An Integrative Memory model of recollection and familiarity to understand memory deficits

Christine Bastin, Gabriel Besson, Jessica Simon, Emma Delhaye, Marie Geurten, Sylvie Willems, & Eric Salmon


Jessica Simon, Psychology and Neuroscience of Cognition, University of Liège, Belgium, j.simon@uliege.be, https://www.psyncog.uliege.be


Marie Geurten, GIGA-Cyclotron Research Centre in vivo imaging & Psychology and Neuroscience of Cognition, University of Liège, Belgium, mgeurten@uliege.be, https://www.psyncog.uliege.be
Sylvie Willems, Psychological and speech therapy consultation center & Psychology and Neuroscience of Cognition, CPLU, University of Liège, Belgium, sylvie.willems@uliege.be, https://www.psyncog.uliege.be


Corresponding author: Christine Bastin, GIGA-Cyclotron Research Centre in vivo imaging, Liège University, Allée du 6 Août, B30 Sart Tilman, 4000 Liège, Belgium, Telephone: 32 4 366 23 69, Fax: 32 4 366 29 46, Email: Christine.Bastin@uliege.be
Short abstract

An Integrative Memory model is proposed to understand the organization of episodic memory and to guide future investigation on the nature of deficits in memory-impaired populations. By integrating currently most influential dual-process views about episodic memory, we describe a neurocognitive architecture of representations and operations underlying recollection and familiarity. Considering the case of Alzheimer’s disease as an application, our framework provides novel predictions about the specific memory mechanisms that could be affected in the course of the pathology.

Long abstract

Humans can recollect past events in details (recollection) and/or know that an object, person or place has been encountered before (familiarity). During the last two decades, there has been intense debate about how recollection and familiarity are organized in the brain. Here, we propose an Integrative Memory model which describes the distributed and interactive neurocognitive architecture of representations and operations underlying recollection and familiarity. In this architecture, the subjective experience of recollection and familiarity arises from the interaction between core systems storing particular kinds of representations shaped by specific computational mechanisms and an attribution system. By integrating principles from current theoretical views about memory functioning, we provide a testable framework to refine the prediction of deficient versus preserved mechanisms in memory-impaired populations. The case of Alzheimer’s disease is considered as an example because it entails progressive lesions starting with limited damage to core systems before invading step-by-step most parts of the model-related network. We suggest a chronological scheme of cognitive impairments along the course of Alzheimer’s disease, where the inaugurating deficit would
relate early neurodegeneration of the perirhinal/anterolateral entorhinal cortex to impaired familiarity for items that need to be discriminated as viewpoint-invariant conjunctive entities. The Integrative Memory model can guide future neuropsychological and neuroimaging studies aiming to understand how such a network allows humans to remember past events, to project into the future and possibly also to share experiences.

Keywords: Alzheimer’s disease, cerebral network, dual-process models of recognition memory, episodic memory, familiarity, fluency, hippocampus, perirhinal cortex, posterior cingulate cortex, recollection.
An Integrative Memory model of recollection and familiarity to understand memory deficits

Episodic memory allows us to remember objects and people that we have encountered as well as details about events that we have personally experienced. It gives us awareness of our past experience, it is crucial to a smooth functioning in our daily life and it permits that we mentally project what might subsequently happen on the basis of our past memories (Tulving 1999). Unfortunately, episodic memory is fragile and can be disrupted by certain conditions. Some people experience memory impairments (amnesia) suddenly after an acute brain damage. Others experience a progressive memory decline because of a neurodegenerative pathology like Alzheimer’s disease.

The understanding of episodic memory mechanisms and how they are implemented in the brain has progressed extensively thanks to research in neuropsychology and neuroimaging. Current theories posit that episodic memories can be retrieved via two processes: recollection which designates the recall of the specific details from the initial experience of the events, including details about the spatiotemporal context, and familiarity which refers to knowing that one has experienced something in the past without recalling details about the encoding episode (Mandler 1980; Tulving 1985; Yonelinas 1994).

In the current article, we will first define the processes of recollection and familiarity in psychological terms. Then, we will summarize the current most influential frameworks that describe their neural substrates. The existing frameworks differ by their focus on cognitive operations versus type of representations, by the emphasis on a specific brain region versus neural systems, and by the assumption that recollection and familiarity processes are localized to a brain region or not. Next, we consider that a more complete understanding of recollection
and familiarity would benefit from combining different accounts into a unified framework that bridges several cognitive and neural mechanisms. So, we propose an integration of principles pertaining to currently separate theories in a neurocognitive architecture of interacting operations and representations within large-scale cerebral networks that allow familiarity and recollection. Such integrative perspective allows to generate new hypotheses about the nature of memory deficits in brain-lesioned populations and neurodegenerative diseases. The last section thus presents predictions relative to recollection and familiarity deficits in memory-impaired populations, with a detailed illustration on Alzheimer’s disease.

1. Recollection and familiarity

In psychological terms, recollection is defined as a retrieval process whereby individuals recall detailed qualitative information about studied events (Montaldi & Mayes 2010; Yonelinas et al. 2010). Some authors consider that there is recollection as soon as one retrieves at least one detail that is not currently perceived, inducing moderate to high confidence that the event actually occurred (Higham & Vokey 2004; Yonelinas et al. 2010), but the amount of details may vary from one trial to the other (Higham & Vokey 2004; Parks & Yonelinas 2007; Wixted & Mickes 2010). These associated details typically represent the context in which an event took place (i.e., place, time, environmental or internal details) (Ranganath 2010). Recollection can be accompanied by a subjective experience of mentally reliving the prior experience with the event, as if one were mentally traveling back in time to re-experience it (Tulving 1985).

In contrast, familiarity is a feeling of oldness indicating that something has been previously experienced. It is thought to support predominantly recognition of single pieces of information (i.e., items such as objects and people) (Ranganath 2010), but associations
between similar types of information could also be recognized as familiar (Mayes et al. 2007). Subjectively, feelings of familiarity are more or less strong feelings that one knows that something has already been encountered, leading to varying degrees of confidence (Tulving 1985; Yonelinas et al. 2010). According to some theories, the feeling of familiarity arises when one interprets enhanced processing fluency of a stimulus as a sign that it was previously encountered (Jacoby et al. 1989; Whittlesea et al. 1990). Fluency is typically defined as the speed and ease with which a stimulus is processed and may arise from many sources (e.g., mere repetition, perceptual clarity, rhyme, predictive context, oral-motor sequence), including past occurrences (Oppenheimer 2008; Reber et al. 2004; Topolinski 2012; Unkelbach & Greifeneder 2013). Because people intuitively know from their earliest years that fluently processed items are more likely to have been encountered previously, a feeling of fluency during a memory task will be likely interpreted as related to prior exposure (Schwarz 2004). However, several conditions have to be fulfilled for fluency to be used to guide memory. First, fluency has to be judged as a diagnostic cue for memory (Westerman et al. 2002). Second, the experienced fluency has to be greater than expected in a given context (i.e., individuals have to be surprised by the ease with which they are able to process an item) and should not be attributed to a more plausible source (e.g., the intrinsic perceptual quality of the stimulus) than past occurrence. Thus, if people appraise past encounter as an improbable source of fluency or if a more plausible source is detected, individuals will disregard fluency as a relevant cue for recognition decisions (Kelley & Rhodes 2002; Miller et al. 2008; Willems & Van der Linden 2006). This disqualification will prevent fluency to give rise to a feeling of familiarity.

Neuropsychological investigation of recollection and familiarity in memory-impaired populations (e.g., normal aging, amnesia, epilepsy, neurodegenerative diseases) as well as
neuroimaging studies examining the neural correlates of recall and recognition memory tasks (using mainly functional magnetic resonance imaging, fMRI) have provided a huge corpus of data that have led to the development of neurocognitive models of episodic memory functioning. Most memory models focus on the role of the medial temporal lobe (MTL) in recollection and familiarity, since seminal neuropsychological work has shown that amnesia arises following MTL damage (Scoville & Milner 1957). Much controversy still surrounds the precise contributions of the different MTL subregions, most notably the hippocampus and the adjacent perirhinal and entorhinal cortices. With the exception of unitary models suggesting that MTL structures contribute to both recollection and familiarity as a function of memory strength (Squire et al. 2007; Wixted & Squire 2011), the majority of models suggest that there is fractionation of memory processes in the MTL by reference to recollection and familiarity. These MTL models can be distinguished as a function of whether they define the role of the hippocampus and adjacent MTL cortices in terms of putative cognitive operations or according to the nature of representations. Most frameworks target the role of anatomical regions (and their functional network), but a few speak at the scale of individual neurons or populations of neurons within a brain region.

MTL process models. These models propose that the different MTL regions have distinct computational properties (Montaldi & Mayes 2010; Norman & O'Reilly 2003). In particular, only the hippocampus is capable of pattern separation (to create distinct memory representations for similar inputs) and pattern completion (once the hippocampus has bound the elements of an episode into a memory trace, subsequent experience of a subset of the elements causes the remaining elements to be reactivated by association). Thanks to these properties, the hippocampus is specialized for recollection of details. In contrast, the perirhinal and parahippocampal cortices extract statistical regularities in repeated inputs by
creating sharper patterns. By contrast with novel inputs that activate weakly a large pattern of units, the sharpness of MTL cortical patterns indexes familiarity (Norman & O'Reilly 2003). The perirhinal cortex would thus encode similarities between events (LaRocque et al. 2013) and support familiarity. At the scale of neurons, some models describe familiarity signals as resulting from decreased firing of perirhinal neurons for repeated stimuli (Bogacz & Brown 2003; Bogacz et al. 2001; Sohal & Hasselmo 2000). This would arise because the number of active neurons that responded to a novel stimulus reduces as the stimulus becomes familiar.

