computations. *Frontiers in Neural Circuits*, **14**, 615626

Chang, Y.-M. (2005). Increased Action Potential Firing Rates of Layer 2/3 Pyramidal Cells in the Prefrontal Cortex are Significantly Related to Cognitive Performance in Aged Monkeys. *Cerebral Cortex*, **15**, 409–18

Charlesworth, L.A., Allen, R.J., Havelka, J., et al. (2016). Who am I? Autobiographical retrieval improves access to self-concepts. *Memory*, **24**, 1033–41

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., et al. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, **88**, 419–31

Christoff, K., Gordon, A.M., Smallwood, J., et al. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, **106**, 8719–24

Christoff, K., Irving, Z.C., Fox, K.C., et al. (2016). Mind-wandering as spontaneous thought: a dynamic framework. *Nature reviews neuroscience*, **17**, 718–31

Christoff, K., Ream, J.M., Gabrieli, J.D. (2004). Neural basis of spontaneous thought processes. *Cortex*, **40**, 623–30

Ciaramelli, E. (2008). The role of ventromedial prefrontal cortex in navigation: A case of impaired wayfinding and rehabilitation. *Neuropsychologia*, **46**, 2099–2105

Ciaramelli, E., Anelli, F., Frassinetti, F. (2021). An asymmetry in past and future mental time travel following vmPFC damage. *Social Cognitive and Affective Neuroscience*, **16**, 315–25

Ciaramelli, E., Braghittoni, D., Di Pellegrino, G. (2012). It Is the Outcome that Counts! Damage to the Ventromedial Prefrontal Cortex Disrupts the Integration of Outcome and Belief Information for Moral Judgment. *Journal of the International Neuropsychological Society*, **18**, 962–71

Ciaramelli, E., De Luca, F., Monk, A.M., et al. (2019). What" wins" in VMPFC: Scenes, situations, or schema? *Neuroscience & Biobehavioral Reviews*, **100**, 208–10

Ciaramelli, E., Ghetti, S. (2007). What are confabulators' memories made of? A study of subjective and objective measures of recollection in confabulation. *Neuropsychologia*, **45**, 1489–1500

Ciaramelli, E., Ghetti, S., Borsotti, M. (2009). Divided attention during retrieval suppresses false recognition in confabulation. *Cortex*, **45**, 141–53

Ciaramelli, E., Ghetti, S., Frattarelli, M., et al. (2006). When true memory availability promotes false memory: evidence from confabulating patients. *Neuropsychologia*, **44**, 1866–77

Ciaramelli, E., di Pellegrino, G. (2011). Ventromedial prefrontal cortex and the future of morality. *Emotion Review*, **3**, 308–9

Ciaramelli, E., Spaniol, J. (2009). Ventromedial prefrontal damage and memory for context: Perceptual versus semantic features. *Neuropsychology*, **23**, 649

Ciaramelli, E., Sperotto, R.G., Mattioli, F., et al. (2013). Damage to the ventromedial prefrontal cortex reduces interpersonal disgust. *Social Cognitive and Affective Neuroscience*, **8**, 171–80

Ciaramelli, E., Treves, A. (2019). A mind free to wander: neural and computational constraints on spontaneous thought. *Frontiers in psychology*, **10**, 39

Clark, D.L., Boutros, N.N., Mendez, M.F. (2010). *The Brain and Behavior: An Introduction to Behavioral Neuroanatomy*. 3rd ed. Cambridge University Press.

Clithero, J.A., Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social cognitive and affective neuroscience*, **9**, 1289–1302

Cobb Scott, J., Woods, S.P., Vigil, O., et al. (2011). Script Generation of Activities of Daily Living in HIV-Associated Neurocognitive Disorders. *Journal of the International Neuropsychological Society*, **17**, 740–45

Cocchi, L., Sale, M.V., L Gollo, L., et al. (2016). A hierarchy of timescales explains distinct effects of local inhibition of primary visual cortex and frontal eye fields. *Elife*, **5**, e15252

Cole, S.N., Morrison, C.M., Barak, O., et al. (2016). Amnesia and future thinking: Exploring the role of memory in the quantity and quality of episodic future thoughts. *British Journal of Clinical Psychology*, **55**, 206–24

Concha, L., Gross, D.W., Beaulieu, C. (2005). Diffusion tensor tractography of the limbic system. *American Journal of Neuroradiology*, **26**, 2267–74

Conway, M.A. (2005). Memory and the self. *Journal of memory and language*, **53**, 594–628

Conway, M.A., Pleydell-Pearce, C.W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, **107**, 261

Conway, M.A., Rubin, D.C. (2019). The structure of autobiographical memory. In: *Theories of memory*. Psychology Press, p. 103–37.

Conway, M.A., Tacchi, P.C. (1996). Motivated confabulation. *Neurocase*, **2**, 325–39

Cooper, R., Shallice, T., Farringdon, J. (1995). Symbolic and continuous processes in the automatic selection of actions. *Hybrid problems, hybrid solutions*, **2**, 27–37

Corcoran, J. (2006). Schemata: the concept of schema in the history of logic. *Bulletin of Symbolic Logic*, **12**, 219–40

Coronel, J.C., Federmeier, K.D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, **84**, 36–43

Coutanche, M.N., Thompson-Schill, S.L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, **143**, 2296–2303

Coutanche, M.N., Thompson-Schill, S.L. (2015). Rapid consolidation of new knowledge in adulthood via fast mapping. *Trends in Cognitive Sciences*, **19**, 486–88

Craik, F.I.M., Moroz, T.M., Moscovitch, M., et al. (1999). In Search of the Self: A Positron Emission Tomography Study. *Psychological Science*, **10**, 26–34

Crovitz, H.F., Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society*, **4**, 517–18

Croxson, P.L., Johansen-Berg, H., Behrens, T.E.J., et al. (2005). Quantitative Investigation of Connections of the Prefrontal Cortex in the Human and Macaque using Probabilistic Diffusion Tractography. *The Journal of Neuroscience*, **25**, 8854–66

Cunningham, S.J., Turk, D.J., Macdonald, L.M., et al. (2008). Yours or mine? Ownership and memory. *Consciousness and cognition*, **17**, 312–18

Dalla Barba, G. (1993). Different patterns of confabulation. *Cortex*, **29**, 567–81

Dalla Barba, G., Brazzarola, M., Marangoni, S., et al. (2017). A longitudinal study of confabulation. *Cortex*, **87**, 44–51

Dalla Barba, G., Brazzarola, M., Marangoni, S., et al. (2020). Confabulation affecting temporal consciousness significantly more than knowing consciousness. *Neuropsychologia*, **140**, 107367

Dalla Barba, G., Mantovan, M.C., Cappelletti, J.Y., et al. (1998). Temporal gradient in confabulation. *Cortex*, **34**, 417–26

Dalton, M.A., Zeidman, P., McCormick, C., et al. (2018). Differentiable processing of objects, associations, and scenes within the hippocampus. *Journal of Neuroscience*, **38**, 8146–59

Damasio, A.R., Tranel, D., Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural brain research*, **41**, 81–94

D'Argembeau, A. (2011). Motivational aspects of future thinking in the ventromedial prefrontal cortex. *From DNA to social cognition*, 81–90

D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Frontiers in human neuroscience*, **7**, 372

D'Argembeau, A. (2020). Zooming in and out on one's life: Autobiographical representations at multiple time scales. *Journal of cognitive neuroscience*, **32**, 2037–55

D'Argembeau, A., Cassol, H., Phillips, C., et al. (2014). Brains creating stories of selves: the neural basis of autobiographical reasoning. *Social Cognitive and Affective Neuroscience*, **9**, 646–52

D'Argembeau, A., Feyers, D., Majerus, S., et al. (2008). Self-reflection across time: cortical midline structures differentiate between present and past selves. *Social cognitive and affective neuroscience*, **3**, 244–52

D'Argembeau, A., Jedidi, H., Balteau, E., et al. (2012). Valuing one's self: medial prefrontal involvement in epistemic and emotive investments in self-views. *Cerebral Cortex*, **22**, 659–67

D'Argembeau, A., Mathy, A. (2011). Tracking the construction of episodic future thoughts. *Journal of experimental psychology: General*, **140**, 258

D'Argembeau, A., Ruby, P., Collette, F., et al. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of cognitive neuroscience*, **19**, 935–44

D'Argembeau, A., Salmon, E. (2012). The neural basis of semantic and episodic forms of self-knowledge: insights from functional neuroimaging. *Advances in experimental medicine and biology*, **739**, 276–90

D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van der Linden, M., et al. (2010a). Modulation of medial prefrontal and inferior parietal cortices when thinking about past, present, and future selves. *Social neuroscience*, **5**, 187–200

D'Argembeau, A., Stawarczyk, D., Majerus, S., Collette, F., Van Der Linden, M., et al. (2010b). The Neural Basis of Personal Goal Processing When Envisioning Future Events. *Journal of Cognitive Neuroscience*, **22**, 1701–13

Davey, C.G., Pujol, J., Harrison, B.J. (2016). Mapping the self in the brain's default mode network. *NeuroImage*, **132**, 390–97

Daviddi, S., Pedale, T., St. Jacques, P.L., et al. (2023). Common and distinct correlates of construction and elaboration of episodic-autobiographical memory: An ALE meta-analysis. *Cortex*, **163**, 123–38

De Luca, F., Benuzzi, F., Bertossi, E., et al. (2018a). Episodic future thinking and future-based decision-making in a case of retrograde amnesia. *Neuropsychologia*, **110**, 92–103

De Luca, F., McCormick, C., Ciaramelli, E., et al. (2019). Scene processing following damage to the ventromedial prefrontal cortex. *NeuroReport*, **30**, 828–33

De Luca, F., McCormick, C., Mullally, S.L., et al. (2018b). Boundary extension is attenuated in patients with ventromedial prefrontal cortex damage. *Cortex*, **108**, 1–12

De Vito, S., Gamboz, N., Brandimonte, M.A., et al. (2012). Future thinking in Parkinson's disease: An executive function? *Neuropsychologia*, **50**, 1494–1501

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, **58**, 17

Delbecq-Derouesne, J., Beauvois, M.F., Shallice, T. (1990). Preserved recall versus impaired recognition: A case study. *Brain*, **113**, 1045–74

Delgado, M.R., Beer, J.S., Fellows, L.K., et al. (2016). Viewpoints: Dialogues on the functional role of the ventromedial prefrontal cortex. *Nature Neuroscience*, **19**, 1545–52

Della Sala, S., Laiacona, M., Spinnler, H., et al. (1993). Autobiographical recollection and frontal damage. *Neuropsychologia*, **31**, 823–39

DeLuca, J. (2000). A cognitive neuroscience perspective on confabulation. *Neuropsychoanalysis*, **2**, 119–32

Denny, B.T., Kober, H., Wager, T.D., et al. (2012). A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, **24**, 1742–52

Dhaka, S., Kashyap, N. (2017). Explicit emotion regulation: Comparing emotion inducing stimuli. *Psychological Thought*, **10**

Diekelmann, S., Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, **11**, 114–26

Do, Q., Hasselmo, M.E. (2021). Neural circuits and symbolic processing. *Neurobiology of Learning and Memory*, **186**, 107552

Douglas, R.J., Martin, K.A., Whitteridge, D. (1989). A canonical microcircuit for neocortex. *Neural computation*, **1**, 480–88

Dudai, Y., Karni, A., Born, J. (2015). The Consolidation and Transformation of Memory. *Neuron*, **88**, 20–32

Dumay, N., Gaskell, M.G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, **18**, 35–39

Durrant, S.J., Cairney, S.A., McDermott, C., et al. (2015). Schema-conformant memories are preferentially consolidated during REM sleep. *Neurobiology of Learning and Memory*, **122**, 41–50

Duval, C., Desgranges, B., de La Sayette, V., et al. (2012). What happens to personal identity when semantic knowledge degrades? A study of the self and autobiographical memory in semantic dementia. *Neuropsychologia*, **50**, 254–65

Ellamil, M., Fox, K.C., Dixon, M.L., et al. (2016). Dynamics of neural recruitment surrounding the spontaneous arising of thoughts in experienced mindfulness practitioners. *Neuroimage*, **136**, 186–96

Elliott, R. (2000). Dissociable Functions in the Medial and Lateral Orbitofrontal Cortex: Evidence from Human Neuroimaging Studies. *Cerebral Cortex*, **10**, 308–17

Elston, G.N., Benavides-Piccione, R., DeFelipe, J. (2001). The pyramidal cell in cognition: a comparative study in human and monkey. *The Journal of Neuroscience*, **21**, RC163

Emmott, C., Alexander, M. (2014). *Schemata*. de Gruyter.

Ersner-Hershfield, H., Wimmer, G.E., Knutson, B. (2009). Saving for the future self: Neural measures of future self-continuity predict temporal discounting. *Social cognitive and affective neuroscience*, **4**, 85–92

Etkin, A., Büchel, C., Gross, J.J. (2015). The neural bases of emotion regulation. *Nature reviews neuroscience*, **16**, 693–700

Faber, M., Mills, C. (2018). The Critical Role of the Hippocampus in Mind Wandering. *The Journal of Neuroscience*, **38**, 6439–41

Fan, L., Zhuang, K., Wang, X., et al. (2023). Exploring the behavioral and neural correlates of semantic distance in creative writing. *Psychophysiology*, **60**, e14239

Farzanfar, D., Spiers, H.J., Moscovitch, M., et al. (2023). From cognitive maps to spatial schemas. *Nature Reviews Neuroscience*, **24**, 63–79

Feinstein, A., Levine, B., Protzner, A. (2000). Confabulation and multiple sclerosis: a rare association. *Multiple Sclerosis Journal*, **6**, 186–91

Fellows, L.K., Farah, M.J. (2005). Dissociable elements of human foresight: a role for the ventromedial frontal lobes in framing the future, but not in discounting future rewards. *Neuropsychologia*, **43**, 1214–21

Feng, S., D'Mello, S., Graesser, A.C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic bulletin & review*, **20**, 586–92

Fenigstein, A., Levine, M.P. (1984). Self-attention, concept activation, and the causal self. *Journal of Experimental Social Psychology*, **20**, 231–45

Ferrari, C., Lega, C., Vernice, M., et al. (2016). The dorsomedial prefrontal cortex plays a causal role in integrating social impressions from faces and verbal descriptions. *Cerebral cortex*, **26**, 156–65

Ferreira, C.S., Charest, I., Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *NeuroImage*, **201**, 115996

Finlay, B.L., Uchiyama, R. (2015). Developmental mechanisms channeling cortical evolution. *Trends in neurosciences*, **38**, 69–76

Flagan, T., Beer, J.S. (2013). *Three ways in which midline regions contribute to selfevaluation. Front. Hum. Neurosci. 7, 450*.

