A Neural Model of Mind Wandering

Matthias Mittner,1,* Guy E. Hawkins,2 Wouter Boekel,2 and Birte U. Forstmann2

The role of the default-mode network (DMN) in the emergence of mind wandering and task-unrelated thought has been studied extensively. In parallel work, mind wandering has been associated with neuromodulation via the locus coeruleus (LC) norepinephrine (LC-NE) system. Here we propose a neural model that links the two systems in an integrative framework. The model attempts to explain how dynamic changes in brain systems give rise to the subjective experience of mind wandering. The model implies a neural and conceptual distinction between an off-focus state and an active mind-wandering state and provides a potential neural grounding for well-known cognitive theories of mind wandering. Finally, the proposed neural model of mind wandering generates precise, testable predictions at neural and behavioral levels.

Mind Wandering and the Brain

Mind wandering, or engaging in trains of thought that are unrelated (or unhelpful) to current task goals, is common in daily life [1]. In recent years mind wandering has received considerable attention in the cognitive neurosciences, with a particular focus on uncovering its neural origins. Because mind wandering appears to be a pervasive state of mental functioning, exploring its underlying mechanisms may tell us much about the human brain. In particular, understanding the causes of the attentional fluctuations that underlie mind wandering can help to identify separate brain states in which information processing is differentially affected.

The DMN (see Glossary) is strongly implicated in mind wandering [2-4]. The DMN is one of the most widely studied intrinsic connectivity networks (ICNs) [5] and includes nodes such as the medial prefrontal cortex (mPFC), the posterior cingulate cortex (PCC), the precuneus, and both angular gyri. These regions are reliably activated in the absence of a task (i.e., resting-state fMRI sessions; for a review see [6]), although it is unlikely that the DMN is a purely task-negative network [7-9]. The DMN is also involved in autobiographical planning and internally guided thoughts [10,11]. Generally, activity in the core DMN nodes is positively related to mind wandering as indicated by introspective thought sampling and attentional lapses in the form of behavioral errors [3,12,13]).

Simultaneously, a second neural system - the LC-NE system - has also been studied as a potential neural modulator of mind wandering [4,14,15]. Norepinephrine is assumed to control an alerting system that produces and maintains optimal levels of vigilance and performance [16,17]. A great deal of research has investigated the role of norepinephrine within the LC-NE system in supporting sustained attention (for reviews see [18,19]) or attentional lapses [20]. The dynamics of the LC-NE system are commonly separated into slow, tonic fluctuations and fast, phasic responses to stimuli that are connected via an inverse U-shape relationship: When tonic LC activity is low or high, performance-relevant phasic responses are attenuated. Measuring these dynamics is difficult because of the small size of the LC (Box 1). Instead, activity of the LC-NE
Probabilistic maps of the LC in standard space can be used to provide an accurate region of interest (ROI) for the investigation of LC signal. However, the position of the LC might vary between individuals to such an extent that standard-space probabilistic LC maps may not provide sufficient spatial precision. This problem is exacerbated by other factors such as age-related alterations in LC signal [54]. To obtain a more precise ROI of the LC, future studies would benefit from acquiring an individual, LC-tailored (e.g., the T1 TSE) sequence for each participant.

An intermediate level of tonic LC activity is likely to be required for optimal information processing; decreased or increased tonic levels are counterproductive in the sense that performance on a primary task suffers. The role of tonic LC-NE activity has been conceptualized in terms of an exploration-exploitation tradeoff [18]. In this framework, intermediate levels of tonic norepinephrine help to efficiently solve the task at hand because transient bursts of norepinephrine allow efficient selection of the most salient action in a multilayered neural network [24]. In this sense, intermediate levels of LC-NE activity are optimal. If tonic LC-NE levels increase relative to the optimum, the brain enters an exploratory mode where incidentally high activations can evoke response patterns that otherwise would not be strong enough to cross the threshold.