**MTL representational models.** These models emphasize the different kinds of information incorporated in representations formed in the hippocampus versus the parahippocampal region (Aggleton & Brown 1999; Davachi 2006; Eichenbaum et al. 2007; Ranganath 2010). Whereas the perirhinal and parahippocampal cortices encode specific constituent elements of an event (e.g., objects, spatial layout), the hippocampus encodes representations of the relationships between the elements. According to the Binding of Item and Context model (Diana et al. 2007; Ranganath 2010), the perirhinal cortex and parahippocampal cortex encode respectively item and context information and the hippocampus encodes representations of item–context associations. Retrieval of item representations in the perirhinal cortex can support familiarity, while context representations and item–context bindings support recollection. As in MTL process models, the hippocampus is important for recollection, but these views consider that the parahippocampal cortex is also important for recollection because it represents contextual information.

**The representational-hierarchical models.** Recently, there has been accumulating evidence that the MTL mediates processes beyond long-term episodic memory. It is also involved in perception and short-term memory. In this view, the role of the MTL would be best described
in terms of how each region represents information rather than in terms of a specific process (Cowell et al. 2006; Graham et al. 2010; Saksida & Bussey 2010). Actually, the MTL is considered an extension of the representational hierarchy of object processing within the ventral visual stream. The complexity of representations increases from posterior occipital areas to the anterior lateral and medial temporal regions. The perirhinal cortex represents the culmination of this object processing pathway, performing the most complex feature computations required to discriminate objects with a high degree of visual feature overlap. In a memory task, the perirhinal cortex can differentiate between objects that share features. Most recent suggestions also posit that the capacity of the perirhinal cortex to distinguish between overlapping item representations makes it a critical region to disambiguate conceptual entities with shared properties, such as living objects (Clarke & Tyler 2015; Inhoff & Ranganath 2015), in various tasks such as naming or recognition memory. As for the hippocampus, its function goes beyond object processing, as it represents relational configurations and scenes that can support performance in a variety of tasks, such as perceptual discrimination of scenes, navigation, imagination, source memory… (Clark & Maguire 2016; Cowell et al. 2010). So, this theoretical approach does not map recollection and familiarity onto specific regions. The role of MTL subregions are rather defined in terms of the type and complexity of representations they contain and all could generate familiarity and recollection (Cowell et al. 2010).

In all these models, the role of another region of the MTL, the entorhinal cortex, is poorly specified. The entorhinal cortex receives the inputs and outputs of other MTL regions, but its anterolateral and posteromedial parts appear to belong to different systems. Indeed, it has been suggested that the anterolateral entorhinal cortex may have functional specialization similar to the perirhinal cortex, whereas the posteromedial entorhinal cortex would support
the same function as the parahippocampal cortex (Keene et al. 2016; Maass et al. 2015; Schultz et al. 2012). Moreover, investigation of connection pathways in the MTL suggests that the hippocampus should not be treated as a unitary region, but has distinct connectivity preference along its anterior-posterior portions and as a function of its subfields (Aggleton 2012; Libby et al. 2012). The perirhinal cortex has preferential connection with anterior CA1 and subiculum, whereas the parahippocampal cortex connects more with the posterior CA1/CA2/CA3/dentate gyrus and subiculum.

**Whole-brain network models.** However, the MTL is not the only region that contribute to recollection and familiarity. As notably evidenced by neuroimaging studies, recollection also involves the posterior cingulate cortex, the retrosplenial cortex, the inferior parietal cortex, the medial prefrontal cortex, anterior nuclei of the thalamus and mammillary bodies (Aggleton & Brown 1999; Ranganath & Ritchey 2012). This network has been labelled the general recollection network (Rugg & Vilberg 2013). The extended cerebral network for familiarity involves, beside the perirhinal cortex, the ventral temporal pole, the dorsolateral prefrontal cortex, the dorsomedial nuclei of the thalamus, and the intraparietal sulcus (Johnson et al. 2013; Kim 2010; Ranganath & Ritchey 2012). Currently, very few theoretical models of recollection and familiarity have integrated these large-scale cerebral memory networks. Recently, however, Ranganath and colleagues (Ranganath & Ritchey 2012; Ritchey et al. 2015) revised the Binding of Item and Context model to suggest that the MTL regions are actually part of two broad memory systems. The perirhinal cortex is considered as a core component of an extended Anterior Temporal system that also includes the ventral temporopolar cortex, lateral orbitofrontal cortex and amygdala. This system may be essential for processing entities (that is, people and things), and would be involved in item familiarity. In contrast, the parahippocampal cortex is considered as core component of an extended
Posterior Medial network that includes the mammillary bodies and anterior thalamic nuclei, presubiculum, the retrosplenial cortex, and the default network (comprising the posterior cingulate cortex, precuneus, lateral parietal cortex and medial prefrontal cortex). It would be involved in tasks that require a mental representation of the relationships between entities, actions and outcomes, such as recollection-based memory tasks. Such models considering the whole-brain network architecture of memory processes are critical given the fundamentally interconnected nature of brain structures.

Currently, yet, some aspects of recollection and familiarity have not been fully integrated in memory models. In particular, current models do not encompass the notion that explicit memory judgments and experiences, such as feelings of remembering and familiarity, arise from attribution mechanisms that interpret memory signals, such as fluency cues (Voss et al. 2012; Whittlesea 2002) and take into account expectations in a particular context (Bodner & Lindsay 2003; McCabe & Balota 2007; Westeman et al. 2002). A line of research considers how feelings of familiarity emerge when previous exposure to some information induces a sense of facilitated processing (i.e., fluency feeling) that is attributed to past occurrence of the information (Westeman et al. 2002; Whittlesea & Williams 2001a, 2001b). Similarly, both fluency signals and attribution mechanisms may also contribute to the experience of recollection (Brown & Bodner 2011; Li et al. 2017; McCabe & Balota 2007).

Here, we propose to integrate the current state of knowledge about the neurocognitive bases of recollection and familiarity by incorporating, into a single model, separate lines of research, namely neural models of recollection and familiarity and attributional models of memory experiences. This Integrative Memory model builds on currently most influential dual-process views of the cognitive and neural bases of recollection and familiarity and takes
into account the highly interconnected nature of the human brain in order to propose a
distributed and interactive neurocognitive architecture of representations and operations
underlying recollection or familiarity.

2. The Integrative Memory model: A neurocognitive architecture of recollection and
familiarity

The notion of recollection and familiarity has been used to refer to processes and subjective
experiences, leading sometimes to confusion between these aspects. In the Integrative
Memory model (Figure 1), we describe recollection and familiarity as the interaction between
core systems that store specific types of representations uniquely shaped by specific
computational operations and make up the content of the memory and an attribution system
framed by the task context that translates content reactivation into a subjective experience.
Recollection emerges preferentially from reactivation of traces from a relational
representation core system, whereas familiarity emerges mainly from reactivation of traces
from the entity representation core system. The distinction between core systems and an
attribution system has two implications. First, the core systems build the memory trace and
damage to these systems induces severe degradation of the content of the memory. In
contrast, the attribution system modulates the use of memory traces as a function of
expectancies, task context and goals, leading to subjective experiences and explicit
judgments. Lesion of the attribution system affects mainly the quality and adequation of the
memory output to the task at hand. Second, although most memory situations generate an
explicit output that matches the content of the memory (e.g., recollection follows reactivation
of a relational representation), this might not always be the case. This means that the
qualitative and subjective experience that one has in a given memory task may dissociate
from the memory reconstructed by a core system. For instance, even if the relational
representation core system reactivates specific item-context details, one may experience a feeling of familiarity. This is because explicit outputs during a memory task (i.e., old/new decisions, confidence judgments and subjective experiences of remembering or knowing) follow from processing the outputs of the relational or entity representation core system in an attribution system. We assume that the attribution mechanisms are common down-stream mechanisms that serve both recollection and familiarity. In this framework, recollection and familiarity are considered as independent processes, in the sense that the underlying memory representation can be retrieved via the entity representation core system only, the relational representation core system only, or via both concomitantly (Jacoby et al. 1997).

Figure 1. Integrative Memory model. OC: occipital cortex; OTC: occipito-temporal cortex; PrC: perirhinal cortex; aTC: anterior temporal cortex; aERC: anterolateral entorhinal cortex; PhC: parahippocampal cortex; OPC: occipito-parietal cortex; pmERC: posteromedial entorhinal cortex; anThal: anterior nuclei of the thalamus; MB: mamillary bodies; RsC:
restrosplenial cortex; vPCC: ventral posterior cingulate cortex; dPCC: dorsal posterior cingulate cortex; oPFC: orbital prefrontal cortex; (v)mPFC: (ventro)medial prefrontal cortex; vPC: ventral parietal cortex; dLPFC: dorsolateral prefrontal cortex; dPC: dorsal parietal cortex; vLPFC: ventrolateral prefrontal cortex; dmThal: dorsomedial nuclei of the thalamus.