Förster, J., Friedman, R.S., Liberman, N. (2004). Temporal construal effects on abstract and concrete thinking: consequences for insight and creative cognition. *Journal of personality and social psychology*, **87**, 177

Fotopoulou, A. (2008). False selves in neuropsychological rehabilitation: The challenge of confabulation. *Neuropsychological rehabilitation*, **18**, 541–65

Fotopoulou, A. (2010). The affective neuropsychology of confabulation and delusion. *Cognitive neuropsychiatry*, **15**, 38–63

Fotopoulou, A., Conway, M.A., Solms, M. (2007). Confabulation: Motivated reality monitoring. *Neuropsychologia*, **45**, 2180–90

Fox, K.C.R., Spreng, R.N., Ellamil, M., et al. (2015). The wandering brain: Meta-analysis of functional neuroimaging studies of mind-wandering and related spontaneous thought processes. *NeuroImage*, **111**, 611–21

Frank, D.W., Dewitt, M., Hudgens-Haney, M., et al. (2014). Emotion regulation: quantitative meta-analysis of functional activation and deactivation. *Neuroscience & Biobehavioral Reviews*, **45**, 202–11

Fransson, P., Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: Evidence from a partial correlation network analysis. *Neuroimage*, **42**, 1178–84

Frederick, S., Novemsky, N., Wang, J., et al. (2009). Opportunity cost neglect. *Journal of Consumer Research*, **36**, 553–61

Frederiks, J.A.M. (1969). Disorders of the body schema

Gage, N.M., Baars, B.J. (2018). Humans Are Social Beings. In: *Fundamentals of Cognitive Neuroscience*. Elsevier, p. 321–56.

Gais, S., Albouy, G., Boly, M., et al. (2007). Sleep transforms the cerebral trace of declarative memories. *Proceedings of the National Academy of Sciences*, **104**, 18778–83

Gallagher, H.L., Frith, C.D. (2003). Functional imaging of 'theory of mind'. *Trends in cognitive sciences*, **7**, 77–83

Gangopadhyay, P., Chawla, M., Dal Monte, O., et al. (2021). Prefrontal–amygdala circuits in social decision-making. *Nature neuroscience*, **24**, 5–18

Gao, R., van den Brink, R.L., Pfeffer, T., et al. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *Elife*, **9**, e61277

Geerligs, L., Gözükara, D., Oetringer, D., et al. (2022). A partially nested cortical hierarchy of neural states underlies event segmentation in the human brain. *eLife*, **11**, e77430

Gelbard-Sagiv, H., Mukamel, R., Harel, M., et al. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science*, **322**, 96–101

Gershman, S.J., Jones, C.E., Norman, K.A., et al. (2013). Gradual extinction prevents the return of fear: implications for the discovery of state. *Frontiers in Behavioral Neuroscience*, **7**

Gherman, S., Philiastides, M.G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *Elife*, **7**, e38293

Ghosh, V.E., Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, **53**, 104–14

Ghosh, V.E., Moscovitch, M., Colella, B.M., et al. (2014). Schema representation in patients with ventromedial PFC lesions. *Journal of Neuroscience*, **34**, 12057–70

Giacometti Giordani, L., Crisafulli, A., Cantarella, G., et al. (2023). The role of posterior parietal cortex and medial prefrontal cortex in distraction and mind-wandering. *Neuropsychologia*, **188**, 108639

Gilbert, D.T., Wilson, T.D. (2007). Prospection: Experiencing the future. *Science*, **317**, 1351–54

Gilbert, S.J., Henson, R.N., Simons, J.S. (2010). The scale of functional specialization within human prefrontal cortex. *Journal of Neuroscience*, **30**, 1233–37

Gilbert, S.J., Spengler, S., Simons, J.S., et al. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *Journal of cognitive neuroscience*, **18**, 932–48

Gilboa, A. (2004). Autobiographical and episodic memory—one and the same?: Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia*, **42**, 1336–49

Gilboa, A. (2010). Strategic retrieval, confabulations, and delusions: Theory and data. *Cognitive Neuropsychiatry*, **15**, 145–80

Gilboa, A., Alain, C., He, Y., et al. (2009). Ventromedial Prefrontal Cortex Lesions Produce Early Functional Alterations during Remote Memory Retrieval. *The Journal of Neuroscience*, **29**, 4871–81

Gilboa, A., Alain, C., Stuss, D.T., et al. (2006). Mechanisms of spontaneous confabulations: a strategic retrieval account. *Brain*, **129**, 1399–1414

Gilboa, A., Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in cognitive sciences*, **21**, 618–31

Gilboa, A., Moscovitch, M. (2021). No consolidation without representation: Correspondence between neural and psychological representations in recent and remote memory. *Neuron*, **109**, 2239–55

Gilboa, A., Moscovitch, M. (2002). The cognitive neuroscience of confabulation: A review and a model. *Handbook of memory disorders*, **2**, 315–42

Gilboa, A., Moscovitch, M. (2017). Ventromedial prefrontal cortex generates pre-stimulus theta coherence desynchronization: a schema instantiation hypothesis. *Cortex*, **87**, 16–30

Gilboa, A., Verfaellie, M. (2010). Introduction—telling it like it isn't: the cognitive neuroscience of confabulation. *Journal of the International Neuropsychological Society*, **16**, 961–66

Gillihan, S.J., Xia, C., Padon, A.A., et al. (2011). Contrasting roles for lateral and ventromedial prefrontal cortex in transient and dispositional affective experience. *Social cognitive and affective neuroscience*, **6**, 128–37

Giordani, L.G., Crisafulli, A., Cantarella, G., et al. (2023). The role of posterior parietal cortex and medial prefrontal cortex in distraction and mind-wandering. *Neuropsychologia*, **188**, 108639

Girn, M., Mills, C., Laycock, E., et al. (2017). Neural dynamics of spontaneous thought: an electroencephalographic study. In: *Augmented Cognition. Neurocognition and Machine Learning: 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 11*. Springer, p. 28–44

Giuliano, A.E., Bonasia, K., Ghosh, V.E., et al. (2021). Differential influence of ventromedial prefrontal cortex lesions on neural representations of schema and semantic category knowledge. *Journal of Cognitive Neuroscience*, **33**, 1928–55

Gluth, S., Sommer, T., Rieskamp, J., et al. (2015). Effective Connectivity between Hippocampus and Ventromedial Prefrontal Cortex Controls Preferential Choices from Memory. *Neuron*, **86**, 1078–90

Godbout, L., Cloutier, P., Bouchard, C., et al. (2004). Script generation following frontal and parietal lesions. *Journal of Clinical and Experimental Neuropsychology*, **26**, 857–73

Godbout, L., Doyon, J. (2000). Defective Representation of Knowledge in Parkinson's Disease: Evidence from a Script-Production Task. *Brain and Cognition*, **44**, 490–510

Godbout, L., Doyon, J. (1995). Mental representation of knowledge following frontal-lobe or postrolandic lesions. *Neuropsychologia*, **33**, 1671–96

Gold, A.L., Morey, R.A., McCarthy, G. (2015). Amygdala–prefrontal cortex functional connectivity during threat-induced anxiety and goal distraction. *Biological psychiatry*, **77**, 394–403

Gold, A.L., Shechner, T., Farber, M.J., et al. (2016). Amygdala–cortical connectivity: associations with anxiety, development, and threat. *Depression and anxiety*, **33**, 917–26

Goldberg, M.C., Spinelli, S., Joel, S., et al. (2011). Children with high functioning autism show increased prefrontal and temporal cortex activity during error monitoring. *Developmental Cognitive Neuroscience*, **1**, 47–56

Goldin, P.R., McRae, K., Ramel, W., et al. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological psychiatry*, **63**, 577–86

Goldstein, N.J., Cialdini, R.B. (2007). The spyglass self: a model of vicarious self-perception. *Journal of personality and social psychology*, **92**, 402

Goodman, G.S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, **12**, 473–95

Grabenhorst, F., Rolls, E.T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in cognitive sciences*, **15**, 56–67

Grafman, J., Sirigu, A., Spector, L., et al. (1993). Damage to the prefrontal cortex leads to decomposition of structured event complexes: *Journal of Head Trauma Rehabilitation*, **8**, 73–87

Graham, L.M. (2007). Need for cognition and false memory in the Deese–Roediger–McDermott paradigm. *Personality and Individual Differences*, **42**, 409–18

Green, A.E., Fugelsang, J.A., Kraemer, D.J.M., et al. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, **1096**, 125–37

Green, A.E., Kraemer, D.J.M., Fugelsang, J.A., et al. (2010). Connecting Long Distance: Semantic Distance in Analogical Reasoning Modulates Frontopolar Cortex Activity. *Cerebral Cortex*, **20**, 70–76

Greicius, M.D., Krasnow, B., Reiss, A.L., et al. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the national academy of sciences*, **100**, 253–58

Greicius, M.D., Supekar, K., Menon, V., et al. (2009). Resting-State Functional Connectivity Reflects Structural Connectivity in the Default Mode Network. *Cerebral Cortex*, **19**, 72–78

Grilli, M.D., Bercel, J.J., Wank, A.A., et al. (2018). The contribution of the left anterior ventrolateral temporal lobe to the retrieval of personal semantics. *Neuropsychologia*, **117**, 178–87

Grilli, M.D., Verfaellie, M. (2016). Experience-near but not experience-far autobiographical facts depend on the medial temporal lobe for retrieval: Evidence from amnesia. *Neuropsychologia*, **81**, 180–85

Grilli, M.D., Verfaellie, M. (2014). Personal semantic memory: insights from neuropsychological research on amnesia. *Neuropsychologia*, **61**, 56–64

Grilli, M.D., Verfaellie, M. (2015). Supporting the self-concept with memory: insight from amnesia. *Social Cognitive and Affective Neuroscience*, **10**, 1684–92

Gronchi, G., Giovannelli, F. (2018). Dual process theory of thought and default mode network: A possible neural foundation of fast thinking. *Frontiers in psychology*, **9**, 1237

Gross, J.J. (2008). Emotion regulation. *Handbook of emotions*, **3**, 497–513

Gross, J.J. (2014). Emotion regulation: Conceptual and empirical foundations. *Handbook of emotion regulation*, **2**, 3–20

Gross, J.J., Thompson, R.A. (2007). Emotion regulation: Conceptual foundations, 2007. *Handbook of emotion regulation*, 3–24

Gruberger, M., Ben-Simon, E., Levkovitz, Y., et al. (2011). Towards a neuroscience of mind-wandering. *Frontiers in human neuroscience*, **5**, 56

Guo, D., Chen, Gang, Yang, J. (2023). Effects of schema on the relationship between post-encoding brain connectivity and subsequent durable memory. *Scientific Reports*, **13**, 8736

Guo, D., Chen, Haoyu, Wang, L., et al. (2023a). Effects of prior knowledge on brain activation and functional connectivity during memory retrieval. *Scientific Reports*, **13**, 13650

Guo, D., Yang, J. (2020). Interplay of the long axis of the hippocampus and ventromedial prefrontal cortex in schema-related memory retrieval. *Hippocampus*, **30**, 263–77

Guo, D., Yang, J. (2023b). Reactivation of schema representation in lateral occipital cortex supports successful memory encoding. *Cerebral Cortex*, **33**, 5968–80

Gupta, R., Tranel, D., Duff, M.C. (2012). Ventromedial prefrontal cortex damage does not impair the development and use of common ground in social interaction: implications for cognitive theory of mind. *Neuropsychologia*, **50**, 145–52

Gusnard, D.A., Akbudak, E., Shulman, G.L., et al. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences*, **98**, 4259–64

Gusnard, D.A., Raichle, M.E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature reviews neuroscience*, **2**, 685–94

Gyurak, A., Gross, J.J., Etkin, A. (2011). Explicit and implicit emotion regulation: A dual-process framework. *Cognition and emotion*, **25**, 400–412

Haber, S.N., Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human Imaging. *Neuropsychopharmacology*, **35**, 4–26

Hahn, B., Ross, T.J., Yang, Y., et al. (2007). Nicotine enhances visuospatial attention by deactivating areas of the resting brain default network. *Journal of Neuroscience*, **27**, 3477–89

Halford, G.S., Busby, J. (2007). Acquisition of structured knowledge without instruction: The relational schema induction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**, 586

Harlow, J.M. (1848). Passage of an iron rod through the head. *The Boston Medical and Surgical Journal (1828-1851)*, **39**, 0_1

Harlow, J.M. (1993). Recovery from the passage of an iron bar through the head. *History of Psychiatry*, **4**, 274–81

Harris, A., Adolphs, R., Camerer, C., et al. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PloS one*, **6**, e21074

Hassabis, D., Kumaran, D., Maguire, E.A. (2007a). Using Imagination to Understand the Neural Basis of Episodic Memory. *The Journal of Neuroscience*, **27**, 14365–74

Hassabis, D., Kumaran, D., Vann, S.D., et al. (2007b). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, **104**, 1726–31

Hassabis, D., Maguire, E.A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, **11**, 299–306

Hassabis, D., Maguire, E.A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 1263–71

Head, H., Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain*, **34**, 102–254

Heatherton, T.F., Wyland, C.L., Macrae, C.N., et al. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience*, **1**, 18–25

Heaton, R.K., Chelune, G.J., Talley, J.L., et al. (2000). Wisconsin Card Sorting Test-Firenze. *Giunti OS*

Heberlein, A.S., Padon, A.A., Gillihan, S.J., et al. (2008). Ventromedial frontal lobe plays a critical role in facial emotion recognition. *Journal of cognitive neuroscience*, **20**, 721–33

Hebscher, M., Barkan-Abramski, M., Goldsmith, M., et al. (2016). Memory, decision-making, and the ventromedial prefrontal cortex (vmPFC): the roles of subcallosal and posterior orbitofrontal cortices in monitoring and control processes. *Cerebral Cortex*, **26**, 4590–4601