Functional connectivity and gain modulation of the LC-NE system may also be linked. Using large-scale simulations, a recent paper [25] showed that increases in neural gain entailed stronger functional connectivity. This finding was validated experimentally: blocks with increased baseline pupil diameter had stronger functional connectivity between brain regions. As neural gain increases there is a shift from widely distributed patterns of neural processing to tightly clustered patterns dominated by the strongest connections. Because this high-gain mode has
also been characterized as facilitating exploration [18], it can be interpreted as an unstable state in which all highly interconnected networks can potentially become dominant and drive behavior. This notion is similar to the ‘network reset’ theory of phasic LC-NE functioning proposed on the basis of experimental work in rodents and nonhuman primates [26]. Therefore, while short, phasic increases in LC-NE promote optimal responding by facilitating action selection, tonically high levels may cause incidental activations in task-unrelated networks to become dominant, hence shifting the focus of attention away from the task.

We argue that recent findings concerning the interaction of different brain networks can help to further specify this view of processing in the high-gain mode of mind wandering [27-29]. A study investigating the convergence of neural networks to local brain areas [27] provided evidence for the simultaneous ‘echoing’ of signals from different ICNs within subparts of specific brain structures. This means that the temporal dynamics of many independent ICNs were locally represented in spatially separate subparts of the PCC [29] and other areas including the mPFC [27], which raises the intriguing possibility that these nodes might serve as a global workspace [30,31]. Notably, the most prominent multinetwork echo structures - the PCC and mPFC - constitute the core nodes of the DMN, which is consistent with existing results that the PCC and mPFC are integrating, transmodal nodes [32]. These findings suggest that the DMN comprises two subnetworks - the dorsal medial (DM) and medial temporal lobe (MTL) subsystems - that are connected and coordinated by two core hub structures, the mPFC and the PCC. This idea is further corroborated by research on the widespread functional and anatomical connectivity of the PCC, supporting its role as a cortical hub [33].

Taking these findings together, rather than being a unified system the core nodes of the DMN might reflect a summation of converging activity from different ICNs. An important implication is a reinterpretation of the frequently observed task-related deactivation of the core DMN nodes during experimental tasks. Rather than being evidence for the direct relationship of DMN activity and mind wandering, task-related DMN deactivations could be a mere side effect of a lower number of functionally specific ICNs being active during the processing of most simple experimental tasks.

A Neural Model of Mind Wandering

We propose a neural model of the emergence of mind wandering that integrates findings regarding ICNs and the LC-NE system. A key feature of our proposal is a movement away from the idea that the core DMN nodes, the PCC and mPFC, are directly involved in mind wandering and toward a reinterpretation of these nodes as integrative, transmodal processing units. These units adjust their activity according to the functionally specific large-scale networks that converge onto them; this would mean that the PCC and mPFC are simply common ‘flags’ of other, broader network processes. Instead, we propose that the driving force behind attentional focus is the LC-NE system: norepinephrine fluctuations determine a global processing state that influences efficiency in solving a task or engaging in mind wandering. As a consequence, our model proposes a fundamental difference between an exploratory ‘off-focus’ state and active mind wandering.

Figure 1 illustrates the proposed model and Table 1 presents predictions of the model. When a participant starts performing an experimental task, engagement and motivation are initially high. This is reflected in an intermediate level of LC activity resulting in optimal neural gain (Figure 1, bottom left). In this state, brain networks that are necessary to efficiently solve the task are active (e.g., the dorsal attention network) while other networks that mainly involve functions unrelated to the task are deactivated (e.g., networks involved in memory retrieval or introspection). Because relatively few networks converge on the PCC and mPFC transmodal hub nodes (only those few activated by a simple experimental task), these nodes show relative deactivation.
Because both core regions (PCC, mPFC) and subnetworks (DM, MTL) of the DMN are weakly involved, this state shows a general deactivation of the DMN.

Attention is limited in duration and constant re-engagement or refocusing of the system is required, previously described as an ‘endogenously controlled refresh system’ [34]. Thus, the focus of attention is periodically broadened, accompanied by a more exploratory state reflected in higher levels of tonic norepinephrine and, hence, high neural gain (Figure 1, top). The off-focus state is also accompanied by higher functional connectivity [25] and higher activity of the DMN. This effect is due to the convergence of the simultaneous activity of many different ICNs involved in the various cognitive functions corresponding to the exploratory nature of the off-focus state. Because high gain increases functional connectivity within and between networks, activity from many networks converges on the PCC and mPFC transmodal hub nodes, resulting in a relative increase of activation in these nodes and episodes of less-efficient task processing [12]. Activation of the transmodal nodes allows selection of a new behavioral goal (i.e., exploitation), which may be to return to task processing or engage in mind wandering.