2.1 Detailed description of the Integrative Memory model

2.1.1 Encoding

Core systems are specialised for encoding and storing specific kinds of representations. The nature of the information that is processed in each core system is determined by the computational operations and level of associativity that characterise its constituent brain regions. Although each core system must be viewed as a representation system rather than as harboring recollection or familiarity processes, we suggest that recollection and familiarity are preferentially associated with specific types of representations: relational representations (centered on the hippocampus) for recollection, and entity representations (centered on the perirhinal cortex) for familiarity. Consistently, fMRI studies examining encoding-related activities observed that hippocampal activity is predictive of subsequent source recollection but uncorrelated with item recognition, and that perirhinal activity predicts item/familiarity-based recognition, but not subsequent recollection (Davachi et al. 2003; Kensinger & Schacter 2006; Ranganath et al. 2004). Recollection of details from the initial experience of an event also relies usually on contextual information that is stored in a context representation core system, but, as detailed below, some contextual tagging of entities occurs and elements of context (e.g., a building) may be subsequently recognized as familiar. Finally, the notion that these objects, people and events have been personally experienced is recorded by the interaction between representation core systems and a self-referential system.
In the *entity representation core system*, encountered entities pertaining to experienced events are encoded. An entity is defined as an exemplar item (i.e., token) from a category (i.e., type) that distinguishes itself from other similar items thanks to its unique configuration of perceptivo-conceptual features. The *entity representation core system* comprises the perirhinal cortex, anterolateral enthorinal cortex, occipitotemporal cortex and anterior temporal cortex.

Of note, even if the entorhinal cortex has a hierarchically higher level of associativity than the perirhinal cortex (Lavenex & Amaral 2000) and recent data speak for a specific role of the anterolateral entorhinal cortex in object-in-context processing (Yeung et al. in press), there are currently not sufficient data to clearly distinguish the role of the perirhinal cortex and the anterolateral entorhinal cortex. Based on studies showing a role for the anterolateral entorhinal cortex in disambiguation of similar objects (Yeung et al. 2017), we will consider here that the perirhinal cortex and anterolateral entorhinal cortex form together a system specialized for entity representation. This system is dedicated to the processing and encoding of single entities (Ranganath & Ritchey 2012), with preferential representation of objects and faces (Kafkas et al. 2017; Martin et al. 2016), unified associations (Haskins et al. 2008) and pairings of similar entities (e.g., two faces) (Hirabayashi et al. 2013; Mayes et al. 2007). It has been suggested to additionally represent the association of a written concrete word with its corresponding object concept (Bruffaerts et al. 2013; Liuzzi et al. 2015).

Critically, the *entity representation core system* is defined by the nature and complexity of the representations it can process and encode for long-term memory after a single exposure to the stimulus. More specifically, in line with the representational-hierarchical view (Cowell et al. 2010; Graham et al. 2010; Saksida & Bussey 2010), there is a hierarchy in terms of the complexity of the representation in the *entity representation core system*. Consider here the example of object processing (Figure 1). While individual features (e.g., shape, texture, color)
are processed in ventral occipitotemporal areas (visual object features), integration of these features into more and more complex entities are achieved as one moves anteriorly along the ventral visual stream. It is at the level of the perirhinal cortex and anterolateral entorhinal that all visual features are integrated in a single complex representation of the object that can be discriminated from other objects with overlapping features. Moreover, the perirhinal cortex may also act as a conceptual binding site. Whereas defining conceptual features such as the category are represented in the anterior temporal areas, the integration of the meaning to object representations will occur in the perirhinal cortex via its interaction with the anterior temporal area (conceptual features) (Martin et al. 2018; Price et al. 2017; Taylor et al. 2011). Indeed, the perirhinal cortex is notably recruited when concepts with confusable features must be distinguished (Clarke & Tyler 2015). For instance, the perirhinal cortex is needed to distinguish between living things during naming (and recognition memory tasks), as living things share a lot of common features and are more easily confusable than non-living things (Kivisaari et al. 2012; Wright et al. 2015). By incorporating features from various sensory and conceptual areas, the perirhinal/anterolateral entorhinal cortex forms unique conjunctive representations of entities allowing to resolve ambiguity in the face of objects with overlapping features and to identify objects in a viewpoint-invariant manner (Erez et al. 2016). These representations rely on a computational property of the perirhinal/anterolateral entorhinal cortex that can be referred to as entity pattern separation, by which similar objects are given separate representations based on specific conjunctions of features, even after a single exposure (Kent et al. 2016). This property allows humans to quickly recognize familiar objects in the stream of resembling objects from the environment.

Given that entities are typically experienced as part of an event, the perirhinal/anterolateral entorhinal cortex also encodes the significance of entities in a context-dependent manner
This is possible thanks to the connections between the perirhinal cortex and the parahippocampal/posteromedial entorhinal cortex, which is part of the context representation core system together with the occipitoparietal cortex and retrosplenial cortex. The parahippocampal cortex represents preferentially buildings and scenes, which often constitute the contextual setting for an event (Bar et al. 2008; Kafkas et al. 2017; Martin et al. 2013; Preston et al. 2010) and the posteromedial entorhinal cortex encodes an internally generated grid of the spatial environment (Doeller et al. 2010). The context representation core system would provide a contextual tagging of the entity, that allows to take into account the background into which the entity occurred and give distinct meanings and values to the entity.

In their article, Inhoff and Ranganath (2015) gave the example of a ticket purchased at a county fair to buy food and rides, whose significance changes beyond the fairgrounds because that same ticket would have little value outside the fair. In addition, we recognize entities that we have personally experienced. Self-reference is also important to define the significance of entities. Via connections of the perirhinal cortex to the orbital prefrontal cortex (Lavenex et al. 2002), the entity representation may also record the self-relatedness of the entity (D'Argembeau et al. 2005; Northoff et al. 2006). Like the contextual significance, self-relatedness of entities may modulate our behavior with regard to the entities. For example, a piece of clothing should lead to different behaviors depending on whether it belongs to me or somebody else.

In brief, entities encountered as part of experienced events are stored in long-term memory in a distributed and hierarchical manner in the entity representation core system. While simple perceptual and conceptual features are represented in occipitotemporal and anterior temporal areas, the conjunctions of multi-modal features are represented as pattern-separated entities.
the perirhinal cortex and the anterolateral entorhinal cortex. Some contextual and self-relatedness tagging via interactions between the entity representation core system and the context representation core system and self-reference system will modulate the significance of entities. The concept of unification is close to the notion of conjunction, with the difference that unification can sometimes be an active encoding strategy whereas conjunction refers to the configurational nature of stimuli. Indeed, unification consists in encoding different pieces of information in a way that integrates them into a single entity (Parks & Yonelinas 2015).

Previous fMRI studies have shown that processing object-color associations by mentally integrating color as an object feature activates the perirhinal cortex (Diana et al. 2010), as does the encoding of word pairs as new compound words (Haskins et al. 2008).

The relational representation core system involves the hippocampus, subiculum, mamillary bodies and the anterior nuclei of the thalamus. It rapidly encodes a detailed representation of the item bound to associated contextual information (Montaldi & Mayes 2010; Ranganath & Ritchey 2012) or more generally complex high-resolution bindings (Yonelinas 2013). In the case of item-context binding, inputs consist in the entity representations from the perirhinal/anterolateral entorhinal cortex entering the hippocampus anteriorly, and context representations (e.g., spatial layout) from the parahippocampal/posteromedial entorhinal cortex entering the hippocampus posteriorly (Ranganath & Ritchey 2012; Staresina et al. 2011). The context representation in the parahippocampal cortex is itself fed by inputs from neocortical regions that represent the specific contents of the context in which the item is embedded (e.g., sounds, visual details, spatial layout…) stored in occipitoparietal sites (visuospatial processing) (Rissman & Wagner 2012), brought to the parahippocampal cortex via the retrosplenial cortex. The self-referential nature of the experienced episodes is also embedded in the memory trace thanks to connection of the hippocampus and retrosplenial...
cortex with the ventromedial prefrontal cortex (Andrews-Hanna et al. 2010). The binding of multimodal and qualitatively different pieces of information occurs in the hippocampus (CA3 via the dentate gyrus) where each unique episode is encoded as a separate representation via \textit{relational pattern separation} (Berron et al. 2016; Leal & Yassa 2018; Montaldi & Mayes 2010; Norman & O'Reilly 2003), so that two very similar events will have two distinct memory traces. For instance, if we attend to two concerts based on the same album of our favourite band, we will still be able to remember the details of each concert as a unique episode.

This pattern-separated representation in the hippocampus constitutes a summary, or an index, of the distributed neocortical representations of the specific details of the episodes (Teyler & Rudy 2007). Contrary to the conjunctive representations in the entity representation core system where components are fused in a frozen integrated trace, the hippocampal representation keeps components separate and flexibly bound (Eichenbaum 2017). This allows the learning of inferences between items that are indirectly related and subsequent flexible use of representations (Eichenbaum & Cohen 2014). So, relational binding and pattern separation are the core computational properties of the relational representation core system.

While the nature of the representations in the entity representation core system makes it specialized for rapidly signaling that objects, faces and simple combinations of those are known (i.e., familiarity judgments), the bound representations in the relational representation core system makes it specialized for reactivating the specific details of experienced events (i.e., recollection). In other words, familiarity and recollection are naturally emerging
processes from the way different brain regions represent the experienced world. But, as will be detailed next, the final explicit memory output will depend on the attribution system.

