Why the mind wanders: How spontaneous thought’s default variability may support episodic efficiency and semantic optimization

Caitlin Mills¹, Arianne Herrera-Bennett², Myrthe Faber⁴, Kalina Christoff¹,²

¹Department of Psychology and ²Centre for Brain Health, University of British Columbia
³Munich Center of the Learning Sciences, Ludwig Maximilians University Munich
⁴Department of Psychology, University of Notre Dame

Abstract

In this chapter, we offer a functional account of why the mind – when free from the demands of a task or the constraints of heightened emotions – tends to wander from one topic to another, in a ceaseless and seemingly random fashion. To account for this, we propose the Default Variability Hypothesis, which builds on William James’ phenomenological account of thought as a form of mental locomotion, as well as on recent advances in cognitive neuroscience and computational modeling. Specifically, the Default Variability Hypothesis proposes that our default mode of mental content production yields the frequent arising of new mental states that have heightened variability of content over time. This heightened variability in the default mode of mental content production may be an adaptive mechanism that (1) enhances episodic memory efficiency through de-correlating individual episodic memories from one another via temporally spaced reactivations, and (2) facilitates semantic knowledge optimization by providing optimal conditions for interleaved learning.

Introduction

Why doesn’t the mind grind to a halt when we are not doing anything? Why does it keep moving instead? And why does this movement tend to proceed in a seemingly haphazard manner, with thoughts jumping from one topic to another, often distant, seemingly unrelated topic—creating a variability in thought content to which the mind seems to default?

More than 100 years ago, William James described thought as a form of mental locomotion. Here we build on James’ phenomenological account and on recent advances in cognitive neuroscience and computational modeling to offer a functional account of why the
mind, when free from the demands of a task or the constraints of heightened emotions, ceaselessly moves from one topic to the next. We introduce the Default Variability Hypothesis, which highlights the continuous change and heightened variability of spontaneous thought’s contents as they unfold over time. The Default Variability Hypothesis proposes that our default mode of mental content production, with its continuous change and heightened variability over time, may be an adaptive mechanism that (1) enhances episodic storage efficiency by helping decorrelate individual episodic memories from one another via temporally spaced reactivations, and (2) facilitates semantic knowledge optimization by providing optimal conditions for interleaved learning.

**Overview of the Default Variability Hypothesis**

People report highly variable moment-to-moment experiences during “resting states” that facilitate spontaneous thought (Hurlburt, Alderson-Day, Fernyhough, & Kühn, 2015). For example, a thought about the scallops one had for dinner the day before might be followed by a memory of a bus ride one took a week ago, followed by an image of a sunny beach. The mental states that form our thought flow need not be events that have actually occurred (Addis, Wong, & Schacter, 2008; Schacter, Addis, & Buckner, 2007). In addition to conjuring up veridical episodic events, details from the past can also be recombined in novel ways to produce episodic mental simulations and other mental states that become part of the stream of thought.

We operationalize content variability as the extent to which consecutive mental states in the stream of thought are episodically and/or semantically distinct from each other. The greater the semantic/episodic distance between consecutive mental states, the more variable thought content would be over time. We propose that a default mode of variability in thought contents serves two purposes: To facilitate efficient encoding of separate episodic events (the *Episodic Efficiency Hypothesis*), and to support the integration and transformation of episodic memories into semantic knowledge (the *Semantic Optimization Hypothesis*).

In what follows, we elucidate the two sub-hypotheses that together make up the Default Variability Hypothesis. We draw on the episodic and semantic memory formation and consolidation literature to explain how these processes are inextricably connected to spontaneous thought’s default variability. Finally, we integrate our hypothesis into existing accounts of mind wandering, and offer some suggestions for empirically testing each sub-hypothesis.
Episodic Efficiency Hypothesis

We propose that a default mode of content variability in the stream of thought improves episodic memory efficiency by optimizing the distinctiveness of different episodic memories. In this section, we give a brief description of two pivotal episodic memory mechanisms – pattern separation and pattern completion – followed by an account of how content variability of spontaneous thought may lead to increased episodic memory efficiency. We propose that this process is two-fold: First, pattern separation processes produce separable (i.e., distinct) episodic memories, through de-correlating (i.e., making distinct) the corresponding activation patterns in the hippocampus and neocortex. A default content variability in spontaneous thought may directly support pattern separation processes via mental simulations (reactivations or novel recombinations) that adaptively separate the memories over time through providing dissimilarities in consecutive representations over time. Second, pattern completion may help strengthen representations of the separately encoded memories through multiple, similar re-instantiations of the same memories (which could be triggered by either external or internal cues).