The probability of engaging in task-unrelated thoughts (or mind wandering) increases when the perceived attractiveness of internal processing exceeds that of actively solving the task (Figure 1, bottom right). This might happen, for example, when thoughts turn to a pressing, subjectively important issue (e.g., a ‘current concern’ [35]) or when motivation has declined due to prolonged exposure to a monotonous task. In this way the concentration on an internal goal during mind wandering is similar to the concentration on an external experimental goal in the on-task state. Neurally, the ICNs corresponding to the functions involved in pursuit of the internal goal are primarily engaged (e.g., the MTL subsystem of the DMN if the content of mind wandering involves projection of the self into the future [32]). As with the on-task state, during mind wandering functional connectivity would be reduced due to the differential engagement of relatively few networks and the PCC and mPFC transmodal nodes are expected to show
Table 1. A Complex Interplay of Brain Systems Gives Rise to Differentiable States of Mind Wandering

<table>
<thead>
<tr>
<th>State</th>
<th>LC-NE</th>
<th>Pupil Diameter</th>
<th>Network Connectivity</th>
<th>Network Activity</th>
<th>Behavior</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>On task</td>
<td>Tonic: optimal</td>
<td>Baseline: intermediate</td>
<td>Intermediate</td>
<td>Task-related networks increased (e.g., DAN); 'internal networks' decreased (e.g., MTL/DM subsystems of DMN); mPFC/PCC decreased</td>
<td>Low variability</td>
<td>Participant is focused on the task and performance is near optimal; distracting thoughts are effectively suppressed</td>
</tr>
<tr>
<td></td>
<td>Phasic response: strong, time locked to stimuli</td>
<td>Response: strong, time locked to stimuli</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off focus</td>
<td>Tonic: high</td>
<td>Baseline: increased</td>
<td>Strong</td>
<td>All ICNs intermediate; mPFC/PCC increased</td>
<td>Increased variability relative to on task</td>
<td>Exploratory state; alternative behavioral/response options are considered; because many functionally specific networks converge on the mPFC/PCC, they show increased activity</td>
</tr>
<tr>
<td>(exploratory)</td>
<td>Phasic response: reduced</td>
<td>Response: reduced</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mind wandering</td>
<td>Tonic: optimal</td>
<td>Baseline: intermediate</td>
<td>Intermediate</td>
<td>Task-related networks decreased (e.g., DAN); internal networks increased (e.g., MTL/DM subsystems of the DMN); mPFC/PCC decreased</td>
<td>High variability relative to on task</td>
<td>Participant focused on internal train of thoughts; only highly automatized behavior still functional, unexpected stimuli result in errors</td>
</tr>
<tr>
<td></td>
<td>Phasic response: strong, time locked to internal events</td>
<td>Response: strong, time locked to internal events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
reduced activity relative to the off-focus state. In addition, we expect to see transient bursts of LC-NE activity, albeit not locked to external stimuli but to internal events and hence difficult to measure.

Over the course of the experiment, the participant switches between the on-task and mind-wandering states. We argue, however, that it is always necessary to proceed through the proposed off-focus state. Recent research has shown that mind wandering comprises a complex, multifaceted pattern involving episodic thought, emotion, executive control, and meta-awareness in a component process account [36] featuring intricate combinations of corresponding neural responses [37]. Furthermore, intention during mind wandering has recently been identified as a key dimension with great explanatory power [38-40]. Our model is less concerned with such precise specification of the mind-wandering state and instead emphasizes the dynamics of the transitions between different attentional states. Qualitatively distinct types of mind wandering are implemented in our model as separate brain states involving different, specialized brain networks. Therefore, the mind-wandering state proposed in our model is not a unitary state: it represents a collection of many different states that share the feature of internally guided cognition, each with potentially different goals, meta-awareness, and emotional associations. As a consequence, shifting between qualitatively different types of mind wandering would also involve a transition through the proposed off-focus state and back to the mind-wandering state.