Hebscher, M., Gilboa, A. (2016). A boost of confidence: The role of the ventromedial prefrontal cortex in memory, decision-making, and schemas. *Neuropsychologia*, **90**, 46–58

Hebscher, M., Wing, E., Ryan, J., et al. (2019). Rapid cortical plasticity supports long-term memory formation. *Trends in Cognitive Sciences*, **23**, 989–1002

Hershfield, H.E., Bartels, D.M. (2018). The future self. *The psychology of thinking about the future*, 89–109

Herwig, U., Kaffenberger, T., Schell, C., et al. (2012). Neural activity associated with self-reflection. *BMC neuroscience*, **13**, 1–12

Hilgetag, C.C., Goulas, A., Changeux, J.-P. (2022). A natural cortical axis connecting the outside and inside of the human brain. *Network Neuroscience*, **6**, 950–59

Hiser, J., Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biological psychiatry*, **83**, 638–47

Hooker, C.I., Bruce, L., Lincoln, S.H., et al. (2011). Theory of Mind Skills Are Related to Gray Matter Volume in the Ventromedial Prefrontal Cortex in Schizophrenia. *Biological Psychiatry*, **70**, 1169–78

Hughes, B.L., Beer, J.S. (2012). Medial orbitofrontal cortex is associated with shifting decision thresholds in self-serving cognition. *NeuroImage*, **61**, 889–98

Humphreys, G.W., Forde, E.M.E. (1998). Disordered action schema and action disorganisation syndrome. *Cognitive neuropsychology*, **15**, 771–812

Hutchison, R.M., Gallivan, J.P. (2018). Functional coupling between frontoparietal and occipitotemporal pathways during action and perception. *Cortex*, **98**, 8–27

Ingvar, D.H. (1985). 'Memory of the future': an essay on the temporal organization of conscious awareness. *Human neurobiology*, **4**, 127–36

Irish, M. (2020). On the Interaction Between Episodic and Semantic Representations – Constructing a Unified Account of Imagination. In: A. Abraham (ed). *The Cambridge Handbook of the Imagination*. Cambridge University Press, p. 447–65

Irish, M., Addis, D.R., Hodges, J.R., et al. (2012). Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. *Brain*, **135**, 2178–91

Irish, M., Halena, S., Kamminga, J., et al. (2015). Scene construction impairments in Alzheimer's disease–A unique role for the posterior cingulate cortex. *Cortex*, **73**, 10–23

Irish, M., Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in behavioral neuroscience*, **7**, 27

Ito, Y., Ueno, T., Kitagami, S., et al. (2013). A computational exploration on the role of semantic memory in episodic future thinking. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*

Jackson, R.L., Bajada, C.J., Lambon Ralph, M.A., et al. (2020). The Graded Change in Connectivity across the Ventromedial Prefrontal Cortex Reveals Distinct Subregions. *Cerebral Cortex*, **30**, 165–80

Jbabdi, S., Lehman, J.F., Haber, S.N., et al. (2013). Human and Monkey Ventral Prefrontal Fibers Use the Same Organizational Principles to Reach Their Targets: Tracing versus Tractography. *The Journal of Neuroscience*, **33**, 3190–3201

Jenkins, A.C., Macrae, C.N., Mitchell, J.P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences*, **105**, 4507–12

Jenkins, A.C., Mitchell, J.P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience*, **6**, 211–18

Jenkins, L.M., Andrewes, D.G., Nicholas, C.L., et al. (2014). Social cognition in patients following surgery to the prefrontal cortex. *Psychiatry Research: Neuroimaging*, **224**, 192–203

John, M.F.S., McClelland, J.L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial intelligence*, **46**, 217–57

Johnson, M.K. (1991). Reality monitoring: Evidence from confabulation in organic brain disease patients. *Awareness of deficit after brain injury: Clinical and theoretical issues*, 176–97

Johnson, M.K., Hashtroudi, S., Lindsay, D.S. (1993). Source monitoring. *Psychological bulletin*, **114**, 3

Johnson, M.K., O'Connor, M., Cantor, J. (1997). Confabulation, memory deficits, and frontal dysfunction. *Brain and cognition*, **34**, 189–206

Johnson, M.K., Raye, C.L. (1998). False memories and confabulation. *Trends in cognitive sciences*, **2**, 137–45

Johnson, M.K., Raye, C.L., Mitchell, K.J., et al. (2006). Dissociating medial frontal and posterior cingulate activity during self-reflection. *Social Cognitive and Affective Neuroscience*, **1**, 56–64

Johnson, S.C., Baxter, L.C., Wilder, L.S., et al. (2002). Neural correlates of self-reflection. *Brain*, **125**, 1808–14

Jones, L.M., Fontanini, A., Sadacca, B.F., et al. (2007). Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, **104**, 18772–77

Jou, J., Flores, S. (2013). How are false memories distinguishable from true memories in the Deese–Roediger–McDermott paradigm? A review of the findings. *Psychological research*, **77**, 671–86

Kalisch, R., Korenfeld, E., Stephan, K.E., et al. (2006). Context-Dependent Human Extinction Memory Is Mediated by a Ventromedial Prefrontal and Hippocampal Network. *The Journal of Neuroscience*, **26**, 9503–11

Kan, I.P., Larocque, K.F., Lafleche, G., et al. (2010). Memory monitoring failure in confabulation: evidence from the semantic illusion paradigm. *Journal of the international Neuropsychological Society*, **16**, 1006–17

Kane, M.J., Brown, L.H., McVay, J.C., et al. (2007). For Whom the Mind Wanders, and When: An Experience-Sampling Study of Working Memory and Executive Control in Daily Life. *Psychological Science*, **18**, 614–21

Kang, P., Lee, J., Sul, S., et al. (2013). Dorsomedial prefrontal cortex activity predicts the accuracy in estimating others' preferences. *Frontiers in human neuroscience*, **7**, 686

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, **107**, 11163–70

Kaping, D., Vinck, M., Hutchison, R.M., et al. (2011). Specific contributions of ventromedial, anterior cingulate, and lateral prefrontal cortex for attentional selection and stimulus valuation. *PLoS biology*, **9**, e1001224

Kaplan, R., Friston, K.J. (2019). Entorhinal transformations in abstract frames of reference. *PLoS biology*, **17**, e3000230

Kapur, N., Coughlan, A.K. (1980). Confabulation and frontal lobe dysfunction. *Journal of Neurology, Neurosurgery, and Psychiatry*, **43**, 461

Karapanagiotidis, T., Bernhardt, B.C., Jefferies, E., et al. (2017). Tracking thoughts: Exploring the neural architecture of mental time travel during mind-wandering. *Neuroimage*, **147**, 272–81

Kelley, W.M., Macrae, C.N., Wyland, C.L., et al. (2002). Finding the Self? An Event-Related fMRI Study. *Journal of Cognitive Neuroscience*, **14**, 785–94

Kenny, D.A., West, T.V. (2010). Similarity and agreement in self-and other perception: A meta-analysis. *Personality and Social Psychology Review*, **14**, 196–213

Kern, R.S., Van Gorp, W.G., Cummings, J.L., et al. (1992). Confabulation in Alzheimer's disease. *Brain and Cognition*, **19**, 172–82

Kerns, J.G., Cohen, J.D., MacDonald III, A.W., et al. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, **303**, 1023–26

Kessels, R.P.C., De Haan, E.H.F., Kappelle, L.J., et al. (2001). Varieties of human spatial memory: a meta-analysis on the effects of hippocampal lesions. *Brain Research Reviews*, **35**, 295–303

Kihlstrom, J.F., Klein, S.B. (2014). The self as a knowledge structure. In: *Handbook of social cognition*. Psychology Press, p. 153–208.

Killingsworth, M.A., Gilbert, D.T. (2010). A wandering mind is an unhappy mind. *Science*, **330**, 932–932

Kim, K., Johnson, M.K. (2015). Activity in ventromedial prefrontal cortex during self-related processing: positive subjective value or personal significance? *Social cognitive and affective neuroscience*, **10**, 494–500

Kim, K., Johnson, M.K. (2012). Extended self: medial prefrontal activity during transient association of self and objects. *Social cognitive and affective neuroscience*, **7**, 199–207

Kim, K., & Johnson, M. K. (2014). Extended self: spontaneous activation of medial prefrontal cortex by objects that are 'mine'. *Social cognitive and affective neuroscience*, **9**(7), 1006-1012.

Kim, S., Dede, A.J.O., Hopkins, R.O., et al. (2015). Memory, scene construction, and the human hippocampus. *Proceedings of the National Academy of Sciences*, **112**, 4767–72

Kimberg, D.Y., Coslett, H.B., Schwartz, M.F. (2007). Power in Voxel-based Lesion-Symptom Mapping. *Journal of Cognitive Neuroscience*, **19**, 1067–80

Klein, S.B. (2012). Self, Memory, and the Self-Reference Effect: An Examination of Conceptual and Methodological Issues. *Personality and Social Psychology Review*, **16**, 283–300

Klein, S.B. (2013). The complex act of projecting oneself into the future. *Wiley Interdisciplinary Reviews: Cognitive Science*, **4**, 63–79

Klein, Stanley B, Cosmides, L., Costabile, K.A., et al. (2002a). Is there something special about the self? A neuropsychological case study. *Journal of Research in Personality*, **36**, 490–506

Klein, S.B., Cosmides, L., Costabile, K.A. (2003). Preserved knowledge of self in a case of Alzheimer's dementia. *Social Cognition*, **21**, 157–65

Klein, Stanley B., Cosmides, L., Tooby, J., et al. (2002b). Decisions and the evolution of memory: multiple systems, multiple functions. *Psychological review*, **109**, 306

Klein, S. B., Kihlstrom, J.F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology. General*, **115**, 26–38

Klein, S.B., Lax, M.L. (2010). The unanticipated resilience of trait self-knowledge in the face of neural damage. *Memory (Hove, England)*, **18**, 918–48

Klein, S.B., Loftus, J. (1993). Behavioral Experience and Trait Judgnents about the Self. *Personality and Social Psychology Bulletin*, **19**, 740–45

Klein, Stanley B., Loftus, J., Kihlstrom, J.F. (2002c). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, **20**, 353–79

Klein, S.B., Loftus, J., Kihlstrom, J.F. (1996). Self-knowledge of an amnesic patient: toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology. General*, **125**, 250–60

Klumpers, F., Raemaekers, M.A., Ruigrok, A.N., et al. (2010). Prefrontal mechanisms of fear reduction after threat offset. *Biological psychiatry*, **68**, 1031–38

Kobayashi, C., Glover, G.H., Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia*, **45**, 1522–32

Koechlin, E., Ody, C., Kouneiher, F. (2003). The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*, **302**, 1181–85

Koenigs, M., Huey, E.D., Calamia, M., et al. (2008a). Distinct regions of prefrontal cortex mediate resistance and vulnerability to depression. *Journal of Neuroscience*, **28**, 12341–48

Koenigs, M., Huey, E.D., Raymont, V., et al. (2008b). Focal brain damage protects against post-traumatic stress disorder in combat veterans. *Nature neuroscience*, **11**, 232–37

Koenigs, M., Young, L., Adolphs, R., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, **446**, 908–11

Kohn, N., Eickhoff, S.B., Scheller, M., et al. (2014). Neural network of cognitive emotion regulation—an ALE meta-analysis and MACM analysis. *Neuroimage*, **87**, 345–55

Kok, R. (2020). Emotion regulation. In: *Encyclopedia of personality and individual differences*. Springer, p. 1281–90.

Kong, J., Loggia, M.L., Zyloney, C., et al. (2010). Exploring the brain in pain: activations, deactivations and their relation. *Pain*, **148**, 257–67

Konu, D., Turnbull, A., Karapanagiotidis, T., et al. (2020). A role for the ventromedial prefrontal cortex in self-generated episodic social cognition. *Neuroimage*, **218**, 116977

Koo, T.K., Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, **15**, 155–63

Koole, S.L., Rothermund, K. (2011). "I feel better but I don't know why": The psychology of implicit emotion regulation. *Cognition and Emotion*, **25**, 389–99

Koole, S.L., Webb, T.L., Sheeran, P.L. (2015). Implicit emotion regulation: Feeling better without knowing why. *Current opinion in psychology*, **3**, 6–10

Kopelman, M. (1999). Retrograde amnesia in patients with diencephalic,temporal lobe or frontal lesions. *Neuropsychologia*, **37**, 939–58

Kopelman, M.D. (2019). Anomalies of autobiographical memory. *Journal of the International Neuropsychological Society*, **25**, 1061–75

Kopelman, M.D., Ng, N., Brouke, O.V.D. (1997). Confabulation extending across episodic, personal, and general semantic memory. *Cognitive Neuropsychology*, **14**, 683–712

Kopelman, M.D., Stanhope, N., Kingsley, D. (1999). Retrograde amnesia in patients with diencephalic, temporal lobe or frontal lesions. *Neuropsychologia*, **37**, 939–58

Koscik, T.R., Tranel, D. (2012). The Human Ventromedial Prefrontal Cortex Is Critical for Transitive Inference. *Journal of Cognitive Neuroscience*, **24**, 1191–1204

Kroes, M.C., Fernández, G. (2012). Dynamic neural systems enable adaptive, flexible memories. *Neuroscience & Biobehavioral Reviews*, **36**, 1646–66

Krueger, F., Barbey, A.K., Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in cognitive sciences*, **13**, 103–9

Kucyi, A., Esterman, M., Riley, C.S., et al. (2016). Spontaneous default network activity reflects behavioral variability independent of mind-wandering. *Proceedings of the National Academy of Sciences*, **113**, 13899–904

Kumaran, D., Summerfield, J.J., Hassabis, D., et al. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, **63**, 889–901

Kurby, C.A., Zacks, J.M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, **12**, 72–79

Kurczek, J., Wechsler, E., Ahuja, S., et al. (2015). Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. *Neuropsychologia*, **73**, 116–26

Kwan, D., Craver, C.F., Green, L., et al. (2015). Cueing the personal future to reduce discounting in intertemporal choice: Is episodic prospection necessary? *Hippocampus*, **25**, 432–43

Kwan, D., Craver, C.F., Green, L., et al. (2013). Dissociations in future thinking following hippocampal damage: evidence from discounting and time perspective in episodic amnesia. *Journal of Experimental Psychology: General*, **142**, 1355