Pattern Separation

Pattern separation is a process via which distinct representations of episodic experiences, and their contextual properties, are indexed in the hippocampus as separate and discrete events (Rolls, 2016). Pattern separation plays a vital role in episodic memory storage and retrieval by helping us create distinct neural representations for individual episodic events. Here, we propose that a fundamental function of a default variability in mental contents over time is to support pattern separation by helping de-correlate distinct memories from one another.

At the neural level, pattern separation is considered to be dependent on hippocampal processes (Leutgeb, Leutgeb, Moser, & Moser, 2007; Rolls, 2016; Yassa & Reagh, 2013). It begins with input from the entorhinal cortex (see Figure 1), which feeds into the granule cells of the dentate gyrus (DG) by way of the perforant path (Witter, 1993, 2007). The DG cells are proposed to serve as a modifiable network that ultimately produces sparse, orthogonalized outputs to CA3. DG granule cells exhibit unique functional properties: they have relatively
sparse firing rates, yet a strong influence on CA3 cells (Jung & McNaughton, 1993; Leutgeb et al., 2007). Moreover, only a very small number of connections are received at each CA3 cell. For example, it is presumed that the small number of connections (approximately 46 mossy fiber connections to each CA3 cell) creates a randomizing effect (i.e., for any given event, a random set of CA3 neurons is activated). In other words, there is an extremely low probability that any two CA3 neurons would receive input from a similar set of DG cells (Kesner, 2007). As a result, event (i.e., episodic) representations should be as highly differentiated as possible from one another (Rolls, 1989, 1989; Rolls & Kesner, 2006; Rolls & Treves, 1990), which affords optimal storage capacity of distinct event representations (Hunsaker & Kesner, 2013; Myers & Scharfman, 2009, 2011, Treves & Rolls, 1992, 1994).

The mechanisms in pattern separation ultimately contribute to the orthogonalization of episodic memory representations, characterized by reduced overlap or redundancy between distinct event representations. Sometimes referred to as dilution or diluted connectivity, orthogonalization is characterized by low levels of correlation between different encoded episodic memories and a low number of synaptic connections between each of the CA3 neurons themselves – as little as one connection between any pair of randomly connected CA3 neurons within the network (Rolls, 2013). Supporting evidence comes from both rodent and human studies suggesting that the DG and CA3 can update the de-correlated network after exposure to even slight deviations in previously encountered contexts or stimulus (Bakker, Kirwan, Miller, & Stark, 2008; Gilbert, Kesner, & Lee, 2001; Leutgeb et al., 2007). For example, novelty is associated with increased firing rates from certain inhibitory neurons in the DG (Nitz & McNaughton, 2004), which may serve as a filtering mechanism for determining when new events should be encoded as such (Jones & McHugh, 2011).

From Pattern Separation in the Hippocampus to Neocortical Competition

Although the exact neural details of pattern separation are still a subject of debate, one thing is clear: the ability to separate different episodic memory patterns requires an efficient storage mechanism. In addition to the sparse encoding in the DG and CA3, efficient storage capacity is also proposed to be achieved through hippocampo-cortical interactions, according to the Hippocampal Indexing Theory (Teyler & DiScenna, 1986). While individual memory traces
are indexed separately in the hippocampus, additional details of the memory are thought to be stored elsewhere in the cortex. Yassa et al. (2013) have described this process as the neocortex being akin to a library where the information is stored and the hippocampus as the librarian who can point to where the information is stored.

Based on this idea, the Competitive Trace Theory (Yassa & Reagh, 2013) makes a specific prediction about the benefits of episodic memory reactivation. According to this theory, certain episodic features of a memory are preserved through multiple reactivations of this memory over time (Figure 2). Every reactivation causes a trace to be re-encoded in the DG, so that this trace does not completely overlap with the traces created by other reactivations or by the original event. Across multiple reactivations over time, some features of the memory will overlap (i.e., will be the same), while others will not (i.e., they will differ; Figure 2). Over time, overlapping features are strengthened with respect to their corresponding hippocampal and neocortical representations, which results in their higher fidelity during retrieval. On the other hand, non-overlapping features compete for representation in the cortex (unlike overlapping features), and mutually inhibit one another through anti-Hebbian learning (i.e., active neurons initiate inhibitory competition, and weakly activated neurons are subsequently inhibited). Therefore, non-overlapping features – which are presumably likely to be the less common and less important features of the memory – will have a reduced likelihood of being retrieved.