2.1.2 Retrieval

*Familiarity-based retrieval.* As illustrated in Figure 2, the typical sequence of operations leading to familiarity starts with the repetition of an encoded entity (Montaldi & Mayes 2010; Ranganath 2010; Voss et al. 2012). For instance, during a recognition memory test, target items are the replication of previously studied items. In our example of the processing of an object item, the repetition of the perceptual and/or conceptual features of the item triggers enhanced *processing fluency* (and reduced activity) in occipitotemporal and anterior temporal areas where these features were first processed (Reber 2013). Several fMRI studies also showed that enhanced processing fluency of items induces a reduction of activity in the perirhinal cortex that predicts familiarity-based memory (Dew & Cabeza 2013; Gonsalves et al. 2005; Meyer et al. 2010). Here, we make the novel hypothesis that the perirhinal and anterolateral entorhinal cortices are sensitive to the repetition of the actual conjunction of features that makes up the specific and viewpoint-invariant representation of the item, associated with a specific meaning, and thus generates enhanced *entity-level processing fluency* and can lead to familiarity for this entity. In addition, any region representing features of the previously encountered object can reactivate these specific features when reexposed to them and thus generates familiarity-based memory through fluency. So, perceptual and conceptual fluency for features arising in occipitotemporal and anterior temporal cortices can also generate familiarity for these features. The dominant type of signal that will contribute to familiarity depends on the characteristics of the memory task (Lanska et al. 2014; Lucas & Paller 2013; Taylor & Henson 2012). For instance, in a task where participants have to rapidly discriminate between old pictures of objects and new pictures of completely different objects
(e.g., Besson et al. 2015), reactivation of simple perceptual (e.g., a small grey fluffy object for
the picture of a grey kitten) or conceptual features (e.g., a feline) is sufficient to successfully
identify studied stimuli. In contrast, if old objects are mixed with very similar objects from the
same category (e.g., Yeung et al. 2013), accurate familiarity-based discrimination will rely on
the reactivation of the studied conjunctions of features. This implies that familiarity may arise
from different regions, depending on the materials (e.g., Kafkas et al. 2017) and demands of
the task, and that lesions to the perirhinal cortex will not necessarily affect all forms of
familiarity.

Figure 2. Main mechanisms supporting familiarity-based retrieval in the example of a lab-based
object recognition memory task with resembling targets and lures.

Besides fluency signals, other signals may also operate in recognition memory tasks. We
focus here on fluency signals because we wish to model recognition memory decisions that
allow to identify a specific stimuli as previously encountered. Item-specific discrimination is a
key property of familiarity in everyday life, as we adapt our behaviour to familiar unique
entities. For instance, we will speak to people we know, we will take our own cup to fetch
some coffee, we will pick up our coat among others in a cloakroom… For all these situations,
we propose that fluency-based familiarity is central. However, feelings of familiarity can arise
from many other sources. Some of them are non-memory, such as affective information
(Duke et al. 2014) or proprioceptive information (Fiacconi et al. 2016) that have been shown to generate a subjective sense of familiarity if manipulated in memory situations. Others are from the memory domain, but supports global matching or similarity judgments when a presented stimulus globally maps onto a stored representation (Norman & O'Reilly 2003). But even then, the involvement of fluency in the emergence of a feeling of familiarity through affective information, proprioceptive information or global matching could not be ruled out (Duke et al. 2014).

Still, whatever its source, enhanced processing fluency in itself is not sufficient to produce familiarity. It has been suggested that fluency only minimally contributes to memory decisions because some patients with amnesia demonstrate chance-level recognition memory (and so, no sign of familiarity), despite successfully completing priming tasks conducted on the same set of stimuli (priming being also driven by fluency) (e.g., Levy et al. 2004). In the same vein, enhancing processing fluency of some stimuli had only small influence on amnesic patients’ memory performance in some studies (Conroy et al. 2005; Verfaellie & Cermak 1999), while others found reliable improvement of recognition memory performance in amnesia following manipulation enhancing processing fluency (Keane et al. 2006). Such findings can be explained if one considers that the transformation of fluency signals into familiarity-based decisions involves complex cognitive and metacognitive mechanisms (Whittlesea & Williams 2000; Willems et al. 2007). Accordingly, our Integrative Memory model argues that one cannot explain familiarity-based memory decisions without considering the role of the attribution system.

Therefore, explicit familiarity judgments and the subjective feeling of familiarity result from attribution of fluency to the prior occurrence of the stimulus (via the attribution system).
The fluency heuristic relies on signal flow from the entity representation core system regions to the attribution system, via connections between the perirhinal cortex and the prefrontal cortex (mainly, orbitofrontal, medial and dorsolateral prefrontal areas (Aggleton & Brown 1999; Lavenex et al. 2002; Libby et al. 2012)). The mechanisms thought to intervene in the attribution system, such as metacognitive and monitoring operations, have been notably associated with the prefrontal cortex in the context of memory tasks (Chua et al. 2014; Henson et al. 1999). Direct involvement in the fluency heuristic comes from electrophysiological studies (i.e., event-related potentials) (Kurilla & Gonsalves 2012; Wolk et al. 2004), notably showing that the attribution of fluency to the past versus the disqualification of fluency as a memory cue was associated with late frontal potentials.

The fluency heuristic involves sophisticated monitoring and metacognitive mechanisms. First, the metacognitive knowledge (supported by medial prefrontal areas) that fluent processing is a sign of prior occurrence exists since childhood (Geurten et al. 2017; Olds & Westerman 2012; Oppenheimer 2008), but this metacognitive heuristic can be unlearned through regular encounter with memory errors, as this might be the case for patients with severe memory problems (Geurten & Willems 2017). Second, the characteristics of the specific task at hand will determine the relevance of using fluency signals. This is determined via several monitoring mechanisms, supported by dorsolateral prefrontal cortex and that may happen at a non-conscious level. Fluency cues will be used if they are expected as diagnostic cues for recognition decisions (Westerman et al. 2002) and if the experienced fluency is salient relative to the context (Jacoby & Dallas 1981; Westerman 2008). People set an internal criterion along the varying dimension of memory strength depending on the task specificities. A feeling of surprise is experienced when the intensity of the fluency signal exceeds this
criterion (Yonelinas et al. 2010). If no alternative source is detected to explain the intensity of this signal, fluency will be attributed to past occurrence and will give rise to a feeling of familiarity. If not, fluency will be disregarded and no feeling of familiarity will arise.

Such an explicit judgment of familiarity occurs when *top-down attention*, supported by the dorsal parietal cortex, is focused on recognition memory decisions. According to the Attention-to-Memory model (Cabeza et al. 2008; Ciaramelli et al. 2008), the dorsal parietal cortex allocates attentional resources to memory retrieval according to the goals of the rememberer, and is often involved in familiarity-based decisions because familiarity may induce low confidence. This is the case in recognition memory paradigms where participants must judge how familiar stimuli are, but this can also occur in daily life (e.g., judging the most familiar brand of an article at the supermarket in order to choose the one usually bought). Yet, this explicit expression of familiarity may be distinguished from the *subjective feeling* of familiarity. Although both often co-occur in memory tasks, so that a participant can gauge how strong is his/her feeling of familiarity during confidence judgments for example, a strong feeling of familiarity may sometimes arise outside of any memory task and capture attention in a bottom-up fashion. One typical example is the butcher-on-the-bus phenomenon where one is surprised by the involuntary strong feeling of knowing the person, albeit in the absence of any recollection.

To come back to the cases where amnesic patients failed to use fluency cues in recognition memory tasks despite preserved perceptual or conceptual fluency, a likely interpretation in the framework of the attribution system considers that this is due to changes in metacognitive knowledge and monitoring in amnesic patients compared to controls (Geurten & Willems 2017). More specifically, because of their continued experience of memory errors in everyday
life, amnesic patients may have modified their metacognitive knowledge so as to unlearn the fluency heuristic (Geurten & Willems 2017; Ozubko & Yonelinas 2014). Additionally, their expectations relative to the origin of fluency feelings may have adapted in a way that makes them readier to detect alternative sources to fluency (Geurten & Willems 2017). Altogether, this will lead them to disqualify fluency as a cue for memory decisions (Conroy et al. 2005; Ozubko & Yonelinas 2014; Verfaellie & Cermak 1999), unless other fluency sources are very difficult to detect (Keane et al. 2006).