Kwan, V.S.Y., Barrios, V., Ganis, G., et al. (2007). Assessing the neural correlates of self-enhancement bias: a transcranial magnetic stimulation study. *Experimental Brain Research*, **182**, 379–85

La Corte, V., Piolino, P. (2016). On the role of personal semantic memory and temporal distance in episodic future thinking: The TEDIFT model. *Frontiers in human neuroscience*, **10**, 385

Laird, A.R., Eickhoff, S.B., Li, K., et al. (2009). Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. *Journal of Neuroscience*, **29**, 14496–505

Landmann, N., Kuhn, M., Piosczyk, H., et al. (2014). The reorganisation of memory during sleep. *Sleep Medicine Reviews*, **18**, 531–41

Lane, P.C., Gobet, F., Cheng, P.C. (2000). Learning-based constraints on schemata

Lebreton, M., Abitbol, R., Daunizeau, J., et al. (2015). Automatic integration of confidence in the brain valuation signal. *Nature neuroscience*, **18**, 1159–67

Lee, H., Bellana, B., Chen, J. (2020). What can narratives tell us about the neural bases of human memory? *Current Opinion in Behavioral Sciences*, **32**, 111–19

Lee, J.L., Nader, K., Schiller, D. (2017). An update on memory reconsolidation updating. *Trends in cognitive sciences*, **21**, 531–45

Leech, R., Kamourieh, S., Beckmann, C.F., et al. (2011). Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *Journal of Neuroscience*, **31**, 3217–24

Leech, R., Sharp, D.J. (2014). The role of the posterior cingulate cortex in cognition and disease. *Brain*, **137**, 12–32

Lehéricy, S., Ducros, M., Van De Moortele, P., et al. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology*, **55**, 522–29

Lent, R., Azevedo, F.A.C., Andrade-Moraes, C.H., et al. (2012). How many neurons do you have? Some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience*, **35**, 1–9

Leopold, A., Krueger, F., dal Monte, O., et al. (2012). Damage to the left ventromedial prefrontal cortex impacts affective theory of mind. *Social Cognitive and Affective Neuroscience*, **7**, 871–80

Leow, W.K., Miikkulainen, R. (1993). Representing visual schemas in neural networks for scene analysis. In: *IEEE International Conference on Neural Networks*. IEEE, p. 1612–17.

Leow, W.K., Miikkulainen, R. (1997). Visual Schemas in Neural Networks for Object Recognition and Scene Analysis. *Connection Science*, **9**, 161–200

Levine, B., Svoboda, E., Hay, J.F., et al. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, **17**, 677–89

Lewis, P.A., Durrant, S.J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, **15**, 343–51

Libby, L.K., Eibach, R.P. (2002). Looking back in time: self-concept change affects visual perspective in autobiographical memory. *Journal of personality and social psychology*, **82**, 167

Liberman, N., Sagristano, M.D., Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of experimental social psychology*, **38**, 523–34

Liberman, N., Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of personality and social psychology*, **75**, 5

Lieberman, M.D. (2010). Social Cognitive Neuroscience. In: S. T. Fiske, D. T. Gilbert, G. Lindzey (eds). *Handbook of Social Psychology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. socpsy001005.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annu. Rev. Psychol.*, **58**, 259–89

Lieberman, M.D., Straccia, M.A., Meyer, M.L., et al. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, **99**, 311–28

van der Linden, M., Berkers, R.M., Morris, R.G., et al. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *Journal of Neuroscience*, **37**, 9474–85

Liu, Z.-X., Grady, C., Moscovitch, M. (2017). Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cerebral cortex*, **27**, 1991–2009

Lockwood, P.L., Wittmann, M.K., Apps, M.A.J., et al. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications*, **9**, 4747

Longe, O., Senior, C., Rippon, G. (2009). The Lateral and Ventromedial Prefrontal Cortex Work as a Dynamic Integrated System: Evidence from fMRI Connectivity Analysis. *Journal of Cognitive Neuroscience*, **21**, 141–54

Luber, B., Lisanby, S.H. (2014). Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *NeuroImage*, **85**, 961–70

Lynch, K., Keane, M.M., Verfaellie, M. (2020). The status of semantic memory in medial temporal lobe amnesia varies with demands on scene construction. *Cortex*, **131**, 114–22

Mackey, S., Petrides, M. (2014). Architecture and morphology of the human ventromedial prefrontal cortex. *European Journal of Neuroscience*, **40**, 2777–96

Mackey, S., Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *European Journal of Neuroscience*, **32**, 1940–50

Macrae, C.N., Moran, J.M., Heatherton, T.F., et al. (2004). Medial prefrontal activity predicts memory for self. *Cerebral cortex*, **14**, 647–54

Madore, K.P., Jing, H.G., Schacter, D.L. (2019). Episodic specificity induction and scene construction: Evidence for an event construction account. *Consciousness and Cognition*, **68**, 1–11

Maguire, E.A. (2001). Neuroimaging studies of autobiographical event memory A. Baddeley, M. Conway, J. Aggleton (eds). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **356**, 1441–51

Maguire, E.A., Hassabis, D. (2011). Role of the hippocampus in imagination and future thinking. *Proceedings of the National Academy of Sciences*, **108**

Maguire, E.A., Mullally, S.L. (2013). The hippocampus: a manifesto for change. *Journal of Experimental Psychology: General*, **142**, 1180

Maguire, E.A., Mummery, C.J. (1999). Differential modulation of a common memory retrieval network revealed by positron emission tomography. *Hippocampus*, **9**, 54–61

Malykhin, N., Concha, L., Seres, P., et al. (2008). Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. *Psychiatry Research: Neuroimaging*, **164**, 132–42

Mancuso, L., Cavuoti-Cabanillas, S., Liloia, D., et al. (2022). Tasks activating the default mode network map multiple functional systems. *Brain Structure and Function*, **227**, 1711–34

Manea, A.M., Zilverstand, A., Hayden, B., et al. (2023). Neural timescales reflect behavioral demands in freely moving rhesus macaques. *bioRxiv*

Markus, H. (1983). Self-knowledge: An expanded view. *Journal of Personality*, **51**, 543–65

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, **35**, 63–78

Markus, H., Smith, J., Moreland, R.L. (1985). Role of the self-concept in the perception of others. *Journal of Personality and social Psychology*, **49**, 1494

Marquine, M.J., Grilli, M.D., Rapcsak, S.Z., et al. (2016). Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. *Neuropsychologia*, **89**, 245–53

Martinelli, P., Sperduti, M., Piolino, P. (2013). Neural substrates of the self-memory system: New insights from a meta-analysis. *Human Brain Mapping*, **34**, 1515–29

Martin-Ordas, G., Atance, C.M., Louw, A. (2012). The role of episodic and semantic memory in episodic foresight. *Learning and Motivation*, **43**, 209–19

Masís-Obando, R., Norman, K.A., Baldassano, C. (2022). Schema representations in distinct brain networks support narrative memory during encoding and retrieval. *Elife*, **11**, e70445

Mason, M.F., Norton, M.I., Van Horn, J.D., et al. (2007). Wandering minds: the default network and stimulus-independent thought. *science*, **315**, 393–95

Mauss, I.B., Bunge, S.A., Gross, J.J. (2007). Automatic emotion regulation. *Social and Personality Psychology Compass*, **1**, 146–67

Mazoyer, B., Zago, L., Mellet, E., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain research bulletin*, **54**, 287–98

McAndrews, M.P., Girard, T.A., Wilkins, L.K., et al. (2016). Semantic congruence affects hippocampal response to repetition of visual associations. *Neuropsychologia*, **90**, 235–42

McClelland, J.L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, **142**, 1190–1210

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419–57

McCormick, C., Barry, D.N., Jafarian, A., et al. (2020). vmPFC Drives Hippocampal Processing during Autobiographical Memory Recall Regardless of Remoteness. *Cerebral Cortex*, **30**, 5972–87

McCormick, C., Ciaramelli, E., De Luca, F., et al. (2018a). Comparing and Contrasting the Cognitive Effects of Hippocampal and Ventromedial Prefrontal Cortex Damage: A Review of Human Lesion Studies. *Neuroscience*, **374**, 295–318

McCormick, C., Dalton, M.A., Zeidman, P., et al. (2021). Characterising the hippocampal response to perception, construction and complexity. *Cortex*, **137**, 1–17

McCormick, C., Maguire, E.A. (2021). The distinct and overlapping brain networks supporting semantic and spatial constructive scene processing. *Neuropsychologia*, **158**, 107912

McCormick, C., Moscovitch, M., Valiante, T.A., et al. (2018b). Different neural routes to autobiographical memory recall in healthy people and individuals with left medial temporal lobe epilepsy. *Neuropsychologia*, **110**, 26–36

McCormick, C., Rosenthal, C.R., Miller, T.D., et al. (2017). Deciding what is possible and impossible following hippocampal damage in humans. *Hippocampus*, **27**, 303–14

McCormick, C., Rosenthal, C.R., Miller, T.D., et al. (2018c). Mind-wandering in people with hippocampal damage. *Journal of Neuroscience*, **38**, 2745–54

McGuire, P.K., Paulesu, E., Frackowiak, R.S., et al. (1996). Brain activity during stimulus independent thought. *Neuroreport*, **7**, 2095–99

McKenzie, S., Eichenbaum, H. (2011). Consolidation and reconsolidation: two lives of memories? *Neuron*, **71**, 224–33

McKenzie, S., Frank, A.J., Kinsky, N.R., et al. (2014). Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron*, **83**, 202–15

McKenzie, S., Robinson, N.T., Herrera, L., et al. (2013). Learning causes reorganization of neuronal firing patterns to represent related experiences within a hippocampal schema. *Journal of Neuroscience*, **33**, 10243–56

Meisner, O.C., Nair, A., Chang, S.W.C. (2022). Amygdala connectivity and implications for social cognition and disorders. In: *Handbook of Clinical Neurology*. Elsevier, p. 381–403.

Mekern, V., Hommel, B., Sjoerds, Z. (2019). Computational models of creativity: a review of single-process and multi-process recent approaches to demystify creative cognition. *Current Opinion in Behavioral Sciences*, **27**, 47–54

Melo, B., Winocur, G., Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology*, **16**, 343–59

Mendez, M.F., Fras, I.A. (2011). The false memory syndrome: experimental studies and comparison to confabulations. *Medical hypotheses*, **76**, 492–96

Mercer, B., Wapner, W., Gardner, H., et al. (1977). A study of confabulation. *Archives of neurology*, **34**, 429–33

Mesulam, M.-M. (2000). Behavioral neuroanatomy. *Principles of behavioral and cognitive neurology*, **2**, 1–120

Metcalf, K., Langdon, R., Coltheart, M. (2007). Models of confabulation: A critical review and a new framework. *Cognitive Neuropsychology*, **24**, 23–47

Metcalf, K., Langdon, R., Coltheart, M. (2010). The role of personal biases in the explanation of confabulation. *Cognitive Neuropsychiatry*, **15**, 64–94

Miller, E.K., Erickson, C.A., Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of neuroscience*, **16**, 5154–67

Mills, C., Herrera-Bennett, A., Faber, M., et al. (2018). Why the mind wanders: How spontaneous thought's default variability may support episodic efficiency and semantic optimization. *The Oxford handbook of spontaneous thought: Mind-wandering, creativity, and dreaming*, 11–22

Mitchell, J.P., Banaji, M.R., Macrae, C.N. (2005). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, **17**, 1306–15

Mitchell, J.P., Heatherton, T.F., Macrae, C.N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*, **99**, 15238–43

Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2006). Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others. *Neuron*, **50**, 655–63

Mitchell, J.P., Schirmer, J., Ames, D.L., et al. (2011). Medial prefrontal cortex predicts intertemporal choice. *Journal of cognitive neuroscience*, **23**, 857–66

Mittner, M., Hawkins, G.E., Boekel, W., et al. (2016). A Neural Model of Mind Wandering. *Trends in Cognitive Sciences*, **20**, 570–78

Mobbs, D., Petrovic, P., Marchant, J.L., et al. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science*, **317**, 1079–83

Mocaiber, I., Pereira, M.G., Erthal, F.S., et al. (2010). Fact or fiction? An event-related potential study of implicit emotion regulation. *Neuroscience Letters*, **476**, 84–88

Monk, A.M., Barnes, G.R., Maguire, E.A. (2020). The Effect of Object Type on Building Scene Imagery—an MEG Study. *Frontiers in Human Neuroscience*, **14**, 592175

Monk, A.M., Dalton, M.A., Barnes, G.R., et al. (2021). The Role of Hippocampal–Ventromedial Prefrontal Cortex Neural Dynamics in Building Mental Representations. *Journal of Cognitive Neuroscience*, **33**, 89–103

Montague, P.R., King-Casas, B., Cohen, J.D. (2006). Imaging valuation models in human choice. *Annu. Rev. Neurosci.*, **29**, 417–48

Mooneyham, B.W., Schooler, J.W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale*, **67**, 11–18

van Moort, M.L., Jolles, D.D., Koornneef, A., et al. (2020). What you read versus what you know: Neural correlates of accessing context information and background knowledge in constructing a mental representation during reading. *Journal of Experimental Psychology: General*, **149**, 2084

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**, 197–211

Moran, J.M., Macrae, C.N., Heatherton, T.F., et al. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of cognitive neuroscience*, **18**, 1586–94

Moretto, G., Sellitto, M., Di Pellegrino, G. (2013). Investment and repayment in a trust game after ventromedial prefrontal damage. *Frontiers in Human Neuroscience*, **7**

Moscovitch, M. (1995). Confabulation.

Moscovitch, M. (1989). Confabulation and the frontal systems: Strategic versus associative retrieval in neuropsychological theories of memory. *Varieties of memory and consciousness: Essays in honour of Endel Tulving*, 133–60

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 / Revue canadienne de psychologie expérimentale*, **62**, 62–79

Moscovitch, M., Cabeza, R., Winocur, G., et al. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, **67**, 105–34

Moscovitch, M., Gilboa, A. (2022). Has the concept of systems consolidation outlived its usefulness? Identification and evaluation of premises underlying systems consolidation. *Faculty Reviews*, **11**

Moscovitch, M., Melo, B. (1997). Strategic retrieval and the frontal lobes: Evidence from confabulation and amnesia. *Neuropsychologia*, **35**, 1017–34

Moscovitch, M., Winocur, G. (2002). The Frontal Cortex and Working with Memory. In: D. T. Stuss, R. T. Knight (eds). *Principles of Frontal Lobe Function*. Oxford University Press, p. 188–209.