The idea that both the hippocampus and neocortex are involved in pattern separation is important for considering how temporally variable mental simulations (reactivations and novel recombinations) can aid in efficient separation. The neocortex, where much of the episodic memory information is stored, is associatively modifiable through competitive learning so that given some input, competition is generated among neural representations (i.e., multiple representations of a memory receive some level of activation, resulting in a competition between them to win total activation, resulting in an action potential). The “winner” of the competition then becomes activated, thus strengthening the association between the input and particular neural activations in the neocortex. Indeed, the mossy fiber system in the DG and its connections to CA3 also exhibit an associative Hebbian learning network (Treves & Rolls, 1994), where concurrent presynaptic activity and postsynaptic action potentials result in a strengthened connection and increased synaptic efficiency (Treves & Rolls, 1994). In turn, this type of
synaptic strengthening supports the sparse coding in the DG and CA3, which may be a sufficient mechanism for orthogonalization in the hippocampus.

The idea that memories can become de-correlated over time in the neocortex also bears relation to other proposed mechanisms. For example, Hulbert & Norman (2015) propose a process similar to pattern separation, called differentiation, where episodic memories can become de-correlated through competitive learning in the hippocampus and cortical regions. Their explanation distinguishes between pattern separation, which is asserted to be automatic, and differentiation, which is driven by competition (in the neocortex) after pattern separation has already occurred (in the hippocampus). Hulbert & Norman (2015) present fMRI evidence that reduced similarity in the hippocampus between memories is correlated with retrieval-induced facilitation, which is the opposite of retrieval-induced forgetting (e.g., impaired memory for related items). This pattern of results supports the idea that when memories are differentiated from one another, they do not hinder retrieval due to similarity. Further support also comes from Favila et al. (2016), who showed that reducing the similarity between memories can be an adaptive process: learning serves to reduce the amount of overlap in hippocampal representations of highly similar stimuli, which in turn prevents interference during subsequent retrieval.

**Role of Content Variability**

How does the content variability inherent to spontaneous thought contribute to episodic memory separation? At a basic level, spontaneous reactivations can provide the foundation for initiating competition between multiple instantiations of a given memory, ultimately preserving the important (i.e. recurring) features of the memory. That is, since each spontaneous thought is re-encoded as a new memory trace (Yassa & Reagh, 2013), competition is generated among the non-overlapping features in the new and previously encoded memories. The idea here is that essential overlapping features that are present in both (or more) versions of the encoded memory will be strengthened and retained for later recall due to spontaneous reactivations. At the same time, spontaneous reactivations are unlikely to be veridical instantiations of the memory. Therefore, the non-overlapping, and perhaps irrelevant, features of that memory will be inhibited and potentially lost over time. See Figure 2 for a graphical example.

A second proposed role of content variability is that memories can become de-correlated through continual shifts in mental content, where memory reactivations are not temporally or
spatially bound from one spontaneous thought to the next (see Figure 3 for an example). The variability of content over time acts to provide a time buffer between overlapping memories. Enough time can pass in between similar memory traces, such that activation from one memory can die down before other related memories are activated, thus avoiding the “fire together, wire together” association rule originally proposed by Hebb.

We argue therefore that default content variability plays a functional role in organizing episodic memories by optimizing de-correlated memories in the hippocampus and neocortex. Although the flow of mental states during spontaneous thought may seem randomly disjointed, spontaneous thoughts are often tied to recent memories, past events, or future plans (Baird, Smallwood, & Schooler, 2011; Klinger & Cox, 2004). Thus, spontaneous mental simulations may play a critical role in separating episodic memories that are important to one’s life, such as episodic experiences that needs to be distinguished from others and should not be grouped with them due to factors such as temporal contiguity.