In initial network models (Aggleton & Brown 1999), the dorsomedial nucleus of the thalamus has been considered as a node within the familiarity system. However, its critical involvement remains unclear because of the divergence of findings relative to a selective impairment of familiarity following lesion to the dorsomedial thalamus (Danet et al. 2017; Edelstyn et al. 2016). Theoretical positions about the role of this region currently diverge. On the one hand, the dorsomedial thalamus could support familiarity, but the loss of inputs to the prefrontal cortex following damage to this region would have wider consequences on cognition, with possible impact on recollection (Aggleton et al. 2011). On the other hand, it could have a general role in several cognitive domains by virtue of its regulatory function of prefrontal cortices, allowing the maintenance of frontal activity over delays necessary to perform complex reflections and decisions (Pergola et al. 2018). In a recognition memory task, the dorsomedial thalamus was found to become critical when interference between stimuli increased (Newsome et al. 2018). Following on this latter view (Pergola et al. 2018), in the Integrative Memory model, we positionned the dorsomedial thalamus as a modulator of prefrontal activity, so that it would support the maintenance of prefrontal activities during tasks demanding in terms of attribution processes (e.g., discrimination between similar interfering stimuli).
Recollection-based retrieval. Figure 3 illustrates the mechanisms involved in recollection-based retrieval. Typically, recollection-based retrieval starts with exposition to partial information from a past episode (either an entity or elements of the context). The partial information cue triggers the reactivation of the complete pattern via pattern completion within the hippocampus (CA3/CA1) (Norman & O'Reilly 2003; Staresina et al. 2013). As the pattern stored in the hippocampus is an index of distributed contents in the neocortex, its reactivation induces the reinstatement of stimulus-specific neocortical representations (Rissman & Wagner 2012; Staresina et al. 2013) in such a way that the contents that were processed when the event was initially experienced and encoded are reactivated at retrieval. So, the sensory-perceptual and visuo-spatial details of the memory (e.g., object features, persons’ characteristics, spatial configuration, sounds…) stored in posterior cerebral areas are brought back. The signal from the hippocampal index is transferred to distributed neocortical sites via the mammillary bodies (connected to the hippocampus by the fornix), the anterior nuclei of the thalamus and the retrosplenial cortex (Brodmann areas BA29 and BA30). In other words, Papez’s circuit is the core pathway for recollecting the content of past experienced episodes (Aggleton & Brown 1999).
Figure 3. Illustration of the main steps for a recollection-based memory judgment in the example of an object recognition memory task (following encoding of objects in various spatial locations).

In addition to strong connections with the hippocampus and anterior thalamus, the retrosplenial cortex is linked to the parahippocampal cortex, occipital areas and adjacent posterior cingulate cortex (BA23 and BA31) (Kobayashi & Amaral 2003; Parvizi et al. 2006; Suzuki & Amaral 1994; Vogt & Pandya 1987; Vogt et al. 1987). The posterior cingulate cortex and the retrosplenial cortex appear to play a pivotal role as interfaces between the hippocampus and the neocortex, thanks to their highly connected nature. Indeed, they have been identified as hubs of connectivity (Hagmann et al. 2008; van den Heuvel & Sporns 2013). However, the different patterns of connection of the retrosplenial cortex and posterior cingulate cortex suggest different contributions (Greicius et al. 2009). As a gateway between
the hippocampus and regions storing the sensory-perceptual details of the memory (especially, visuo-spatial information in the parahippocampal cortex and occipitoparietal cortex), the retrosplenial is a key region for enabling cortical reinstatement of the content of memories. It is part of the context representation core system, and its damage should prevent content reactivation and leads to amnesia (Aggleton 2010; Valenstein et al. 1987; Vann et al. 2009).

In contrast, the posterior cingulate cortex sits outside the context and relational representation core systems because it does not contribute to recollecting the content of episodes like the retrosplenial cortex does. Intracranial recordings from posterior cingulate sites in epileptic patients show enhanced gamma band activity specific to autobiographical remembering (Foster et al. 2012), but perturbation of posterior cingulate neurons by electric brain stimulation in the intracranial electrodes do not produce any observable behavioral responses, nor any subjective experience in the participants (Foster & Parvizi 2017). By contrast, electrical stimulation of the MTL evokes a subjective experience of déjà vu/déjà vécu, reminiscence of scenes or of visual details of known objects (Barbeau et al. 2005; Bartolomei et al. 2004). This suggests that the posterior cingulate cortex does not store any content related to experienced memories, but rather plays a supportive role during recollection. More specifically, the posterior cingulate cortex contributes to the quality of recollection and the subjective experience of remembering thanks to its central position as hub of connectivity. A distinction is made between the ventral and dorsal posterior cingulate cortex (Vogt et al. 2006). While the ventral posterior cingulate cortex connects notably with the inferior parietal cortex and ventromedial prefrontal cortex, the dorsal posterior cingulate cortex has main connections with the superior parietal cortex and the dorsolateral prefrontal cortex (Bzdok et al. 2015; Leech et al. 2011; Parvizi et al. 2006; Vogt et al. 2006).
The ventral posterior cingulate cortex is part of the default mode network (Leech & Sharp 2014; Margulies et al. 2009), which has been associated with various internally-directed cognitive functions, such as episodic memory retrieval, self-referential processing, and mentalizing (Buckner et al. 2008). During recollection, the ventral posterior cingulate cortex will support pattern completion by allowing the reactivation of the self-referential character of memories for personally experienced events via its connection to the ventromedial prefrontal cortex (D'Argembeau 2013). It should be noted that recollection can occur in the absence of self-referential feeling as illustrated by the case of a patient who remembered personally experienced events with contextual details, but who had the feeling that these events did not belong to him (Klein & Nichols 2012). However, the lack of self-referential character in recollected memories would prevent them from inducing the subjective feeling of travelling back in time to reexperience one’s past (Tulving 1985). Then, the sudden recovery of the whole memory trace on the basis of a simple cue (i.e., ecphory) captures bottom-up attention and engages the ventral attention network, more specifically the ventral parietal cortex (supramarginal gyrus and angular gyrus (Attention-to-Memory model, Cabeza et al. 2012), via the ventral posterior cingulate cortex connection.

As for the dorsal posterior cingulate cortex, it is thought to be a transitional zone of connectivity, linking the default mode network and a frontoparietal network involved in executive control (Leech & Sharp 2014). In our Integrative Memory model, this frontoparietal network corresponds to the attribution system interacting with attention. Of note, the retrosplenial cortex also has direct connections with the dorsolateral prefrontal cortex (Kobayashi & Amaral 2003; Vann et al. 2009), suggesting that the posterior cingulate gyrus as a whole acts as a gateway between the hippocampally-centered relational representation
core system and the frontoparietal attribution and attention system. So, we propose that the posterior cingulate gyrus hub of connectivity, comprising the retrosplenial cortex and posterior cingulate cortex, has a pivotal role in the integration of all the recollection-related operations and contents. It would act as a relay node allowing activation to spread from the relational representation core system throughout the entity representation core system, context representation core system, self-referential system and the attribution system. Dysfunction of this node would disintegrate the network, preventing the full reinstatement of the memory. Consistently, Bird et al. (2015) have shown that the posterior cingulate gyrus allows the reinstatement of episodic details and the strength of the posterior cingulate reinstatement activity correlated with the amount of details that the participants could subsequently recall.

Finally, in order for the individual to report an “old” judgment based on a recollective experience, attribution mechanisms should come into play, taking into account the task context and memorability expectations (metacognitive knowledge and monitoring) (McCabe & Balota 2007). We assume that the fundamental cognitive operations are the same as in the case of familiarity, but the nature of representations on which this applies differs. Here, the attribution system will assess notably the amount of recollected details (Johnson et al. 2009) and their relevance (Bodner & Lindsay 2003). This implies that, even if an individual recollects qualitative details about an event, s/he may report a familiarity-based recognition decision if the retrieved information is judged irrelevant or insufficient to succeed at the task and to be qualified as recollection (e.g., Remember response) (Bodner & Lindsay 2003). In addition, the criterion for recollection will depend on task context. For instance, in McCabe and Balota’s (2007) study, medium-frequency words were intermixed with high-frequency or low-frequency words at test. Remember responses were greater for medium-frequency targets when they were tested among high-frequency, as compared with low-frequency words. This
suggests that participants are more likely to experience recollection when targets exceed an expected level of memorability in the context of words that were relatively less distinctive.

In line with the hypothesis that the posterior cingulate gyrus contributes to consciousness (Vogt & Laureys 2005), an additional hypothesis of the Integrative Memory model is that the spread of activation throughout distributed brain regions via the posterior cingulate gyrus hub, the catching-up of attention related to ephory and the high diagnosticity of such signal in terms of evidence of past experience is equivalent to a mobilisation of a global neuronal workspace (Dehaene & Naccache 2001; Vatansever et al. 2015) that conveys consciousness of remembering and a feeling of re-experiencing (i.e., autonoetic consciousness). In this view, autonoetic consciousness would thus be an emerging property of integrated reactivation of the representation core systems together with the attribution system, where the posterior cingulate gyrus plays a central role.

Table 1 summarizes the key computational operations and the corresponding types of content that can be represented thanks to these properties according to the Integrative Memory model. We distinguish the core systems that create the memory trace, that will become available for familiarity- and recollection-based memory decisions (as well as other cognitive functions as described in section 2.3), and the subjective experience of remembering and knowing which are emerging psychological phenomena arising from the interaction between the core systems representations and the cognitive operations of the attribution system.
Table 1. Main computational/cognitive operations and associated representations/psychological consequences in the Integrative Memory model.

<table>
<thead>
<tr>
<th>Entity representation core system</th>
<th>Relational representation core system</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Computational operations</strong></td>
<td><strong>Representations</strong></td>
</tr>
<tr>
<td>Hierarchical integration</td>
<td>From features to conjunctions of features</td>
</tr>
<tr>
<td>Entity pattern separation</td>
<td>Unique conjunctive representations of objects, people and simple associations</td>
</tr>
<tr>
<td>Fluency due to prior exposure</td>
<td>Reactivated features/conjunctions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feeling of familiarity</th>
<th>Remembering</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive operations</strong></td>
<td><strong>Psychological phenomenon</strong></td>
</tr>
<tr>
<td>Fluency heuristic (i.e., attribution)</td>
<td>Feeling of familiarity</td>
</tr>
<tr>
<td>Metacognitive knowledge &amp; monitoring</td>
<td>Modulation of use of the fluency heuristic</td>
</tr>
<tr>
<td>Top-down attention</td>
<td>Explicit judgment of familiarity</td>
</tr>
</tbody>
</table>

2.2 Interactions within the Integrative Memory model

Although the core systems that represent the memory traces generating recollection and familiarity are independent, it is important to consider that these systems interact. This will
occur when the representations from the entity and context representation core systems are used to create relational associations in the relational representation core system and then can be subsequently reinstated during pattern completion. For instance, fMRI studies have shown that covert retrieval of the context previously associated with an item activated the parahippocampal cortex when probed with the item alone, whereas perirhinal-related representations of the item was activated by presenting the associated context, with the hippocampus coordinating the reinstatement (Diana et al. 2013; Staresina et al. 2012; Wang et al. 2013).