Motzkin, J.C., Philippi, C.L., Wolf, R.C., et al. (2015). Ventromedial Prefrontal Cortex Is Critical for the Regulation of Amygdala Activity in Humans. *Biological Psychiatry*, **77**, 276–84

Mudrik, L., Lamy, D., Deouell, L.Y. (2010). ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*, **48**, 507–17

Mullally, S.L., Hassabis, D., Maguire, E.A. (2012). Scene construction in amnesia: An fMRI study. *Journal of Neuroscience*, **32**, 5646–53

Mullally, S.L., Maguire, E.A. (2014). Memory, imagination, and predicting the future: A common brain mechanism? *The Neuroscientist*, **20**, 220–34

Mullally, S.L., Vargha-Khadem, F., Maguire, E.A. (2014). Scene construction in developmental amnesia: an fMRI study. *Neuropsychologia*, **52**, 1–10

Murayama, K., Izuma, K., Aoki, R., et al. (2016). "Your Choice" motivates you in the brain: the emergence of autonomy neuroscience. *Recent developments in neuroscience research on human motivation*, 95–125

Murayama, K., Matsumoto, M., Izuma, K., et al. (2015). How Self-Determined Choice Facilitates Performance: A Key Role of the Ventromedial Prefrontal Cortex. *Cerebral Cortex*, **25**, 1241–51

Murray, R.J., Schaer, M., Debbané, M. (2012). Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self-and other-reflection. *Neuroscience & Biobehavioral Reviews*, **36**, 1043–59

Murray, S., Krasich, K., Schooler, J.W., et al. (2020). What's in a task? Complications in the study of the task-unrelated-thought variety of mind wandering. *Perspectives on Psychological Science*, **15**, 572–88

Myers-Schulz, B., Koenigs, M. (2012). Functional anatomy of ventromedial prefrontal cortex: implications for mood and anxiety disorders. *Molecular Psychiatry*, **17**, 132–41

Nahum, L., Bouzerda-Wahlen, A., Guggisberg, A., et al. (2012). Forms of confabulation: dissociations and associations. *Neuropsychologia*, **50**, 2524–34

Naim, M., Boboeva, V., Kang, C.J., et al. (2018). Reducing a cortical network to a Potts model yields storage capacity estimates. *Journal of Statistical Mechanics: Theory and Experiment*, **2018**, 043304

Nawa, N.E., Ando, H. (2020). Effective connectivity during autobiographical memory search. *Brain and Behavior*, **10**, e01719

Nawa, N.E., Ando, H. (2019). Effective connectivity within the ventromedial prefrontal cortex-hippocampus-amygdala network during the elaboration of emotional autobiographical memories. *NeuroImage*, **189**, 316–28

Nedjam, Z., Dalla Barba, G., Pillon, B. (2000). Confabulation in a patient with fronto-temporal dementia and a patient with Alzheimer's disease. *Cortex*, **36**, 561–77

Nelson, K. (2015). Developing self-knowledge from autobiographical memory. In: *The mental representation of trait and autobiographical knowledge about the self*. Psychology Press, p. 111–21.

Nieuwenhuis, I.L.C., Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural Brain Research*, **218**, 325–34

Norman, D.A., Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In: *Consciousness and self-regulation: Advances in research and theory volume 4*. Springer, p. 1–18.

Northoff, G., Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in cognitive sciences*, **8**, 102–7

Northoff, G., Heinzel, A., De Greck, M., et al. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*, **31**, 440–57

Nuthall, G. (2000). The role of memory in the acquisition and retention of knowledge in science and social studies units. *Cognition and instruction*, **18**, 83–139

Nyberg, L., Kim, A.S., Habib, R., et al. (2010). Consciousness of subjective time in the brain. *Proceedings of the National Academy of Sciences*, **107**, 22356–59

O'Callaghan, C., Shine, J.M., Hodges, J.R., et al. (2019). Hippocampal atrophy and intrinsic brain network dysfunction relate to alterations in mind wandering in neurodegeneration. *Proceedings of the National Academy of Sciences*, **116**, 3316–21

O'Doherty, J.P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, **1239**, 118–29

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**, 1369–80

Öngür, D., Ferry, A.T., Price, J.L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comparative Neurology*, **460**, 425–49

Ongur, D., Price, J. L. (2000). The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans. *Cerebral Cortex*, **10**, 206–19

Pace-Schott, E.F., Nave, G., Morgan, A., et al. (2012). Sleep-dependent modulation of affectively guided decision-making. *Journal of Sleep Research*, **21**, 30–39

Paller, K.A., Voss, J.L. (2004). Memory reactivation and consolidation during sleep. *Learning & Memory*, **11**, 664–70

Papagno, C., Baddeley, A. (1997). Confabulation in a dysexecutive patient: implication for models of retrieval. *Cortex*, **33**, 743–52

Pardilla-Delgado, E., Payne, J.D. (2017). The Deese-Roediger-McDermott (DRM) task: A simple cognitive paradigm to investigate false memories in the laboratory. *JoVE (Journal of Visualized Experiments)*, e54793

Parfit, D. (1971). Personal identity. *The Philosophical Review*, **80**, 3–27

Paulus, M.P., Frank, L.R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *NeuroReport*, **14**, 1311–15

Pelham, B.W. (1991). On confidence and consequence: The certainty and importance of self-knowledge. *Journal of personality and social psychology*, **60**, 518

Perry, A., Lwi, S.J., Verstaen, A., et al. (2016). The role of the orbitofrontal cortex in regulation of interpersonal space: evidence from frontal lesion and frontotemporal dementia patients. *Social Cognitive and Affective Neuroscience*, **11**, 1894–1901

Pestrides, M. (1991). Functional specialization within the dorsolateral frontal cortex for serial order memory. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **246**, 299–306

Peters, S.L., Fellows, L.K., Sheldon, S. (2017). The ventromedial frontal lobe contributes to forming effective solutions to real-world problems. *Journal of cognitive neuroscience*, **29**, 991–1001

Phelps, E.A., Delgado, M.R., Nearing, K.I., et al. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron*, **43**, 897–905

Philippi, C.L., Bruss, J., Boes, A.D., et al. (2021). Lesion network mapping demonstrates that mind-wandering is associated with the default mode network. *Journal of Neuroscience Research*, **99**, 361–73

Philippi, C.L., Duff, M.C., Denburg, N.L., et al. (2012a). Medial PFC damage abolishes the self-reference effect. *Journal of cognitive neuroscience*, **24**, 475–81

Philippi, C.L., Feinstein, J.S., Khalsa, S.S., et al. (2012b). Preserved self-awareness following extensive bilateral brain damage to the insula, anterior cingulate, and medial prefrontal cortices

Philippi, C.L., Tranel, D., Duff, M., et al. (2015). Damage to the default mode network disrupts autobiographical memory retrieval. *Social Cognitive and Affective Neuroscience*, **10**, 318–26

Piaget, J. (1926). *Language and Thought of the Child: Selected Works vol 5*. Routledge.

Piaget, J. (1952). *The origins of intelligence in children.* New York: W W Norton & Co.

Picard, L., Mayor-Dubois, C., Maeder, P., et al. (2013). Functional independence within the self-memory system: New insights from two cases of developmental amnesia. *Cortex*, **49**, 1463–81

Picó-Pérez, M., Alemany-Navarro, M., Dunsmoor, J.E., et al. (2019). Common and distinct neural correlates of fear extinction and cognitive reappraisal: a meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews*, **104**, 102–15

Pierce, J.L., Kostova, T., Dirks, K.T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of general psychology*, **7**, 84–107

Platt, J., Spivack, G. (1975). *Manual for the means-ends problem-solving procedure (MEPS) : a measure of interpersonal cognitive problem-solving skill*.

Poerio, G.L., Sormaz, M., Wang, H.-T., et al. (2017). The role of the default mode network in component processes underlying the wandering mind. *Social cognitive and affective neuroscience*, **12**, 1047–62

Poldrack, R.A., Wagner, A.D., Prull, M.W., et al. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, **10**, 15–35

Preston, A.R., Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, **23**, R764–73

Price, J.L., Drevets, W.C. (2010). Neurocircuitry of Mood Disorders. *Neuropsychopharmacology*, **35**, 192–216

Pronin, E., Ross, L. (2006). Temporal differences in trait self-ascription: When the self is seen as an other. *Journal of personality and social psychology*, **90**, 197

Ptak, R., Schnider, A. (1999). Spontaneous confabulations after orbitofrontal damage: The role of temporal context confusion and self-monitoring. *Neurocase*, **5**, 243–50

Pullen, E., Morris, R.G., Kerr, S., et al. (2006). Exploration of social rule violation in patients with focal prefrontal neurosurgical lesions using a virtual reality simulation. *International Journal on Disability and Human Development*, **5**

Race, E., Keane, M.M., Verfaellie, M. (2011). Medial temporal lobe damage causes deficits in episodic memory and episodic future thinking not attributable to deficits in narrative construction. *Journal of Neuroscience*, **31**, 10262–69

Raichle, M.E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, **38**, 433–47

Raichle, M.E., Snyder, A.Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage*, **37**, 1083–90

Ramanan, S., Piguet, O., Irish, M. (2018). Rethinking the role of the angular gyrus in remembering the past and imagining the future: the contextual integration model. *The Neuroscientist*, **24**, 342–52

Rameson, L.T., Satpute, A.B., Lieberman, M.D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, **50**, 701–8

Rangel, A., Clithero, J.A. (2014). The computation of stimulus values in simple choice. *Neuroeconomics*, 125–48

Rangel, A., Hare, T. (2010). Neural computations associated with goal-directed choice. *Current opinion in neurobiology*, **20**, 262–70

Rapcsak, S.Z., Kaszniak, A.W., Reminger, S.L., et al. (1998). Dissociation between verbal and autonomic measures of memory following frontal lobe damage. *Neurology*, **50**, 1259–65

Raykov, P.P., Keidel, J.L., Oakhill, J., et al. (2020). The brain regions supporting schema-related processing of people's identities. *Cognitive Neuropsychology*, **37**, 8–24

Reagh, Z.M., Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, **14**, 1279

Reber, J., Feinstein, J.S., O'Doherty, J.P., et al. (2017). Selective impairment of goal-directed decision-making following lesions to the human ventromedial prefrontal cortex. *Brain*, **140**, 1743–56

Reggev, N., Bein, O., Maril, A. (2016). Distinct neural suppression and encoding effects for conceptual novelty and familiarity. *Journal of Cognitive Neuroscience*, **28**, 1455–70

Renoult, L., Davidson, P.S., Palombo, D.J., et al. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends in cognitive sciences*, **16**, 550–58

Renoult, L., Tanguay, A., Beaudry, M., et al. (2016). Personal semantics: Is it distinct from episodic and semantic memory? An electrophysiological study of memory for autobiographical facts and repeated events in honor of Shlomo Bentin. *Neuropsychologia*, **83**, 242–56

Ritter, S.M., Strick, M., Bos, M.W., et al. (2012). Good morning creativity: task reactivation during sleep enhances beneficial effect of sleep on creative performance. *Journal of Sleep Research*, **21**, 643–47

Roberts, R.P., Schacter, D.L., Addis, D.R. (2018). Scene construction and relational processing: Separable constructs? *Cerebral Cortex*, **28**, 1729–32

Roberts, S., Henry, J.D., Molenberghs, P. (2019). Immoral behaviour following brain damage: a review. *Journal of neuropsychology*, **13**, 564–88

Robertson, E.M. (2012). New Insights in Human Memory Interference and Consolidation. *Current Biology*, **22**, R66–71

Robin, J., Moscovitch, M. (2017). Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. *Current Opinion in Behavioral Sciences*, **17**, 114–23

Roe, A.W. (2019). Columnar connectome: toward a mathematics of brain function. *Network Neuroscience*, **3**, 779–91

Roehri, N., Bréchet, L., Seeber, M., et al. (2022). Phase-Amplitude Coupling and Phase Synchronization Between Medial Temporal, Frontal and Posterior Brain Regions Support Episodic Autobiographical Memory Recall. *Brain Topography*, **35**, 191–206

Rogers, T.B., Kuiper, N.A., Kirker, W.S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, **35**, 677–88

Roland, P., Gulyas, B. (1994). Visual imagery and visual representation. *Trends in Neurosciences*, **17**, 281–87

Rolls, E.T. (2022). The hippocampus, ventromedial prefrontal cortex, and episodic and semantic memory. *Progress in Neurobiology*, **217**, 102334

Rolls, E.T., Deco, G., Huang, C.-C., et al. (2022). The human orbitofrontal cortex, vmPFC, and anterior cingulate cortex effective connectome: emotion, memory, and action. *Cerebral Cortex*, **33**, 330–56

Rolls, E.T., Hornak, J., Wade, D., et al. (1994). Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery & Psychiatry*, **57**, 1518–24

Roman, M., Brownell, H.H., Potter, H.H., et al. (1987). Script knowledge in right hemisphere-damaged and in normal elderly adults. *Brain and Language*, **31**, 151–70

Romero, K., Barense, M.D., Moscovitch, M. (2019). Coherence and congruency mediate medial temporal and medial prefrontal activity during event construction. *NeuroImage*, **188**, 710–21

Rorden, C., Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural neurology*, **12**, 191–200

Rosenbaum, R.S., Halilova, J.G., Kwan, D., et al. (2023). Temporal Construal Effects Are Independent of Episodic Future Thought. *Psychological Science*, **34**, 75–86

Rotshtein, P., Henson, R.N., Treves, A., et al. (2005). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature neuroscience*, **8**, 107–13

Rourke, L., Cruikshank, L.C., Shapke, L., et al. (2016). A neural marker of medical visual expertise: implications for training. *Advances in Health Sciences Education*, **21**, 953–66

Roy, M., Shohamy, D., Wager, T.D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, **16**, 147–56

Rumelhart, D.E. (2017). Schemata: The building blocks of cognition. In: *Theoretical issues in reading comprehension*. Routledge, p. 33–58.

Rumelhart, D.E. (1980). Schemata: The building blocks ofcognition. *Theoretical Issues in Reading Comprehension, edited by Rand J. Spiro, Bertram C. Bruce, and William F. Brewer. Hillsdale, NJ: Lawrence Erlbaum Associates*

Rumelhart, D.E., Lindsay, P.H., Norman, D.A. (1972). A process model for long-term memory.