Implications of the Model

Our model integrates converging empirical findings into a cogent theory that lays the foundation for the next wave of hypothesis-driven research into the neural underpinning of mind wandering. First, the model leads to specific, testable predictions at the neural and behavioral level. Second, it provides a working hypothesis to resolve opposing views. Third, it is consistent with existing, largely qualitative, cognitive perspectives on mind wandering.

The model leads to a set of predictions to guide future research into mind wandering in particular and attention in general (Table 1). The most important implication of the model is that the assumption of a unified concept of mind wandering is an oversimplification. Observed behavior and brain activity studied under the label of mind wandering might arise from the proposed off-focus state or the active mind-wandering state. As a consequence, studies must carefully specify which of these phenomena is being investigated.

The model also resolves previous inconsistencies in the exploration-exploitation tradeoff as well as other paradoxical findings recently described in relation to the DMN. Several recent studies found that activity in the DMN is inversely related to measures of behavioral variability (i.e., poorer task performance) [41-43]. For example, in a finger-tapping task increased tap variability was associated with reduced DMN activity [43] although behavioral variability is consistently associated with mind wandering [44,45]. Finger tapping does not feature any external stimulation and therefore is prone to high levels of mind wandering. Our model links an active state of mind wandering to reduced activation in the PCC/mPFC and poor task performance (Table 1), which would explain these finger-tapping findings. By contrast, studies that found a positive correlation between mind wandering and PCC/mPFC activity used sustained-attention tasks where mind wandering was sampled with thought probes [2-4,13]. In such a setting, it is likely that episodes of mind wandering are relatively sparse and most thought probes where participants indicated that they were off task were likely to mirror the state we described as off focus (exploratory) in this model.

The tripartite model describes mind wandering (but not the transient off-focus state) as an active, goal-oriented state in which internally guided cognition is pursued. This conceptualization fits
well with findings indicating that brain networks involved in cognitive control (e.g., the fronto-parietal network) are also active during episodes of mind wandering [3,46-48], indicating that these networks are involved in actively guiding internal trains of thoughts or protecting it against external stimuli.

The model distinguishes between different on-task, off-focus, and mind-wandering states, which aligns with the general consensus among researchers that there are different stages of mind wandering. One popular theory of mind wandering is the perceptual decoupling hypothesis. Several studies have shown that mind wandering results in loss of sensitivity to sensory stimuli [11,49] and that the DMN is involved in this process [10]. The model we propose here can be interpreted as a neural implementation of the perceptual decoupling hypothesis, where coupling with the visual and saliency networks is reduced in favor of the networks involved in mind wandering.

Our neural model is also consistent with an insightful introspection of the phenomenon that proposes a hierarchical set of qualitatively different levels of mind wandering [50]. These authors proposed that an episode of mind wandering starts with a shallow detachment from the current

**Box 2. Cognitive Effects of Mind Wandering**

Mind wandering impairs performance in ongoing behavioral tasks, leading to higher error rates and more variable response times (e.g., [45,50]). Recent work has attempted to understand mind wandering-induced changes in behavior as the observed output of a change in latent task processing, via quantitative cognitive process models (for a review see [57]). Quantitative cognitive process models decompose observed variables, such as choices and response times, into latent components of processing that are typically of greater theoretical interest, such as information processing efficiency and cautiousness. In this way, cognitive models can address questions regarding how and why mind wandering affects observed performance during task completion.

Recent work has implemented cognitive process models in a model-based cognitive neuroscience framework. This allows mind wandering to be conceived as a neural state or process - as outlined in this Opinion article - that affects the latent components of cognitive process models, which in turn affects observed behavior. To date only one study has taken the first step toward an integrated model-based cognitive neuroscience of mind wandering [4] (Figure I). The general approach in this study can be extended to test empirically the tripartite neural model of mind wandering proposed in the main text. This extension requires the development of a dynamical component (e.g., a hidden Markov assumption [58,59]) that describes the transitions between the three states and an experimental paradigm that can discriminate the off-focus and mind-wandering states. This will allow not only experimental validation of the neural theory of mind wandering, but also quantitative study of the effect of the three neural states on cognition and behavior (cf., Table 1).