Moreover, at the level of memory outputs from the attribution system, familiarity and recollection can interact (Kurilla & Westerman 2008, 2010; Mandler et al. 1969; Whittlesea 1997). Notably, a feeling of familiarity can trigger an active search in memory to recollect specific details about some event. For instance, when seeing a familiar face in the crowd, one often wishes to remember one’s past interactions with that person. Typically, we will elaborate retrieval cues, with the support of the ventrolateral prefrontal cortex (i.e., cue specification, Figure 1), trying to specify contextual information associated to the face until we find an appropriate cue that will trigger pattern completion in the hippocampus (Ciaramelli et al. 2008). Alternatively, recollection acts as a control over familiarity. For instance, when some aspects of a stimulus feels familiar, remembering that they were actually part of another memory allows to correctly reject the current stimulus (e.g., recombined pairs in associative memory tasks, exclusion trials in the Process Dissociation Procedure).

Also, expectancies induced by the task characteristics can shift the balance between recollection and familiarity as outputs. For instance, some materials like pictures induce high expectations in terms of memorability compared to other kinds of materials (i.e.,
distinctiveness heuristic). In this case, participants think that they will recollect many perceptual details. If they do not for a given stimulus, they will consider it is new even if they experience fluency feelings. As recollection was anticipated but not familiarity, fluency cues are disregarded because of the absence of recollection (Dodson & Schacter 2001; Ghetti 2003).

Finally, the individual may set specific goals for a given memory situation, that will generate a retrieval mode orientating attention towards the search for particular types of information. This will rely on the interaction between the dorsolateral prefrontal cortex and dorsal parietal cortex (Cabeza et al. 2008; Lepage et al. 2000). For instance, an individual may favor global processing of information leading to familiarity versus analytic processing leading to recollection of details (Whittlesea & Price 2001; Willems et al. 2008), or may even search for specific types of details (Bodner & Lindsay 2003; Bodner & Richardson-Champion 2007).

2.3 Beyond recollection and familiarity
In the Integrative Memory model, similarly to models emphasizing the nature of representations used for memory, core systems store specific contents that serve in memory tasks to retrieve the objects, people, actions, settings… that have been experienced. But the same representations can also be used to perform other tasks. Indeed, perceptual discrimination between entities with overlapping features and their maintenance in short-term memory have been found to involve the perirhinal/anterolateral entorhinal cortex (Barense et al. 2016; Graham et al. 2010). Naming and conceptual discrimination of such entities also rely on perirhinal integrity (Clarke & Tyler 2015). Similarly, the hippocampus uses relational representations in navigation, short-term memory, perceptual discrimination, imagination… (Clark & Maguire 2016; Lee et al. 2012; Yonelinas 2013). So, even if recollection and
familiarity recruit relational and entity representations, they are not the only functions to do so. This has implications for the pattern of deficits arising from damage to these core systems (see section 3.1).

Actually, the whole architecture described in the Integrative Memory model may not be uniquely mnemonic in nature. For instance, the interaction between fluent processing of repeated items in the entity representation core system and the attribution system may lead to affective judgments. This is well illustrated by the mere repetition effect in which repeated items are judged more pleasant and preferred over non-repeated items (Willems et al. 2007). Moreover, the default network, that overlaps partly with the relational representation core system, self-reference system, posterior cingulate gyrus hub of connectivity, ventral parietal cortex and regions from the attribution system involved in metacognition, is also recruited during imagination of future events, mind wandering, and reflection about one’s and other’s mental states (Andrews-Hanna et al. 2010). This network may have an adaptive role by which the brain uses past experiences to simulate possible future scenarios in order to prepare humans to react to upcoming events (Buckner et al. 2008). Additionally, the combined use of the default network and the frontoparietal network (corresponding to interacting core and attribution systems here) supports creative thinking (Madore et al. 2019). Thus, the core systems provide the building blocks that are reconstructed and recombined depending on the individual’s goals, with the help of the attribution system.

The very facts that consistent impairments are observed following brain damage and that the same brain regions are activated when different individuals perform a given task suggest that the neural networks underlying cognitive functions are common to all individuals. The purpose of theoretical models, like the Integrative Memory model and others, is precisely to
reveal the universal neurocognitive architecture of memory. Beyond anatomical similarity of memory functioning, one may wonder about the social role of such organisation. Regarding memory, it appears that, when individuals recall a given event (e.g., a TV show episode) with their own words, the pattern of cerebral activation is more similar between people recalling the same event than between recall and actual perception (Chen et al. 2017). This suggests that perceived events are transformed when entering memory in a systematic way that is shared across humans. If true, this would mean that the main purpose of our memory-related neurocognitive scaffolding is not only to allow each individual to remember the events that s/he experienced, but more widely to communicate and share beliefs about the past with other people (Mahr & Csibra 2017) and to facilitate the creation of collective memories that build the social identity of human groups (Halbwachs 1980; Hirst et al. 2018).

2.4 Novelty of the Integrative Memory model compared to current models of recollection and familiarity

As indicated by its name, the Integrative Memory model does not have the ambition to propose a novel framework, but rather to integrate some principles from currently most influential theories. There are therefore a lot of similarities with existing models, although some differences exist. The Integrative Memory model borrows to representational models the idea that memory processes arise from the use of particular types of representations. The entity representation core system relies on hypotheses from the Representational Hierarchical view (Cowell et al. 2006, 2010; Saksida & Bussey 2010) and the Emergent Memory account (Graham et al. 2010). Like the Emergent Memory account, we consider that memory emerges from hierarchically organized representations distributed throughout the brain. The consequence of this is that familiarity can arise from the reactivation of any of these representations (including outside the MTL). In turn, the relational representation core system
builds on relational theories about the role of the hippocampus, by suggesting that the hippocampus flexibly binds disparate pieces of information (Aggleton & Brown 1999; Eichenbaum & Cohen 2014; Eichenbaum et al. 2007). However, our view departs slightly from another representation-based model, Binding of Item and Context (Diana et al. 2007; Ranganath 2010), which posits that the perirhinal cortex supports familiarity for items in general, whereas recollection will rely on context representation in the parahippocampal cortex and item-context binding in the hippocampus. Indeed, we rather propose that the perirhinal cortex is specifically tuned for the representation of complex conjunctive entities, but not items of lower levels of complexity. Moreover, the context representation core system can support familiarity for scene and buildings.

Contrary to process-based models, the Integrative Memory model does not localise recollection and familiarity processes themselves to certain regions, but conceptualises them as processes emerging from the interaction between specific kinds of representation and attribution mechanisms. Yet, in line with process-models like the Convergence, Recollection, and Familiarity Theory and the Complementary Learning Systems (Montaldi & Mayes 2010; Norman & O'Reilly 2003), we consider that the core systems have unique computational properties (e.g., entity versus relational pattern separation) that contribute to shaping the content of stored information. The combination of computational properties and the associated representations makes certain core systems more tune to either recollection or familiarity. But the ultimate memory output will depend on attribution mechanisms.

The network organization of the Integrative Memory model clearly resonates with the Posterior Medial Anterior Temporal (PMAT) framework (Ranganath & Ritchey 2012; Ritchey et al. 2015), but here we separate the network into several subsystems rather than in
two systems. The two views share the idea that this neurocognitive architecture not only supports episodic memory, but also other functions like perception, navigation and semantic processing. In the PMAT framework, the ventromedial prefrontal cortex is a site of convergence between prefrontal and MTL components of the Anterior Temporal and Posterior Medial systems. This region would provide the value of item and bound representations, exercise some control over the representations, notably to select the relevant content as a function of the situation, and to help with the integration of new information within existing representations. Similar ideas figure in the Integrative Memory model, notably by suggesting that the self-representation system (involving the orbitofrontal and ventromedial prefrontal cortex) interacts with the core systems to provide some self-referential tagging, thus modulating the value of the representations in core systems. Close to the idea of control over mnemonic traces, we also include the prefrontal cortex in the attribution system. Although both the PMAT framework and the Integrative Memory model include the retrosplenial cortex and posterior cingulate cortex, their role is conceived slightly differently. In the PMAT framework, both the retrosplenial cortex and posterior cingulate cortex form parts of the Posterior Medial system that allows individuals to orient in time, space and situation. Here, we suggest that the retrosplenial cortex is an integral part of a core system dedicated to storing visuo-spatial and contextual information. In contrast, the posterior cingulate cortex acts as a relay node during cortical reinstatement of the memory trace and, by connecting all systems within the network, including the attribution system, it would contribute to the subjective experience of mentally reliving the episode. This is an original hypothesis of the Integrative Model suggesting a key role of the posterior cingulate gyrus in autonoetic consciousness.
Finally, the articulation of the model around the interaction between core systems and the attribution system is probably the most novel aspect of the Integrative Memory model. Currently, no recollection/familiarity neurocognitive framework has taken into account the principles from attribution theories. A first proposal relating the fluency heuristic to the perirhinal cortex has however been formulated by Dew and Cabeza (2013). We expand it by suggesting that reactivation of any component of the hierarchically represented item (i.e., object, face, building, word, simple association) will generate a fluency signal which is interpreted by the attribution system in the light of metacognitive knowledge. Similarly, reactivated patterns of complex representations via the hippocampus are also evaluated through the glasses of metacognitive knowledge before being attributed to the past. Because the mapping of attribution processes with cerebral regions is still to be confirmed, a lot remains to be learned about the exact neurocognitive mechanisms involved in the attribution system. For now, we integrated theories about control mechanisms over memory to propose a mechanistic account of the attribution system. Notably, the Attention-to-Memory model (Cabeza et al. 2008; 2012; Ciaramelli et al. 2008) is key in describing the role of the parietal and prefrontal regions in attention and monitoring mechanisms.