Rumelhart, D.E., Ortony, A. (2017). The representation of knowledge in memory 1. In: *Schooling and the acquisition of knowledge*. Routledge, p. 99–135.

Rushworth, M.F., Behrens, T.E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature neuroscience*, **11**, 389–97

Rushworth, M.F., Noonan, M.P., Boorman, E.D., et al. (2011). Frontal cortex and reward-guided learning and decision-making. *Neuron*, **70**, 1054–69

Russo, E., Treves, A. (2012). Cortical free-association dynamics: Distinct phases of a latching network. *Physical Review E*, **85**, 051920

Ryom, K.I., Boboeva, V., Soldatkina, O., et al. (2021). Latching dynamics as a basis for short-term recall P. E. Latham (ed). *PLOS Computational Biology*, **17**, e1008809

Ryom, K.I., Treves, A. (2023). Speed Inversion in a Potts Glass Model of Cortical Dynamics. *PRX Life*, **1**, 013005

Sajonz, B., Kahnt, T., Margulies, D.S., et al. (2010). Delineating self-referential processing from episodic memory retrieval: common and dissociable networks. *Neuroimage*, **50**, 1606–17

Sala, S.D., Laiacona, M., Spinnler, H., et al. (1993). Autobiographical recollection and frontal damage. *Neuropsychologia*, **31**, 823–39

Sara, S.J., Hars, B. (2006). In memory of consolidation. *Learning & Memory*, **13**, 515–21

Sawczak, C., McAndrews, M.P., O'Connor, B.B., et al. (2022). I remember therefore I am: Episodic memory retrieval and self-reported trait empathy judgments in young and older adults and individuals with medial temporal lobe excisions. *Cognition*, **225**, 105124

Schacter, D.L., Addis, D.R. (2020). Memory and imagination: Perspectives on constructive episodic simulation. *The Cambridge handbook of the imagination*, 111–31

Schacter, D.L., Addis, D.R., Buckner, R.L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, **8**, 657–61

Schacter, D.L., Addis, D.R., Hassabis, D., et al. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, **76**, 677–94

Schacter, D.L., Benoit, R.G., De Brigard, F., et al. (2015). Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of Learning and Memory*, **117**, 14–21

Schacter, D.L., Benoit, R.G., Szpunar, K.K. (2017). Episodic future thinking: mechanisms and functions. *Current Opinion in Behavioral Sciences*, **17**, 41–50

Schaeffer, D.J., Hori, Y., Gilbert, K.M., et al. (2020). Divergence of rodent and primate medial frontal cortex functional connectivity. *Proceedings of the National Academy of Sciences*, **117**, 21681–89

Schank, R.C. (1983). *Dynamic memory: A theory of reminding and learning in computers and people*. cambridge university press.

Schank, R.C., Abelson, R.P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Schlichting, M.L., Mumford, J.A., Preston, A.R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, **6**, 8151

Schlichting, M.L., Preston, A.R. (2016). Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiology of learning and memory*, **134**, 91–106

Schlichting, M.L., Preston, A.R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current opinion in behavioral sciences*, **1**, 1–8

Schlichting, M.L., Preston, A.R. (2014). Memory reactivation during rest supports upcoming learning of related content. *Proceedings of the National Academy of Sciences*, **111**, 15845–50

Schmitz, T.W., Johnson, S.C. (2007). Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neuroscience & Biobehavioral Reviews*, **31**, 585–96

Schmitz, T.W., Johnson, S.C. (2006). Self-appraisal decisions evoke dissociated dorsal—ventral aMPFC networks. *NeuroImage*, **30**, 1050–58

Schmitz, T.W., Kawahara-Baccus, T.N., Johnson, S.C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage*, **22**, 941–47

Schneider, B., Koenigs, M. (2017). Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia*, **107**, 84–93

Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience*, **4**, 662–71

Schnider, A. (2001). Spontaneous confabulation, reality monitoring, and the limbic system — a review. *Brain Research Reviews*, **36**, 150–60

Schnider, A. (2000). Spontaneous confabulations, disorientation, and the processing of 'now'. *Neuropsychologia*, **38**, 175–85

Schnider, A., Ptak, R. (1999). Spontaneous confabulators fail to suppress currently irrelevant memory traces. *Nature neuroscience*, **2**, 677–81

Schnider, A., Ptak, R., von Däniken, C., et al. (2000). Recovery from spontaneous confabulations parallels recovery of temporal confusion in memory. *Neurology*, **55**, 74–83

Schnider, A., Von Däniken, C., Gutbrod, K. (1996). The mechanisms of spontaneous and provoked confabulations. *Brain*, **119**, 1365–75

Sebastian, C.L., Fontaine, N.M.G., Bird, G., et al. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, **7**, 53–63

Seghier, M.L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *The Neuroscientist*, **19**, 43–61

Sekeres, M.J., Winocur, G., Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience letters*, **680**, 39–53

Seli, P., Beaty, R.E., Cheyne, J.A., et al. (2018). How pervasive is mind wandering, really?,. *Consciousness and Cognition*, **66**, 74–78

Seli, P., Risko, E.F., Smilek, D., et al. (2016). Mind-Wandering With and Without Intention. *Trends in Cognitive Sciences*, **20**, 605–17

Sellitto, M., Ciaramelli, E., di Pellegrino, G. (2010). Myopic discounting of future rewards after medial orbitofrontal damage in humans. *Journal of Neuroscience*, **30**, 16429–36

Sestieri, C., Corbetta, M., Romani, G.L., et al. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: functional and topographic analyses. *Journal of Neuroscience*, **31**, 4407–20

Shallice, T., Cooper, R.P. (2012). The organisation of mind. *Cortex*, **48**, 1366–70

Shamay-Tsoory, S.G., Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, **45**, 3054–67

Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., et al. (2003). Characterization of empathy deficits following prefrontal brain damage: the role of the right ventromedial prefrontal cortex. *Journal of cognitive neuroscience*, **15**, 324–37

Shamay-Tsoory, S.G., Tomer, R., Berger, B.D., et al. (2005). Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, **18**, 55–67

Sharma, S., Arain, Mathur, et al. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment*, 449

Shaw, R.E., Hazelett, W.M. (2014). Schemas in cognition. In: *Event Cognition*. Psychology Press, p. 57–70.

Sheldon, S., McAndrews, M.P., Moscovitch, M. (2011). Episodic memory processes mediated by the medial temporal lobes contribute to open-ended problem solving. *Neuropsychologia*, **49**, 2439–47

Shulman, G.L., Corbetta, M., Buckner, R.L., et al. (1997). Common Blood Flow Changes across Visual Tasks: I. Increases in Subcortical Structures and Cerebellum but Not in Nonvisual Cortex. *Journal of Cognitive Neuroscience*, **9**, 624–47

Silvers, J.A., Moreira, J.F.G. (2019). Capacity and tendency: A neuroscientific framework for the study of emotion regulation. *Neuroscience letters*, **693**, 35–39

Singer, T., Tusche, A. (2014). Understanding others: Brain mechanisms of theory of mind and empathy. In: *Neuroeconomics*. Elsevier, p. 513–32.

Sio, U.N., Monaghan, P., Ormerod, T. (2013). Sleep on it, but only if it is difficult: Effects of sleep on problem solving. *Memory & Cognition*, **41**, 159–66

Sirigu, A., Zalla, T., Pillon, B., et al. (1995). Selective impairments in managerial knowledge following pre-frontal cortex damage. *Cortex*, **31**, 301–16

Smallwood, J. (2013). Distinguishing how from why the mind wanders: A process–occurrence framework for self-generated mental activity. *Psychological Bulletin*, **139**, 519–35

Smallwood, J., Andrews-Hanna, J. (2013). Not all minds that wander are lost: the importance of a balanced perspective on the mind-wandering state. *Frontiers in psychology*, **4**, 441

Smallwood, J., Bernhardt, B.C., Leech, R., et al. (2021). The default mode network in cognition: a topographical perspective. *Nature reviews neuroscience*, **22**, 503–13

Smallwood, J., Karapanagiotidis, T., Ruby, F., et al. (2016). Representing representation: Integration between the temporal lobe and the posterior cingulate influences the content and form of spontaneous thought. *PloS one*, **11**, e0152272

Smallwood, J., Schooler, J.W. (2006). The restless mind. *Psychological bulletin*, **132**, 946

Smallwood, J., Schooler, J.W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, **66**, 487–518

Smith, E.R. (1984). Attributions and other inferences: Processing information about the self versus others. *Journal of Experimental Social Psychology*, **20**, 97–115

Smith, R., Keramatian, K., Smallwood, J., et al. (2006). Mind-wandering with and without awareness: an fMRI study of spontaneous thought processes. In: *Proceedings of the 28th annual conference of the Cognitive Science Society*. Erlbaum Vancouver, p. 804–9.

Smith, V., Mitchell, D.J., Duncan, J. (2018). Role of the Default Mode Network in Cognitive Transitions. *Cerebral Cortex*, **28**, 3685–96

Sommer, T. (2017). The emergence of knowledge and how it supports the memory for novel related information. *Cerebral Cortex*, **27**, 1906–21

Sommer, T., Hennies, N., Lewis, P.A., et al. (2022). The assimilation of novel information into schemata and its efficient consolidation. *Journal of Neuroscience*, **42**, 5916–29

Sonuga-Barke, E.J., Castellanos, F.X. (2007). Spontaneous attentional fluctuations in impaired states and pathological conditions: a neurobiological hypothesis. *Neuroscience & Biobehavioral Reviews*, **31**, 977–86

Sotres-Bayon, F., Bush, D.E., LeDoux, J.E. (2004). Emotional perseveration: an update on prefrontal-amygdala interactions in fear extinction. *Learning & memory*, **11**, 525–35

Sotres-Bayon, F., Quirk, G.J. (2010). Prefrontal control of fear: more than just extinction. *Current opinion in neurobiology*, **20**, 231–35

Spalding, K.N., Jones, S.H., Duff, M.C., et al. (2015). Investigating the Neural Correlates of Schemas: Ventromedial Prefrontal Cortex Is Necessary for Normal Schematic Influence on Memory. *The Journal of Neuroscience*, **35**, 15746–51

Spalding, K.N., Schlichting, M.L., Zeithamova, D., et al. (2018). Ventromedial prefrontal cortex is necessary for normal associative inference and memory integration. *Journal of Neuroscience*, **38**, 3767–75

Spalla, D., Cornacchia, I.M., Treves, A. (2021). Continuous attractors for dynamic memories. *Elife*, **10**, e69499

Spinnler, H., Tognoni, G. (1987). Italian Group on the Neuropsychological Study of Ageing: Italian standardization and classification of neuropsychological tests. *Ital J Neurol Sci*, **6**, 1–120

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. *Journal of Cognitive Neuroscience*, **22**, 1112–23

Spreng, R.N., Mar, R.A., Kim, A.S.N. (2009). The Common Neural Basis of Autobiographical Memory, Prospection, Navigation, Theory of Mind, and the Default Mode: A Quantitative Meta-analysis. *Journal of Cognitive Neuroscience*, **21**, 489–510

Squire, L.R., Genzel, L., Wixted, J.T., et al. (2015). Memory Consolidation. *Cold Spring Harbor Perspectives in Biology*, **7**, a021766

Srull, T.K., Gaelick, L. (1983). General principles and individual differences in the self as a habitual reference point: An examination of self-other judgments of similarity. *Social Cognition*, **2**, 108–21

Stawarczyk, D., D'Argembeau, A. (2015). Neural correlates of personal goal processing during episodic future thinking and mind-wandering: An ALE meta-analysis. *Human brain mapping*, **36**, 2928–47

Stawarczyk, D., Majerus, S., Maquet, P., et al. (2011). Neural Correlates of Ongoing Conscious Experience: Both Task-Unrelatedness and Stimulus-Independence Are Related to Default Network Activity S. Gilbert (ed). *PLoS ONE*, **6**, e16997

Stendardi, D., Biscotto, F., Bertossi, E., et al. (2021). Present and future self in memory: the role of vmPFC in the self-reference effect. *Social Cognitive and Affective Neuroscience*, **16**, 1205–13

Stendardi, D., Giordani, L.G., Gambino, S., et al. (2023). Who am I really? The ephemerality of the self-schema following vmPFC damage. *Neuropsychologia*, **188**, 108651

Sterpenich, V., Albouy, G., Darsaud, A., et al. (2009). Sleep Promotes the Neural Reorganization of Remote Emotional Memory. *The Journal of Neuroscience*, **29**, 5143–52

Stevens, F.L., Hurley, R.A., Taber, K.H. (2011). Anterior Cingulate Cortex: Unique Role in Cognition and Emotion R. A. Hurley, L. A. Hayman, K. H. Taber (eds). *The Journal of Neuropsychiatry and Clinical Neurosciences*, **23**, 121–25

Stickgold, R., Scott, L., Rittenhouse, C., et al. (1999). Sleep-Induced Changes in Associative Memory. *Journal of Cognitive Neuroscience*, **11**, 182–93

Stickgold, R., Walker, M.P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature neuroscience*, **16**, 139–45

St-Laurent, M., Moscovitch, M., Levine, B., et al. (2009). Determinants of autobiographical memory in patients with unilateral temporal lobe epilepsy or excisions☆. *Neuropsychologia*, **47**, 2211–21

Stockwell, P. (2019). *Cognitive poetics: An introduction*. routledge.