**Figure I. Overview of a Model-Based Cognitive Neuroscience Approach to Mind Wandering [4].** Neural data (fMRI, pupil diameter) were preprocessed to extract theoretically relevant features for use in a machine-learning classifier. Self-reported ratings of mind wandering were obtained during task completion for use as training labels in the classifier. After training the classifier, behavioral trials were assigned to on-task or off-task states, permitting quantitatively precise tests of the neural and behavioral signatures of the two states. Reproduced, with permission, from [4].
task not unlike the partial detachment of our off-focus state. This state has also been referred to as ‘tuning out’ [51] and has been described as allowing almost unimpaired performance in the primary experimental task, albeit characterized by increased variability. In a second, deeper state of mind wandering, participants continue performing the task on a superficial level while actively engaging in task-unrelated thoughts: ‘zoning out’, which corresponds to our exploitation-like state when internal goals are being pursued. The deepest level of mind wandering features an almost total lack of responsiveness to task-related stimuli, which in our model would correspond to a strong commitment to internal goals resulting in highly impaired performance.

Concluding Remarks

The neural correlates of attentional fluctuations and mind wandering are complex, involving regionally specific activity fluctuations, dynamic connectivity fluctuations, and neuromodulatory effects. We argue that understanding this complex pattern of results necessitates theoretical and methodological integration of all relevant effects in a comprehensive model (see Outstanding Questions). Here we proposed an empirically and theoretically driven framework that has the potential to explain results from all of these measures. We believe that focusing on one of the neural measures in isolation can lead to an oversimplified pattern of results. It is essential for future studies to simultaneously collect data reflecting the involvement of the various neural components, which will require the development of better neuroimaging protocols. It is, for example, notoriously difficult to measure blood oxygen level-dependent (BOLD) activity in the LC using fMRI (Box 1), although this is highly desirable to better understand the impact of the LC-NE system on mind wandering in particular and goal-directed cognition more generally. It is also necessary to develop sophisticated methods of analysis that integrate the separate measures in a formal framework and relate them to behavior (Box 2). Comprehensive data-based models of mind wandering will also be useful to those who are not studying the intricate phenomenon of mind wandering. In experiments designed to investigate other cognitive processes (e.g., decision making [52]), mind wandering will inevitably occur and obfuscate the phenomenon under investigation. In these cases, isolating and eliminating this source of noise using a suitable model (of mind wandering) can reveal new insights into the cognitive constructs under investigation.

Acknowledgments

This work was supported by a Vidi grant from the Dutch Organization for Scientific Research (NWO) (B.U.F.) and a starter grant from the European Research Council (ERC) (B.U.F.). The authors thank Sander Neuwenhuis, Quincy Rondeel, Sarah Hilles, and Gábor Csíkszaik for helpful discussions.

References


Outstanding Questions

What are suitable experimental paradigms to empirically discriminate the off-focus and mind-wandering states? Can a model-selection procedure based on, for example, a hidden Markov assumption provide evidence for the dynamical switching mediated by the off-focus state?

How are the components of the component process account of mind wandering [36] related to the off-focus and mind-wandering states? How are executive control and meta-awareness related to these states?

Can the phenomenon of mind blanking [60] be explained in terms of prolonged time in the off-focus state?

Is it possible to replace introspective measures of mind wandering with more objective, neural-based measures?

How are the identified electrophysiological and neuroimaging correlates of mind wandering related and can they be simultaneously measured and modeled?

What is the best way to quantify tonic and phasic LC-NE parameters using pupilometric signals? Can these measures be validated using in vivo imaging of the human LC?

On what timescale do the temporal dynamics of the human attentional system operate? Is it possible to capture them using dynamical extensions of cognitive process models?

Can mind wandering be actively influenced by pharmacology or brain stimulation and what are the implications for related psychopathological conditions?