**Pattern Completion**

Pattern completion is a process via which completion of a whole memory of an event or experience is generated from recall of any of its parts. In other words, partial or degraded cues can trigger the respective stored event representation, which then serves to reactivate the original episodic memory and its accompanying features, including the context in which it was originally experienced (Marr, 1971). The phenomenological qualities of the original event – including even elements such as the emotional tone of the initial experience – can be recaptured and reinstated, and in this sense vividly *re-experienced* by the individual. As such, pattern completion is central to the notion of episodic memory retrieval, in that it supports not only the recall of the information surrounding a given event, but also taps into the fundamental conscious feeling of reliving a moment as a specific, rich and unique subjective episode (Nadel & Moscovitch, 1997, 1997; Teyler & Rudy, 2007). Pattern completion therefore elicits a sense of autonoetic consciousness (e.g., the ability to mentally put ourselves in other situations – past, present, and imagined – and reflect on them), a hallmark of episodic awareness (James, 1890; Tulving, 2002), which is necessary for episodic memory retrieval.

At the neural level, the hippocampus and the surrounding structures of the medial temporal lobe (MTL) are considered to be among the key neural substrates underlying the
reinstatement of episodic memories (see Figure 1). While pattern separation is thought to be mediated by the DG, areas CA1 and CA3 have been implicated as more central components of pattern completion (see Figure 1). Incoming information stems from the entorhinal cortex, and perforant path projections onto CA3 cells initiate retrieval in CA3 (without passing through the DG). The process of pattern completion itself is principally subserved by the CA3 autoassociative network architecture (a network which can essentially retrieve a memory from partial information about the memory itself; Marr, 1971). This autoassociative CA3 architecture is considered to operate as a single attractor network (Rolls, 2013). Because of that, a retrieval cue need not be very strong in order to produce accurate recall – the retrieval process itself is taken over by the CA3 recurrent autoassociative system (Rolls, 2013; Treves & Rolls, 1992).

Completion is then carried out via CA3 projections to CA1 neurons, which then results in divergent backprojections from CA1 to the entorhinal cortex and subsequent neocortical areas. These backprojections occur through a Hebbian-like competitive learning network (i.e., associative learning where similar firing patterns result in strengthened connections), so that inputs from CA3 generate competition among the cells in CA1. Subsequently, cells with the strongest activation in CA1 instigate a winner-take-all effect, thereby strengthening that specific pattern and suppressing shared activation among other memory representations that were not completed. Thus, an anti-Hebbian effect takes place when the active neurons initiate inhibitory competition, thereby depressing activations from weakly activated neurons.

CA1 projections act as efficient retrieval cues (even partial or degraded), ultimately eliciting activity in those areas of the cerebral cortex that initially supplied input to the hippocampus. In other words, those areas of the brain that served to generate the initial episodic experience are again recruited upon retrieval. In this way, pattern completion can be conceptualized as “a reverse hierarchical series of pattern association networks implemented by the hippocampo-cortical backprojections, each one of which performs some pattern generalization, to retrieve a complete pattern of cortical firing in higher-order cortical areas” (Rolls, 2013, p.1).

**Pattern Completion as a Source of Continuously Generated Mental Content**

In addition to strengthening the existing memories, we also propose that pattern completion might serve as a source of the ceaseless change in mental content (i.e., the frequent
generation of new mental states). Indeed, a similar idea was proposed before by O’Neill, Pleydell-Bouverie, Dupret, & Csicsvari (2010), where pattern completion in the CA3 was suggested to be well-suited for promoting reactivation during rest. Further, there is evidence that spontaneous reactivation of a memory can be triggered by partial cues from the memory’s retrieval context, as evidenced by qualitative overlap between thought content and its cue (Berntsen, 1996; Berntsen & Hall, 2004).

We therefore consider the possibility that memories recalled during pattern completion might provide partial or degraded cues that may then serve to trigger further pattern completions, thus facilitating a continuous change in mental contents. For example, one might see a chocolate cupcake. Chocolate may then become a cue to complete a memory of a birthday party with chocolate cake. The cue of birthday might lead to completing a memory of the ninja turtles, and green may serve as a cue of to remember a favorite green t-shirt. This continuous cue provision and pattern completion tendency may help us understand why the mind keeps moving, with novel mental contents emerging repeatedly.

If, as we propose here, there is a bias for consecutive spontaneous mental simulations to be de-correlated via pattern separation processes, these partial cues are likely to trigger patterns that are at least somewhat dissimilar to the immediately preceding pattern that was triggered. This might be one reason spontaneous thought exhibits a heightened variability over time, while at the same time allowing for thematic relationships or other partial associations to be present among consecutive mental states. In turn, the completed patterns may also work together with pattern separation processes to further differentiate episodic events by strengthening the hippocampal-neocortical representations of an episodic memory when it is reactivated.