Stone, V.E., Baron-Cohen, S., Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of cognitive neuroscience*, **10**, 640–56

Stuss, D.T., Gallup Jr, G.G., Alexander, M.P. (2001). The frontal lobes are necessary fortheory of mind'. *Brain*, **124**, 279–86

Sugiura, M., Shah, N.J., Zilles, K., et al. (2005). Cortical representations of personally familiar objects and places: functional organization of the human posterior cingulate cortex. *Journal of cognitive neuroscience*, **17**, 183–98

Sui, J., He, X., Humphreys, G.W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, **38**, 1105–17

Sui, J., Humphreys, G.W. (2015). The integrative self: How self-reference integrates perception and memory. *Trends in cognitive sciences*, **19**, 719–28

Sui, J., Humphreys, G.W. (2017). The ubiquitous self: what the properties of self-bias tell us about the self. *Annals of the New York Academy of Sciences*, **1396**, 222–35

Sui, J., Rotshtein, P., Humphreys, G.W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences*, **110**, 7607–12

Summerfield, J.J., Hassabis, D., Maguire, E.A. (2010). Differential engagement of brain regions within a 'core'network during scene construction. *Neuropsychologia*, **48**, 1501–9

Svoboda, E., McKinnon, M.C., Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, **44**, 2189–2208

Sweegers, C.C.G., Coleman, G.A., Van Poppel, E.A.M., et al. (2015). Mental Schemas Hamper Memory Storage of Goal-Irrelevant Information. *Frontiers in Human Neuroscience*, **9**

Symons, C.S., Johnson, B.T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, **121**, 371–94

Szpunar, K.K., Spreng, R.N., Schacter, D.L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences*, **111**, 18414–21

Szpunar, K.K., Watson, J.M., McDermott, K.B. (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, **104**, 642–47

Takashima, A., Nieuwenhuis, I.L.C., Rijpkema, M., et al. (2007). Memory trace stabilization leads to large-scale changes in the retrieval network: A functional MRI study on associative memory. *Learning & Memory*, **14**, 472–79

Takashima, A., Petersson, K.M., Rutters, F., et al. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences*, **103**, 756–61

Takeuchi, T., Tamura, M., Tse, D., et al. (2022). Brain region networks for the assimilation of new associative memory into a schema. *Molecular Brain*, **15**, 24

Talairach, T., Tournoux P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. Georg Thieme Verlag, Stuttgart.

Talland, G.A. (1965). Deranged memory: A psychonomic study of the amnesic syndrome. *(No Title)*

Tallberg, I.-M., Almkvist, O. (2001). Confabulation and Memory in Patients With Alzheimers Disease. *Journal of Clinical and Experimental Neuropsychology*, **23**, 172–84

Tamminen, J., Payne, J.D., Stickgold, R., et al. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. *The Journal of Neuroscience*, **30**, 14356–60

Tanguay, A.F., Palombo, D.J., Love, B., et al. (2023). The shared and unique neural correlates of personal semantic, general semantic, and episodic memory. *ELife*, **12**, e83645

Tanguay, A.N., Benton, L., Romio, L., et al. (2018). The ERP correlates of self-knowledge: Are assessments of one's past, present, and future traits closer to semantic or episodic memory? *Neuropsychologia*, **110**, 65–83

Taylor, S.E., Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, **103**, 193

Teffer, K., Semendeferi, K. (2012). Human prefrontal cortex. In: *Progress in Brain Research*. Elsevier, p. 191–218.

Teghil, A., Boccia, M. (2024). Brain connectivity patterns associated with individual differences in the access to experience-near personal semantics: a resting-state fMRI study. *Cognitive, Affective, & Behavioral Neuroscience*, 1–13

Tenenbaum, J.B., Kemp, C., Griffiths, T.L., et al. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, **331**, 1279–85

Thaiss, L., Petrides, M. (2008). Autobiographical memory of the recent past following frontal cortex or temporal lobe excisions. *European Journal of Neuroscience*, **28**, 829–40

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, **1**, 39–60

Tranel, D. (2002). Emotion, decision making, and the ventromedial prefrontal cortex. *Principles of frontal lobe function*, **338**, 353

Tremblay, R., Lee, S., Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*, **91**, 260–92

Treves, A. (2005). Frontal latching networks: a possible neural basis for infinite recursion. *Cognitive Neuropsychology*, **22**, 276–91

Tripathi, V., Garg, R. (2022). Weak task synchronization of default mode network in task based paradigms. *NeuroImage*, **251**, 118940

Trope, Y., Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, **117**, 440

Trope, Y., Liberman, N. (2003). Temporal construal. *Psychological review*, **110**, 403

Tse, D., Langston, R.F., Kakeyama, M., et al. (2007). Schemas and memory consolidation. *Science*, **316**, 76–82

Tse, D., Takeuchi, T., Kakeyama, M., et al. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science*, **333**, 891–95

Tsuchida, A., Fellows, L.K. (2012). Are you upset? Distinct roles for orbitofrontal and lateral prefrontal cortex in detecting and distinguishing facial expressions of emotion. *Cerebral cortex*, **22**, 2904–12

Tulving (1993). Self-knowledge of an amnesic individual is represented abstractly. In: *The mental representation of trait and autobiographical knowledge about the self*. Psychology Press.

Tulving, E. (1972). Episodic and semantic memory. *Organization of memory*, **1**, 1

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology*, **53**, 1–25

Tulving, E. (1985). How many memory systems are there? *American psychologist*, **40**, 385

Tupak, S.V., Dresler, T., Guhn, A., et al. (2014). Implicit emotion regulation in the presence of threat: neural and autonomic correlates. *Neuroimage*, **85**, 372–79

Turk, D.J., Van Bussel, K., Brebner, J.L., et al. (2011a). When "it" becomes "mine": Attentional biases triggered by object ownership. *Journal of cognitive neuroscience*, **23**, 3725–33

Turk, D.J., Van Bussel, K., Waiter, G.D., et al. (2011b). Mine and me: Exploring the neural basis of object ownership. *Journal of cognitive neuroscience*, **23**, 3657–68

Van Damme, I., d'Ydewalle, G. (2010). Confabulation versus experimentally induced false memories in Korsakoff patients. *Journal of Neuropsychology*, **4**, 211–30

Van Kesteren, M.T., Beul, S.F., Takashima, A., et al. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. *Neuropsychologia*, **51**, 2352–59

Van Kesteren, M.T., Brown, T.I. (2014). The medial prefrontal cortex and the deceptiveness of memory. *Journal of Neuroscience*, **34**, 13569–70

Van Kesteren, Marlieke T.R., Rijpkema, M., Ruiter, D.J., et al. (2010a). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *Journal of Neuroscience*, **30**, 15888–94

Van Kesteren, Marlieke T. R., Fernández, G., Norris, D.G., et al. (2010b). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, **107**, 7550–55

Van Kesteren, M.T.R., Meeter, M. (2020). How to optimize knowledge construction in the brain. *npj Science of Learning*, **5**, 5

Van Kesteren, M.T., Rignanese, P., Gianferrara, P.G., et al. (2020). Congruency and reactivation aid memory integration through reinstatement of prior knowledge. *Scientific Reports*, **10**, 4776

Van Kesteren, M.T.R., Ruiter, D.J., Fernández, G., et al. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, **35**, 211–19

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, **30**, 829–58

Vandekerckhove, M., Plessers, M., Van Mieghem, A., et al. (2014). Impaired facial emotion recognition in patients with ventromedial prefrontal hypoperfusion. *Neuropsychology*, **28**, 605

Vassena, E., Krebs, R.M., Silvetti, M., et al. (2014). Dissociating contributions of ACC and vmPFC in reward prediction, outcome, and choice. *Neuropsychologia*, **59**, 112–23

Verfaellie, M., Race, E., Keane, M.M. (2012). Medial temporal lobe contributions to future thinking: Evidence from neuroimaging and amnesia. *Psychologica Belgica*, **52**, 77

Verfaellie, M., Wank, A.A., Reid, A.G., et al. (2019). Self-related processing and future thinking: Distinct contributions of ventromedial prefrontal cortex and the medial temporal lobes. *Cortex*, **115**, 159–71

Vianna, D.M.L., Brandão, M.L. (2003). Anatomical connections of the periaqueductal gray: specific neural substrates for different kinds of fear. *Brazilian Journal of Medical and Biological Research*, **36**, 557–66

Viard, A., Chételat, G., Lebreton, K., et al. (2011). Mental time travel into the past and the future in healthy aged adults: an fMRI study. *Brain and cognition*, **75**, 1–9

Victor, M. (1976). The Wernicke-Korsakoff syndrome. *Handbokk of Clinical Neurology*, **28**, 243–70

Victor, M., Yakovlev, P.I. (1955). SS Korsakoff's psychic disorder in conjunction with peripheral neuritis: a translation of Korsakoff's original article with brief comments on the author and his contribution to clinical medicine. *Neurology*, **5**, 394–394

Viol, A., Treves, A., Ciaramelli, E. (2021). Navigating through the ebbs and flows of language. *Current Opinion in Neurobiology*, **70**, 130–36

Wagner, A.D., Schacter, D.L., Rotte, M., et al. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, **281**, 1188–91

Wagner, D.D., Chavez, R.S., Broom, T.W. (2019). Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, **10**, e1482

Wagner, D.D., Haxby, J.V., Heatherton, T.F. (2012). The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdisciplinary Reviews: Cognitive Science*, **3**, 451–70

Wagner, I.C., van Buuren, M., Kroes, M.C., et al. (2015). Schematic memory components converge within angular gyrus during retrieval. *elife*, **4**, e09668

Wagner, U., Gais, S., Haider, H., et al. (2004). Sleep inspires insight. *Nature*, **427**, 352–55

Walker, M.P., Liston, C., Hobson, J.A., et al. (2002). Cognitive flexibility across the sleep–wake cycle: REM-sleep enhancement of anagram problem solving. *Cognitive Brain Research*, **14**, 317–24

Wallis, J.D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience*, **15**, 13–19

Wang, S.-H., Tse, D., Morris, R.G. (2012). Anterior cingulate cortex in schema assimilation and expression. *Learning & memory*, **19**, 315–18

Wang, T., Huang, X.T. (2014). The role of episodic and semantic memory in episodic future thinking. *Advances in Psychology*, **4**, 496–505

Wang, T., Yue, T., Huang, X.T. (2016). Episodic and semantic memory contribute to familiar and novel episodic future thinking. *Frontiers in psychology*, **7**, 1746

Wank, A.A., Robertson, A., Thayer, S.C., et al. (2022). Autobiographical memory unknown: Pervasive autobiographical memory loss encompassing personality trait knowledge in an individual with medial temporal lobe amnesia. *Cortex*, **147**, 41–57

Warren, D.E., Jones, S.H., Duff, M.C., et al. (2014). False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory. *Journal of Neuroscience*, **34**, 7677–82

Warren, D.E., Kurczek, J., Duff, M.C. (2016). What relates newspaper, definite, and clothing? An article describing deficits in convergent problem solving and creativity following hippocampal damage: Hippocampal Damage Impairs Convergent Problem Solving. *Hippocampus*, **26**, 835–40

Watkins, E.R. (2008). Constructive and unconstructive repetitive thought. *Psychological bulletin*, **134**, 163

Webb, T.L., Totterdell, P., Ibar, D.N.H. (2015). Foundations and extensions for the extended model: More on implicit and explicit forms of emotion regulation. *Psychological Inquiry*, **26**, 123–29

Weilbächer, R.A., Gluth, S. (2016). The interplay of hippocampus and ventromedial prefrontal cortex in memory-based decision making. *Brain sciences*, **7**, 4

Weinstein, E.A. (1991). Anosognosia and denial of illness. *Awareness of deficit after brain injury: Clinical and theoretical issues*, 240–57

Weinstein, E.A., Kahn, R.L. (1955). Denial of illness: Symbolic and physiological aspects.

Welborn, B.L., Lieberman, M.D. (2015). Person-specific Theory of Mind in Medial pFC. *Journal of Cognitive Neuroscience*, **27**, 1–12

Wen, T., Mitchell, D.J., Duncan, J. (2020). The Functional Convergence and Heterogeneity of Social, Episodic, and Self-Referential Thought in the Default Mode Network. *Cerebral Cortex*, **30**, 5915–29

Westmacott, R., Leach, L., Freedman, M., et al. (2001). Different Patterns of Autobiographical Memory Loss in Semantic Dementia and Medial Temporal Lobe Amnesia: a Challenge to Consolidation Theory. *Neurocase*, **7**, 37–55

Wheeler, M.A., Stuss, D.T., Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological bulletin*, **121**, 331

Wheeler, M.E., Petersen, S.E., Buckner, R.L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, **97**, 11125–29

Willis, M.L., Palermo, R., McGrillen, K., et al. (2014). The nature of facial expression recognition deficits following orbitofrontal cortex damage. *Neuropsychology*, **28**, 613–23

Wing, E.A., D'Angelo, M.C., Gilboa, A., et al. (2021). The role of the ventromedial prefrontal cortex and basal forebrain in relational memory and inference. *Journal of Cognitive Neuroscience*, **33**, 1976–89

Wing, E.A., Marsh, E.J., Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia*, **51**, 2360–70

Wixted, J.T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, **55**, 235–69

Wolf, R.C., Philippi, C.L., Motzkin, J.C., et al. (2014). Ventromedial prefrontal cortex mediates visual attention during facial emotion recognition. *Brain*, **137**, 1772–80

Wolf, R.C., Pujara, M., Baskaya, M.K., et al. (2016). Emotion recognition deficits associated with ventromedial prefrontal cortex lesions are improved by gaze manipulation. *cortex*, **82**, 255–62

Wood, J.N., Tierney, M., Bidwell, L.A., et al. (2005). Neural correlates of script event knowledge: A neuropsychological study following prefrontal injury. *Cortex*, **41**, 796–804

Wrigley, S.N., Brown, G.J. (2004). A computational model of auditory selective attention. *IEEE Transactions on Neural Networks*, **15**, 1151–63

Xu, X., Yuan, H., Lei, X. (2016). Activation and Connectivity within the Default Mode Network Contribute Independently to Future-Oriented Thought. *Scientific Reports*, **6**, 21001

Yang, Y., Chen, Z., Zhang, R., et al. (2020). Neural substrates underlying episodic future thinking: A voxel-based morphometry study. *Neuropsychologia*, **138**, 107255

Yaoi, K., Osaka, M., Osaka, N. (2015). Neural correlates of the self-reference effect: evidence from evaluation and recognition processes. *Frontiers in human neuroscience*, **9**, 383

Yasin, S., Fierst, A., Keenan, H., et al. (2022). Self-Enhancement and the Medial Prefrontal Cortex: The Convergence of Clinical and Experimental Findings. *Brain Sciences*, **12**, 1103

Yeshurun, Y., Nguyen, M., Hasson, U. (2021). The default mode network: where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, **22**, 181–92

Yeterian, E.H., Pandya, D.N., Tomaiuolo, F., et al. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex*, **48**, 58–81

Yin, S., Bi, T., Chen, A., et al. (2021). Ventromedial prefrontal cortex drives the prioritization of self-associated stimuli in working memory. *Journal of Neuroscience*, **41**, 2012–23