3. The Integrative Memory model to understand recollection and familiarity deficits

3.1 Damage to core systems versus attribution system

According to the Integrative Memory model, the dissociation of recollection and familiarity in patients with lesions selective to the hippocampus or perirhinal/entorhinal cortex (Barbeau et al. 2011; Bowles et al. 2010; Brandt et al. 2016) would arise because the core representations are damaged. Hippocampal lesions prevent the encoding and retrieval of relational representations, and so hamper the possibility to re-experience all details from a past episode. In contrast, perirhinal/entorhinal lesions affect the creation of entity representations that
cannot be subsequently felt familiar. Yet, in the latter case, our view predicts that not all forms of familiarity should be impaired following perirhinal/entorhinal damage, but more specifically familiarity-based discrimination between stimuli that share features and require a conjunctive trace. This should be tested by directly manipulating the level of confusability of targets and distractors in recognition memory tests. Moreover, if the same representations support episodic recognition memory, short-term memory and perception, a patient with a selective hippocampal lesion should present with impaired use of relational representations in a variety of tasks, beyond the episodic memory domain. For instance, KA, a patient with developmental amnesia and atrophy of the hippocampus, fornix, mammillary bodies and anterior thalamic nuclei (i.e., the relational representation core system) shows both impaired source memory and impaired relational binding in short-term memory, with preserved conjunctive binding in short-term memory (Jonin et al. 2018a; 2018b). In contrast, a patient with a selective perirhinal/entorhinal lesion should be impaired in the processing of entities across recognition memory, perception and short-term memory. Recently, Lacot et al. (2017) reported that JMG, who had damage to the MTL cortices preserving the right hippocampus, failed on a variety of recognition memory tasks involving objects and abstract pictures, but had good visual recall for spatial patterns and geometric figures and good scene recognition performance. Interestingly, recollection as well as familiarity for objects and abstract pictures were affected, emphasizing the idea that the nature of representations carried by the perirhinal/entorhinal cortex versus hippocampus matters more than the memory processes to explain JMG’s profile.

With regard to the context representation core system, severe anterograde amnesia associated with mild retrograde amnesia has been described following lesion to the retrosplenial cortex (for reviews, Aggleton 2010; Vann et al. 2009). As the retrosplenial cortex connects the
hippocampus to regions storing the sensory-perceptual details of memories, patients with retrosplenial amnesia should present with a recollection deficit (Aggleton 2010) as well as poor scene recognition and perceptual discrimination. The paucity of retrosplenial amnesia cases has not allowed to test this prediction yet, but the finding that the case described by Valenstein et al. (1987) showed impaired recall of paired associates and complex figure, but relatively preserved recognition memory for faces could speak for intact familiarity-based memory for entities in the face of deficient relational memory.

Finally, frontal lesions should affect the adequacy of the explicit output with regard to the characteristics of the task or the quality of the subjective experience because of disruption of the attribution system. Impairment of monitoring of memory output has been described following lateral prefrontal cortex lesions, leading notably to false recognitions (Schacter 1997), and metamemory abilities are negatively affected by frontal lesions, especially to medial prefrontal cortex (Pannu & Kaszniak 2005). If such monitoring and metacognitive mechanisms are common to recollection and familiarity processes, as we suggest, both should be impacted by prefrontal lesions. A few studies have assessed the consequences of frontal lesions on recollection and familiarity. Their results were inconclusive, with some studies reporting deficits only in recollection (Anderson et al. 2011; Stamenova et al. 2017; Wheeler & Stuss 2003), others indicating deficits only in familiarity (Aly et al. 2011; MacPherson et al. 2008) and a few studies describing deficits in both recollection and familiarity (Duarte et al. 2005; Kishiyama et al. 2009). This inconsistency is perhaps not surprising if one takes into account the fact that prefrontal areas are not supporting recollection and familiarity processes per se, but rather expectations, selection and monitoring processes operating on the reactivated content in answer to the specificities of the task at hand. In this view, a prefrontal lesion could reduce the expression of recollection/familiarity-based memory outputs in some
conditions, but not others. We believe that the effect of frontal lesions on recollection and familiarity would be best apprehended by examining variation in memory outputs as a consequence of manipulations affecting expectations and criterion setting.

3.2 Recollection and familiarity in the course of Alzheimer’s disease

The case of Alzheimer’s disease (AD) is interesting to consider in the light of the Integrative Memory model for at least two reasons. First, this is a progressive disease in which the trajectory of cognitive decline is related to the spreading of neurofibrillary tangles and neuronal loss (Jack et al. 2013) starting in the anterolateral entorhinal and perirhinal cortices (Braak & Braak 1995; Braak & Del Tredici 2015), corresponding to Braak’s stage I. As hippocampal pathology comes later (Braak’s stages III), Braak’s stage I represents a unique model of selective MTL lesions targeting the entity representation core system, which is very rarely encountered in other neuropsychological populations (Barbeau et al. 2011; Bowles et al. 2007). Moreover, in amnestic Mild Cognitive Impairment (aMCI), a diagnostic entity at high risk of developing AD (Albert et al. 2011), dysfunction of the posterior cingulate gyrus appears as a prominent feature (Chetelat et al. 2003; Dunn et al. 2014; Salmon et al. 2008). More specifically, although both the ventral and dorsal posterior cingulate cortex show hypometabolic activity in aMCI (Mutlu et al. 2016), the retrosplenial cortex demonstrates the most consistent reduction in metabolism (Nestor et al. 2003). Two mechanisms may contribute to such retrosplenial hypometabolism in aMCI: atrophy of the region (Pengas et al. 2010; Scahill et al. 2002), and distant consequence of the pathology affecting the hippocampus and anterior nuclei of the thalamus (Braak & Braak 1991; Villain et al. 2008), this loss of inputs being aggravated by disruption of cingulum bundle that connects the hippocampus to the retrosplenial cortex (Villain et al. 2008). Patients with aMCI would
therefore be a target population to relate changes to recollection-based memory to the pathology of regions within the relational and context representation core systems.

Second, even if several studies have investigated the integrity of recollection and familiarity in aMCI and AD, the exact profile of deficits is still unclear. Most puzzling is the fact that there is no consensus regarding the fate of familiarity in aMCI and AD. Whereas a number of studies have reported preserved familiarity in these populations, an almost equal number of studies have shown that familiarity is deficient (for reviews, Koen & Yonelinas 2014; Schoemaker et al. 2014). Of course, this divergence across studies can be partly explained by differences in terms of methods (i.e., paradigm used to assess recollection and familiarity, nature of the materials…) and characteristics of the patients (i.e., severity of cognitive decline, cognitive domains affected beyond memory, heterogeneity of the aMCI population…). A systematic evaluation of these factors that modulate familiarity performance in aMCI and AD would actually be warranted. Yet, we propose that considering the complex multi-faceted nature of familiarity may shed some light on these divergent findings.

One fact that could contribute to explain why it is difficult to get a clear picture of the exact profile of impairment of recollection and familiarity from current findings in aMCI and AD may precisely relate to the progressive nature of the pathology. Because key regions within the cerebral architecture of the Integrative Memory model are affected at different stages of the disease, we hereby propose a hypothetical scheme for the chronological pattern of deficits in the course of AD (see Figure 4). More specifically, some dimensions of recollection and familiarity processes may become dysfunctional at a specific stage of the disease depending on the brain regions most affected at that time. Progression stages will mainly refer to Braak’s neuropathological stages (Braak & Braak 1991, 1995), in association with hypometabolism
and atrophy. Amyloid burden does not appear to correlate with cognitive decline, but provides the background that defines Alzheimer’s pathological changes (Jack et al. 2018). Figure 4 indicates when a deficit starts to appear, assuming that already existing impairments are still present and exacerbated by increased pathological burden.

Figure 4. Chronological unfolding of deficits to specific mechanisms as a function of neuropathological changes in the course of Alzheimer’s disease.