Cascades of thought might spontaneously arise within the hippocampus and propagate throughout the brain (Ellamil et al., 2016). Some of these thoughts may end up being experienced consciously, whereas others may fail to reach awareness. This account is consistent with the notion of thoughts shifting in and out of the foreground of one’s focus of attention, and the accompanying subjective experience of competing or co-existing streams of thought. It also speaks to the ease with which we experience a high level of content variability from one moment to the next, a large proportion of which may unfold spontaneously from partial cues in the internal or external environment.
Semantic Optimization Hypothesis

How do we transform our fragmented episodic experiences into a meaningful understanding of our world – or what scientists call ‘semantic knowledge’? Prominent consolidation models, such as The Standard Consolidation Theory (Scoville & Milner, 1957; Squire, 1992; Squire & Alvarez, 1995; Squire & Zola, 1998) and The Multiple Trace Theory (Nadel & Moscovitch, 1997) – although not in full theoretical agreement – share a central assumption with regard to this episodic-to-semantic transformation: at the neural level, episodic memories for events are primarily hippocampus-dependent, whereas semantic memories rely primarily on neocortical substrates. Here, we propose that default variability not only supports the organization of episodic memory in the hippocampus and neocortex, but also supports the organization of semantic memory by providing the conditions necessary for efficient episodic-to-semantic transformation.

The creation of semantic knowledge out of episodic experiences is a gradual process that occurs across multiple instantiations (McClelland, McNaughton, & O’Reilly, 1995). Variability across instantiations plays an important role in semantic knowledge acquisition. A combination of similarity and dissimilarity across representations facilitate the extraction of regularities and the development of categorization (Gelman & Markman, 1986; Sloutsky, 2003). Similarity (i.e. overlapping features that should be extracted for meaning making) provides evidence for regularities within a category, whereas dissimilarity (i.e., specific differences in individual events) provides contrasting evidence that helps identify category boundaries. Moreover, the experience of repeated events in various contexts aids the encoding of relationships between its typical elements (Avrahami & Kareev, 1994). Over multiple subsequent exposures, these event elements are stored together in one schema, affording economical representations of semantic concepts (Nadel, Hupbach, Gomez, & Newman-Smith, 2012). As part of the Default Variability Hypothesis, we propose that spontaneous thought’s heightened content variability serves to support and optimize semantic abstraction by providing multiple mental simulations that are both similar and dissimilar in nature. A default mode of content variability in spontaneous thought may therefore provide a mechanism for generating contextually variable episodic simulations (both veridical and novel recombinations). The similarity in consecutive mental simulations can
provide the basis for abstracting general meaning and overarching categories through multiple exposures, while the dissimilarities can help ensure that one specific instance is not overlearned (e.g., if you only saw one breed of dog, you may not realize another breed was also a dog).

Aside from the variability of consecutive representations, gradual exposure is also considered to play a critical role. Gradual exposure, also referred to as *interleaved learning*, affords optimal semantic abstraction (McClelland et al., 1995). Based on evidence from connectionist models, interleaved learning is theorized to critically support the progressive refinement of stable representations at the conceptual level. Semantic representations resulting from interleaved learning are optimally flexible in assuming and reflecting “the aggregate influence of the entire ensemble of patterns” elicited across events, while simultaneously being resilient to large modifications due to exposure to a single episodic trial (e.g., catastrophic interference; McClelland, et al. 1995, p. 429).

Accuracy in neocortical conceptual representation formation is argued to be a function of both sample size (i.e., number of experiences being aggregated across) and learning rate, whereby a slower rate allows for a greater number of interleaved samples to be factored into each computed estimate (White, 1989). For example, after enough gradual exposure to the meaning of cat, a child would be less susceptible to fundamental misunderstandings of the cat category (e.g., classifying a small dog as a cat). However, if a child is shown 50 pictures of cats in one day, he or she may confuse a small dog for a cat one week later. Interleaved learning is assumed to operate by “basically causing the network to take a running average over a larger number of recent examples,” (McClelland, et al., 1995, p. 437). We propose that a default content variability in spontaneous thought serves as a mechanism for increasing the opportunities for interleaved episodic-to-semantic transformation. By combining spontaneous reactivations that are highly variable from moment to moment, but also have recurring themes over time (e.g., particular things that are relevant to goals or current currents), spontaneous thought may optimize the conditions for episodic-to-semantic abstraction and semantic memory organization overall.