Yu, L.Q., Kan, I.P., Kable, J.W. (2020). Beyond a rod through the skull: A systematic review of lesion studies of the human ventromedial frontal lobe. *Cognitive Neuropsychology*, **37**, 97–141

Zacharia, A.A., Kaur, S., Sharma, R. (2022). Altered functional connectivity: A possible reason for reduced performance during visual cognition involving scene incongruence and negative affect. *IBRO Neuroscience Reports*, **13**, 533–42

Zacks, J.M. (2020). Event Perception and Memory. *Annual Review of Psychology*, **71**, 165–91

Zadbood, A., Chen, J., Leong, Y.C., et al. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cerebral cortex*, **27**, 4988–5000

Zaki, J., Ochsner, K.N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience*, **15**, 675–80

Zangwill, O.L. (1953). Disorientation for age. *Journal of Mental Science*, **99**, 698–701

Zanini, S. (2008). Generalised script sequencing deficits following frontal lobe lesions. *Cortex*, **44**, 140–49

Zeithamova, D., Dominick, A.L., Preston, A.R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, **75**, 168–79

Zhang, L., Gläscher, J. (2020). A brain network supporting social influences in human decision-making. *Science advances*, **6**, eabb4159

Zhang, S., Wang, W., Zhornitsky, S., et al. (2018). Resting State Functional Connectivity of the Lateral and Medial Hypothalamus in Cocaine Dependence: An Exploratory Study. *Frontiers in Psychiatry*, **9**, 344

Zhou, J., Jia, C., Montesinos-Cartagena, M., et al. (2021). Evolving schema representations in orbitofrontal ensembles during learning. *Nature*, **590**, 606–11

Zimmer, C. (2011). 100 Trillion Connections. *Scientific American*, **304**, 58–63

Zysset, S., Huber, O., Ferstl, E., et al. (2002). The anterior frontomedian cortex and evaluative judgment: an fMRI study. *Neuroimage*, **15**, 983–91

Zysset, S., Huber, O., Samson, A., et al. (2003). Functional specialization within the anterior medial prefrontal cortex: a functional magnetic resonance imaging study with human subjects. *Neuroscience Letters*, **335**, 183–86

Appendix A

Non-scripted events

- Finding your lost watch – from the moment you realize you have lost it to when you find it;
- Making friends in your new neighbourhood – from moving to a new neighbourhood to making at least a friend there;
- Entertain and baby-sit three children aged 4, 7 and 11 for a few hours – from when a relative of yours brings them to you, until s/he collects them;
- Retrieve a soccer ball that your niece/nephew threw on a tree – from when you realize the ball is stuck, to when you successfully retrieve it.

Scripted events

1. Going out for dinner with friends – from the decision to go out, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Make a reservation/call the restaurant; Get ready to go out; Go to the restaurant; Order food; Eat; Pay; Return home.	Decide to go out for dinner; Choose the restaurant. Take your wallet/key/phone;	Call/invite your friends; Leave the house; Meet up with your friends; Enter the restaurant; Sit at your table.

2. Going to the movies – from the decision to go, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Choose the movie Get ready to go out; Get snacks; Get the tickets; Look for the cinema hall/enter the hall; Watch the movie; Return home.	Decide to go to the movies; Take your wallet/key/phone; Meet up with your friends; Go to the cinema; Sit down in the hall.	Choose a time for the movie; Leave the house; Enter the building; Exit the cinema hall; Take your car to go home.

3. Attend a wedding – from getting ready, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Get ready/get dressed; Go to the venue; Attend the function; Go to the restaurant; Eat/attend the lunch or dinner; Return home.	Take your wallet/purse/wedding gift; Leave the house; Party/dance.	Meet up/greet friends; Congratulate the bride and groom; Say goodbye.

4. Shop for groceries – from the decision to go, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Write down a shopping list; Go to the grocery store; Get a shopping trolley; Shop/choose the products; Get in the check out line/queue; Pay; Return home.	Decide to go grocery shopping; Look in fridge/pantry to see what you need; Take your wallet/key/phone; Leave the house; Enter the grocery store; Put your shopping in shopping bags.	Get ready to go out; Take your shopping bags; Park your car; Wander the aisles; Place your shop on the conveyor belt; Put your shopping in the car; Exit the grocery store; Put the trolley back in its place.

5. Going to the doctors – from when you book the appointment, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Book an appointment; Go to the doctor's office; Sit in the waiting room/wait for your turn; Have the check-up/get examined; Return home.	Call your doctor; Get ready to go out; Leave the house; Leave the doctor's office.	Take your wallet/key/phone; Enter the doctor's office; Check-in at the doctor's office; Talk to the doctor/explain your symptoms; Get a diagnosis/get medication; Pay; Say goodbye when leaving.

6. Taking a shower – from the decision to have a shower, to when you are dry and dressed.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Decide to shower; Turn the water on; Undress; Enter the shower; Wash yourself up/soap and rinse; Exit the shower; Use the bathrobe/towel; Dry yourself; Dry your hair; Get dressed.	Go in the bathroom; Wait for hot water; Wash your hair/shampoo and rinse.	Get wet; Turn the water off.

7. Christmas day – from waking up to going to sleep.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Wake up; Have breakfast; Open presents; Have lunch; Go to sleep.	Get ready to go out/get dressed; Exchange Christmas greetings; Have dinner.	Cook; Go to the family Christmas lunch; Play bingo/play games; Say goodbye to relatives; Return home.

8. Get ready for work/school – from waking up to arriving at work/school.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Wake up/get up; Wash up/wash your face/brush your teeth; Have breakfast; Get dressed; Leave the house; Arrive at work/school.	Go to the bathroom; Take your wallet/key/phone.	Hear the alarm clock; Sit down at your desk.

9. Shop for clothes – from the decision to go, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Decide to go shopping; Try the clothes on; Check-out/queue at the check-out line; Pay; Return home.	Take your wallet/key/phone; Leave the house; Go to the shops; Enter the shop; Choose clothes to try on; Exit the shop.	Get dressed (to go out); Wander around different shops; Look at the clothes; Look for your size; Go to the dressing rooms; Look at your reflection/see whether the clothes fit well; Choose which clothes to buy.

10. Going a day at the swimming pool - from the decision to go, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Decide to go to the swimming pool; Take your swimming bag/towels/swim cap/key/wallet; Leave the house; Get changed/put swimsuit on; Go to the pool; Swim/do laps; Have a shower before leaving the pool; Return home.	Dive/get in the water; Get out the water; Get dry and dressed after swimming. .	Get dressed (to go out); Pay at the entrance; Enter the building; Sunbathe/relax on the deckchair.

11. Go to the hairdresser/barber – from the decision to go, to coming back home.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Decide to go to the hairdresser/barber; Set an appointment; Get to the salon; Get a haircut/wash/hair dye; Pay; Return home.	Call the salon; Get ready to go out; Leave the house; Enter the salon; Wait for your turn; Explain/request a haircut/blow-dry/hair dye.	Take your key/wallet/phone; Exit the salon.

12. Have a minor car accident (*e.g.* a rear ending) – from the collision to when you drive away after having dealt with the other person.

Major (≥65%)	Minor (45-64%)	Trivial (25-44%)
Be rear-ended; Get out your car; Assess the damage; Reach an agreement with the other person; Fill insurance documents (italian "CID"); Get back in your car; Drive away.	Stop/pull over; Make sure nobody is injured; Exchange personal info with the other driver.	Get the documents to fill; Talk with the other driver.

Appendix B

Potts model details

A Potts neural network is an autoassociative memory network comprised of N Potts units, which model patches of cortex as they contribute to retrieve distributed long-term memory traces addressed by their contents (Treves, 2005). Each Potts unit has S active states, indexed as 1, 2, · · ·, S, representing local attractors in that patch, and one quiet state, the 0 state. The N units interact with each other via tensor connections, that represent associative long-range interactions through axons that travel through the white matter (Braitenberg & Almut Schuz, 1991), while local, within-grey-matter inter actions are assumed to be governed by attractor dynamics in each patch. The values of the tensor components are pre-determined by the Hebbian learning rule, which can be construed as derived from Hebbian plasticity at the synaptic level (Naim et al., 2018).

$$J_{ij}^{kl} = \frac{c_{ij}}{c_m a(1-\frac{a}{S})} \sum_{\mu=1}^{p} \left(\delta_{\xi_i^\mu k} - \frac{a}{S}\right)\left(\delta_{\xi_j^\mu l} - \frac{a}{S}\right)(1-\delta_{k0})(1-\delta_{l0}), \tag{6}$$

where c_{ij} is either 1 if unit j gives input to unit i or 0 otherwise, allowing for asymmetric connections between units, and the δ's are the Kronecker symbols. The number of input connections per unit is c_m. The p distributed activity patterns which represent memory items are assigned, in the simplest model, as composition of local attractor states $\{\xi^\mu{}_i\}$ (i = 1, 2, · · ·, N and μ = 1, 2, · · ·, p). The variable $\xi^\mu{}_i$ indicates the state of unit i in pattern μ and is randomly sampled, independently on the unit index i and the pattern index μ, from {0, 1, 2, · · ·, S} with probability

$$P(\xi_i^\mu = k) = \frac{a}{S}(1-\delta_{k,0}) + (1-a)\delta_{k,0}. \tag{7}$$

Constructed in this way, patterns are randomly correlated with each other. We use these randomly correlated memory patterns $\{\xi^{\mu}_i\}_{\mu=1...p}$ in this study. The parameter a is the sparsity of patterns – fraction of active units in each pattern; the average number of active units in any pattern μ is therefore given by Na.

Local network dynamics within a patch are taken to be driven by the "current" that the unit I in state k receives

$$h_i^k(t) = \sum_{j \neq i}^{N} \sum_{l=1}^{S} J_{ij}^{kl} \sigma_j^l(t) + w \left[\sigma_i^k(t) - \frac{1}{S} \sum_{l=1}^{S} \sigma_i^l(t) \right], \tag{8}$$

where the local feedback w, introduced in Russo & Treves (2012), models the depth of attractors in a patch, as shown in Naim et al. (2018) - it helps the corresponding Potts unit converge to its most active state. The activation along each state for a given Potts unit is updated with a *soft max* rule

$$\begin{aligned} \sigma_i^k(t) &= \frac{\exp[\beta r_i^k(t)]}{\sum_{k=1}^{S} \exp[\beta r_i^k(t)] + \exp\{\beta[U + \theta_i^A(t) + \theta_i^B(t)]\}} \quad \text{if } k > 0, \\ \sigma_i^0(t) &= \frac{\exp\{\beta[U + \theta_i^A(t) + \theta_i^B(t)]\}}{\sum_{k=1}^{S} \exp[\beta r_i^k(t)] + \exp\{\beta[U + \theta_i^A(t) + \theta_i^B(t)]\}} \quad \text{if } k = 0, \end{aligned} \tag{9}$$

where U is a fixed threshold common for all units and β is an effective inverse "temperature" (noise level). Note that σ_i^k takes continuous values in (0, 1) and that $\sum^S{}_{k=0} \sigma_i^k = 1$ for any i. The variables r_i^k, θ_i^A and θ_i^B parameterize, respectively, the state-specific potential, fast inhibition and slow inhibition in patch i. The state-specific potential r_i^k integrates the state-specific current h_i^k by

$$\tau_1 \frac{dr_i^k(t)}{dt} = h_i^k(t) - \theta_i^k(t) - r_i^k(t), \tag{10}$$

where the variable θ_i^k is a specific threshold for unit *i* and for state *k.*

Taking the threshold θ_i^k to vary in time to model adaptation, *i.e.* synaptic or neural fatigue selectively affecting the neurons active in state *k*, and not all neurons subsumed by Potts unit *i*

$$\tau_2 \frac{d\theta_i^k(t)}{dt} = \sigma_i^k(t) - \theta_i^k(t), \tag{11}$$

the Potts network additionally expresses latching dynamics, the key to its possible role in modelling temporal schemata.

The unit-specific thresholds θ_i^A and θ_i^B describe local inhibition, which in the cortex is relayed by at least 3 main classes of inhibitory interneurons (Tremblay et al., 2016) acting on $GABA_A$ and $GABA_B$ receptors, with widely different time courses, from very short to very long. Formally in our model, θ_i^A denotes fast, $GABA_A$ inhibition and θ_i^B denotes slow, $GABA_B$ inhibition and they vary in time in the following way:

$$\tau_A \frac{d\theta_i^A(t)}{dt} = \gamma_A \sum_{k=1}^{S} \sigma_i^k(t) - \theta_i^A(t), \tag{12}$$

$$\tau_B \frac{d\theta_i^B(t)}{dt} = (1 - \gamma_A) \sum_{k=1}^{S} \sigma_i^k(t) - \theta_i^B(t), \tag{13}$$

where one sets $\tau_A < \tau_1 \ll \tau_2 \ll \tau_B$ and the parameter γ_A sets the balance of fast and slow inhibition. Specifically in this work, we set these parameters as $\tau_A = 10$, $\tau_B = 10^5$, $\tau_1 = 20$ and $\gamma_A = 0.5$.

Appendix C

Simulation details

We have used an asynchronous updating, where one unit is updated at a time with a random order. Updating all Potts units in the network once is our measuring unit of simulation time: all timescales of the model are measured with this unit. We stop the simulation after updating the entire network 10000 times (except for Fig. 5). Then, we cut out the first 3 patterns in the sequence to remove the effect of initialization. Every stored memory is used as a cue with its full representation. In order to compute the probability $P^{\mu\nu}{}_{\gamma}(z)$ in Eq. (5), we have run $p \times 1000$ simulations for each condition. For each memory pattern, we take 40% of its active units and flip them into different states. We prepare 1000 corrupted versions of each memory by repeating this procedure 1000 times. Each of these corrupted versions is used as a cue in each simulation, which is terminated after 12 transitions. Unless specified explicitly, parameters of the Potts model are set as in Table 1.

Symbol	Meaning	Default value
N	number of Potts units	256
c_m	number of presynaptic units	50
S	number of states per unit	7
p	number of memory patterns	50
a	sparsity of patterns	0.25
λ	relative coupling strength	0.5
U	global threshold	0.1
τ_2	adaptation timescale	200
w	self-reinforcement term	1.1
β	inverse "temperature"	11

Table 1. Parameters of the network