In stage I, neurofibrillary tangles and neuropil threads are limited to the transentorhinal cortex, corresponding to the medial portion of the perirhinal cortex and the anterolateral entorhinal cortex (BA 35) (Taylor & Probst 2008). Even if individuals in that stage have evidence of amyloid pathology, they are asymptomatic and cognitively healthy. Nevertheless, we predict that the presence of neuropathology in the perirhinal/anterolateral entorhinal cortex would affect the ability of stage I individuals to create complex conjunctive representations (i.e., entity pattern separation). This may not impact everyday life functioning, as the difficulty would be limited to specific cases requiring fine-grained entity-level discrimination, but could be seen in tasks that particularly probe rapid familiarity-based recognition of entities among very similar stimuli (Besson et al. 2017). Familiarity for non-conjunctive information and recollection should remain intact.
Stage II is still called a “transentorhinal stage” because it is characterized mainly by an aggravation of transentorhinal pathology. While most individuals in stage II are asymptomatic, a few aMCI patients may harbor stage II neuropathology (Petersen et al. 2006). Whereas impaired familiarity for entity should be relatively modest in stage I, it would be more clearly established in stage II, proportionally to neuropathology in the perirhinal/anterolateral entorhinal cortex. Of note, Braak and Braak (1991) reported mild changes to the anterior nuclei of the thalamus. As an important relay node in the Papez’s circuit (Aggleton & Brown 1999), the anterior thalamic nuclei participate to the recollection of the content of past experienced episodes (i.e. relational representation core system). Therefore, slightly impoverished recollection should accompany these early pathological changes in anterior thalamus. Still, poor familiarity for entities should dominate.

Stage III corresponds to invasion of neurofibrillary tangles in the hippocampus, initially affecting CA1 and the subiculum. Individuals in stage III present significant memory decline compared to stage II, but no general cognitive decline (Grober et al. 1999). Most aMCI patients are indeed in this stage (Petersen et al. 2006). Moreover, in aMCI, hippocampal atrophy is typically found, especially in CA2-3 (Hanseeuw et al. 2011), and a consistent hypometabolism is observed in the retrosplenial cortex (Nestor et al. 2003). Altogether, this would lead to dysfunction of the relational and context representation core systems, affecting the ability to bind item and context information in pattern-separated representations. Reinstatement of encoded patterns should also be impaired. Thus, recollection deficits are expected to occur, together with familiarity for entities as in earlier stages. Familiarity for non-conjunctive information still remains intact and poor recollection may dominate the memory profile in stage III/aMCI patients.
Stage IV starts to include demented AD patients, with more and more severe cognitive impairment as one moves towards stages V to VI (Braak & Braak 1991; Grober et al. 1999; Petersen et al. 2006). In stage IV, in addition to pathology affecting most parts of the hippocampal formation, mild changes to the isocortex can be seen. From stage V onwards, isocortex becomes more and more severely and widely affected, with a spreading of tau pathology to connected regions (Brettschneider et al. 2015). Initial isocortical pathology is notably present in the retrosplenial cortex and inferior occipitotemporal cortex (Braak & Braak 1991), and then transmits progressively to connected regions. On FDG-PET, one can see a typical pattern of temporoparietal, ventromedial prefrontal and posterior cingulate hypometabolism from the mild AD stage (Herholz et al. 2002). Frontal atrophy also emerges, although later (Salat et al. 2001).

In stage IV and early stage V, impaired recollection should dominate as more and more regions contributing to the creation and reinstatement of complex relational representations are lesioned, notably the retrosplenial cortex. Consistently, an fMRI study indicated that, when AD patients used residual recollection capacities, they activated the posterior cingulate gyrus as healthy controls did, but showed impaired functional connectivity between the posterior cingulate gyrus and the hippocampus, inferior parietal cortex and dorsolateral prefrontal cortex (Genon et al. 2013). Moreover, decreased self-reference memory effects suggest that the connection between the content of the memory trace and the self-reference system is reduced in early AD (Genon et al. 2014), in relation to atrophy in the medial prefrontal cortex and posterior cingulate gyrus (Wong et al. 2017). In addition, the difficulty to process entities, in memory tasks as well as in others such as perceptual tasks, would still correlate with the degree of pathology in perirhinal/anterolateral entorhinal cortex, as suggested by recent data in mild AD (Bastin et al. 2014; Delhaye et al. 2019; Kivisaari et al.
2013; Yeung et al. 2017). Finally, creation and reactivation of traces for simple items and associations associated with feelings of familiarity should remain possible, but mild pathology to inferior occipitotemporal could disturb processing of certain categories of stimuli that depend on this area, such as faces.

Yet, when attribution still functions relatively well (likely stages III, IV, and maybe early stage V), in the context of objective (recollection) memory difficulties in everyday life, we predict that fluctuant familiarity may be observed because of metacognitive changes. Indeed, studies in amnesia suggested that awareness of memory deficits in everyday life may modify metacognitive assessment of fluency cues and lead to increased disqualification of such cues (Geurten & Willems 2017; Ozubko & Yonelinas 2014). Along those lines, for early stages of AD, the interaction between anosognosia and reliance on fluency cues would be a promising avenue for research. In aMCI and AD, awareness of memory difficulties is variable among patients (Starkstein 2014). It may be that patients who acknowledge their memory deficits may be less prone to rely on fluency-based feeling of familiarity and would discard them more readily, whereas anosognosic patients would not and may even demonstrate an over-reliance on fluency-based familiarity.

As soon as the disease progresses and impacts regions belonging to the attribution system (late stage V and stage VI), the interaction between (already degraded) contents and attribution becomes severely compromised, and both recollection and familiarity should appear impaired in recognition memory tasks. In particular, patients will lose the ability to subjectively relive past events (i.e., autonoetic consciousness) and this deficit should correlate with decreased functional connectivity between the posterior cingulate hub and the rest of the network. Moreover, less efficient attribution processes could modify the transformation of
reactivated content into subjective judgments and feelings. For instance, in AD, despite the
deficit in their recollection abilities, patients show increased false recollections (Gallo et al.
2010) and have a continued willingness to adopt retrieval strategies oriented towards
recollection (Willems et al. 2008). To explain false recollections, it can be hypothesized that
patients relax their criterion for experiencing recollection and even very partial recollection
may trigger a feeling of remembering.

This hypothetical scheme of chronological changes of key mechanisms within the Integrative
Memory model suggests that cross-sectional comparisons of groups of aMCI or AD patients
on general measures of recollection and familiarity would provide divergent results because
of variability in the pathological status (in terms of Braak’s stage) of patients and in the type
of mechanisms that tasks preferentially assessed. An ideal approach to test the unfolding of
memory impairments would be longitudinal assessments of individuals from the
asymptomatic stages of AD with tasks designed to measure specific mechanisms such as
entity pattern separation, relational pattern separation, pattern completion…, with a possibility
to relate these cognitive changes to the progressive topography of neurodegeneration, using
for instance tau-PET tracers (Schwarz et al. 2016).

4. Conclusions

The Integrative Memory Model presents a framework of the computational mechanisms and
their cerebral bases allowing encoding and retrieval of events in interacting core systems and
attribution system. This model is not opposed to existing influential neurocognitive models of
memory, but rather complements them by integrating many of their principles into a single
view. This integration allows generating novel hypotheses. For instance, we propose that
disruption of the posterior cingulate connectivity hub, where the retrosplenial cortex, ventral
and dorsal posterior cingulate cortex play distinct but complementary roles, will alter cortical reinstatement of encoded details and autonoetic consciousness. Another novel prediction is that reactivation of traces in core systems does not lead obligatorily to the subjective experience of recollection and familiarity. Subjective feelings and final explicit recognition judgements necessitate late attribution mechanisms. The intervention of such attribution processes could explain why apparently impaired use of familiarity may result from a strategic disqualification of fluency cues in people with memory problems.

We propose the Integrative Memory Model with the aim to move forward research on the nature of the memory deficits in brain-damaged populations as we believe that future studies should be framed according to the latest advances in the knowledge about memory functioning. In particular, we consider the course of Alzheimer’s disease as a good example for applying a reading based on our framework by suggesting a scheme of chronological dysfunction of specific mechanisms as a function of the topographical progression of neuropathological changes. Such scheme suggests that the predominant impairment in terms of recollection and familiarity will shift in the course of the disease, from very early neurodegeneration of the perirhinal cortex leading to deficits in tasks that rely on the representation of items as viewpoint-invariant conjunctive wholes (e.g., familiarity for entities) with increasingly dominant impairment of recollection when pathology invades several regions from the relational representation core system.

The model could also serve as the basis to understand memory impairments in various conditions, such as healthy aging, amnesia, temporal lobe epilepsy, or frontotemporal dementia. The impact of brain damage affecting specific regions on recollection and familiarity would provide important tests of the neurocognitive architecture of memory and
could lead to reconsideration of the model if some hypothesized region-mechanism links were falsified by experiments. Also, the proposed neurocognitive architecture is certainly not comprehensive and should evolve to incorporate other mechanisms. For instance, the notion of time which is critical in episodic memory (Eichenbaum 2013) is not considered here. Finally, a broader avenue for research is to examine the role that such architecture plays beyond memory, in allowing humans to prepare for future events and to share the past.
References


Acknowledgements

This work was supported by the University of Liège, the Fonds Léon Frédéricq, SAO-FRA (grant S#14003 and #2017/0008), Federation Wallonie-Bruxelles Special Funds for Research (grant #FSRC-14/11), the F.R.S.-FNRS (grant T0009.19), the Alzheimer Association (grant 2016-NIRG-394141) and the Inter-University Attraction Pole P7/11. C. Bastin is a research associate of the F.R.S.-FNRS. C. Bastin and S. Willems thank Pr. Martial Van der Linden for mentoring their first steps in recollection/familiarity studies.