Another important property of interleaved learning is that it can deter *catastrophic interference* (the loss of previously learned information due to the introduction of new information; McCloskey & Cohen, 1989). As commonly portrayed through the *AB-AC paradigm* (for more details, see McClelland, McNaughton, & O’Reilly, 1995), newly learned associations...
(AC) can exhibit retroactive interference upon a previously acquired set of associations (AB). In this example, AC can interfere with the ability to recall AB later—because AC has “replaced” our concept of the AB association. Avoiding catastrophic interference means that we can actually distinguish AB and AC as different instances in an overarching category, rather than letting exposure to one harm memory of the other.

If “what one learns about something is stored in the connection weights among the units activated in representing it” (McClelland, et al., 1995, p. 433), then abstraction or generalization is only possible to the extent to which conceptual pattern representations overlap (Hinton, Mclelland, & Rumelhart, 1986). Therefore, it is imperative that a system is not only capable of – but also capitalizes upon – the ability to extract shared properties among concepts, while simultaneously minimizing catastrophic interference. McClelland, McNaughton, and O’Reilly (1995) aptly highlight the existence of two independent yet complementary learning systems that meet both these needs: rapid acquisition at the hippocampal level (pattern separation and completion), paired with gradual interleaved learning consolidation at the neocortical level.

Taken together, we propose that the content variability that characterizes spontaneous thought supports episodic-to-semantic transformation and semantic memory organization by providing both increased variability and frequency over time in a set of samples – novel combinations of elements and reinstatements of episodic memories in new contexts – that facilitates the rapid extraction of regularities and generalization across them. In this way, spontaneous thought’s default variability plays a critical role in optimizing the efficient abstraction and organization of semantic memory.

**Other Potential Benefits of Spontaneous Thought**

**Novel association formation**

Another key prediction of the semantic optimization hypothesis is that novel association formations can arise out of MTL activity. Fox, Andrews-Hanna & Christoff’s (2016) expanded account of the hippocampal indexing theory (Teyler & DiScenna, 1986; Teyler & Rudy, 2007) suggests that the generation of novel thought is supported by the same mechanisms involved in spontaneous reactivation of memory traces. This is consistent with the idea that the organization
of memory traces in MTL regions is considered associative (Moscovitch, 1995, p. 279). In other words, immediate temporal contiguity or simultaneity will largely dictate which combinations of cues and ensuing memory reactivations will arise together.

In this way, novel thought patterns which are constructive or generative in nature can be potentially facilitated by spontaneous mental simulations, whereby randomization of emergent thought patterns might in part promote more flexible, as opposed to fixated, thinking (Fox, Kang, Lifshitz, & Christoff, 2016). In fact, it has been shown that noise or variability in attractor networks is indeed beneficial for decision-making and memory, because it causes them to be non-deterministic, which in turn can cultivate new problem solutions and creativity (Deco, Rolls, & Romo, 2009; Rolls, 2013, 2014). Furthermore, spontaneous thought has also been recognized as supporting many constructive cognitive functions (Fox & Christoff, 2014; Fox, Kang, et al., 2016; McMillan, Kaufman, & Singer, 2013; Smallwood & Andrews-Hanna, 2013), including generation of creative solutions and ideas to present problems (Baird et al., 2012; Campbell, 1960; Simonton, 1999), simulated thinking (Rice & Redcay, 2015; Spiers & Maguire, 2006), and coordination and planning of future goals (Smallwood & Andrews-Hanna, 2013; Spreng, Stevens, Chamberlain, Gilmore, & Schacter, 2010).

According to the Default Variability Hypothesis, mind-wandering and spontaneous thought activity can be considered not only a mechanism involved in consolidating past episodes, but also processing ongoing current concerns and upcoming future events, a system which is expected and theorized to continuously update and integrate new information into existing semantic knowledge.

**Reconciling Existing Mind Wandering Frameworks**

Several theories of spontaneous thought and mind-wandering have been proposed, yet there is a lack of consensus about functional role(s) and underlying mechanisms. Smallwood (2013) recently attempted to differentiate two accounts: the first explained why the spontaneous onset of unconstrained self-generated mental activity arises (deemed ‘occurrence’ hypotheses), while the second explained how the continuity of internal thought is maintained once initiated (i.e., ‘process’ accounts).

Although his process-occurrence model does help unify the various accounts under one framework, it falls short of providing a functional reason for why the mind evolved to wander.
Instead, Smallwood suggests that the prospective consolidation hypothesis might be a possible explanation for the source and function of internally-generated thought (Smallwood, 2013). The prospective consolidation hypothesis suggests that “a core function of the hippocampal system is to make predictions about upcoming events” (Buckner, 2010, p. 42). Our hypothesis extends this idea to also include the reactivation of past and current information. Specifically, the Default Variability Hypothesis provides further insight into the question of why we have evolved to produce spontaneous thought marked by heightened variability of content over time. First and foremost, we propose a functional account for why spontaneous thought is such a prevalent and ongoing experience in daily waking life, and the mechanisms that support this ongoing mental activity. Second, we suggest that from one moment to the next, high levels of content variability – thoughts that seem unrelated to each, or only loosely related – are capable of arising quickly, ranging and shifting between past and current episodic reactivations to future-related simulated events. Finally, the current account takes a step away from the traditional task-centric literature, and suggests that this ongoing mental activity persists in both the presence and absence of external input.

As such, the content of spontaneous thought itself may be partially determined by simple random probability of thought pattern reactivation, as determined by any incoming externally- or internally-generated partial cues, paired with the effect of constraints acting upon the cognitive system within each given moment (Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). These constraints might be a function of salience, whether personal or perceptual in nature (e.g., the current concerns hypothesis; Klinger & Cox, 2004), the effect of attentional control (e.g., the executive failure hypothesis; McVay & Kane, 2009), a result of one’s capacity to identify the contents of one’s consciousness (e.g., the meta-awareness hypothesis; Schooler, 2002), or most likely the outcome of a combination of all of those, functioning to different degrees. In this way, it can be expected that mental contents are constantly emerging from within hippocampal structures, whereby the extent to which they are transformed into thoughts and unfold throughout the rest of the brain, and the extent to which they are likely to be experienced consciously, is determined by the level and specificity of those constraints.

**Conclusion**
Our minds frequently tend to ‘wander’ about, shaping a spontaneous thought flow marked by heightened content variability over time. Since the signature of free movement and content variability are likely to come at a considerable metabolic cost (Laughlin, de Ruyter van Steveninck, & Anderson, 1998; Plaçais & Preat, 2013), there is likely some evolutionary advantage of the dynamic nature of human thought. Thus, this chapter has attempted to introduce an account of the neural and cognitive evolutionary benefits of spontaneous thought and its inherent content variability.

Specifically, the Default Variability Hypothesis proposes that mind wandering is characterized by content variability and continuous movement which supports both efficient episodic storage (Episodic Efficiency Hypothesis) and semantic knowledge abstraction (Semantic Optimization Hypothesis). The Episodic Efficiency Hypothesis suggests that the reactivations and recombinations underlying content variability play a critical role in pattern separation by helping to de-correlate memories in the hippocampus and neocortex. Pattern completion, on the other hand, is proposed to strengthen the separated episodic memory representations, while also being a potential source of continuous mental content, where one activated memory serves as a partial cue for the next. In addition, the Semantic Optimization Hypothesis maintains that content variability supports episodic-to-semantic abstraction through multiple mental simulations that are both similar and dissimilar: The similarities provide the opportunity for repeated exposures so that concepts and categories can be strengthened over multiple exposures, while the dissimilarities mitigate the danger of overlearning a single instance. Through mental simulations, stemming from novel recombinations as well as reactivations, semantic abstraction is optimized due to increased variability and frequency over time in a set of samples containing similar yet dissociable information.
References


https://doi.org/10.1111/j.1467-9280.2008.02043.x


https://doi.org/10.1126/science.1152882


https://doi.org/10.3389/fnsys.2013.00074


https://doi.org/10.1016/j.nlm.2015.07.008


https://doi.org/10.1038/nrn2213


https://doi.org/10.1016/j.neuroimage.2010.06.016


Figure 1. Diagram representing the pathways for Pattern Completion and Pattern Separation from the cortex to the Medial Temporal Lobe (and back to the cortex).
Figure 2. Graphic illustration of how overlapping features are preserved and non-overlapping features are suppressed.
Figure 3. Examples of low variability in thought (top), corresponding to clustered learning, and highly variable thought content (bottom), corresponding to de-correlated memories via temporally spaced memories in spontaneous thoughts.