UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Toward a model-based cognitive neuroscience of mind wandering

Hawkins, G.E.; Mittner, M.; Boekel, W.; Heathcote, A.; Forstmann, B.U.

DOI
10.1016/j.neuroscience.2015.09.053

Publication date
2015

Document Version
Final published version

Published in
Neuroscience

License
CC BY

Link to publication

Citation for published version (APA):
Hawkins, G. E., Mittner, M., Boekel, W., Heathcote, A., & Forstmann, B. U. (2015). Toward a model-based cognitive neuroscience of mind wandering. *Neuroscience*, *310*, 290-305. https://doi.org/10.1016/j.neuroscience.2015.09.053

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:16 Oct 2026

Neuroscience 310 (2015) 290–305

NEUROSCIENCE FOREFRONT REVIEW

TOWARD A MODEL-BASED COGNITIVE NEUROSCIENCE OF MIND WANDERING

G. E. HAWKINS,[a]* M. MITTNER,[b] W. BOEKEL,[a] A. HEATHCOTE[c] AND B. U. FORSTMANN[a]

[a] *Amsterdam Brain and Cognition Center, University of Amsterdam, Nieuwe Achtergracht 129, 1018 WS Amsterdam, The Netherlands*

[b] *Department of Psychology, University of Tromsø, Huginbakken 32, 9037 Tromsø, Norway*

[c] *School of Medicine – Division of Psychology, University of Tasmania, Social Sciences Building, Private Bag 30, Sandy Bay, Tasmania 7005, Australia*

Abstract—People often ''mind wander'' during everyday tasks, temporarily losing track of time, place, or current task goals. In laboratory-based tasks, mind wandering is often associated with performance decrements in behavioral variables and changes in neural recordings. Such empirical associations provide descriptive accounts of mind wandering – *how* it affects ongoing task performance – but fail to provide true explanatory accounts – *why* it affects task performance. In this perspectives paper, we consider mind wandering as a neural state or process that affects the parameters of quantitative cognitive process models, which in turn affect observed behavioral performance. Our approach thus uses cognitive process models to bridge the explanatory divide between neural and behavioral data. We provide an overview of two general frameworks for developing a model-based cognitive neuroscience of mind wandering. The first approach uses neural data to segment observed performance into a discrete mixture of latent task-related and task-unrelated states, and the second regresses single-trial measures of neural activity onto structured trial-by-trial variation in the parameters of cognitive process models. We discuss the relative merits of the two approaches, and the research questions they can answer, and highlight that both approaches allow neural data to provide additional constraint on the parameters of cognitive models, which will lead to a more precise account of the effect of mind wandering on brain and behavior. We conclude by summarizing prospects for mind wandering as conceived within a model-based cognitive neuroscience framework, highlighting the opportunities for its continued study and the benefits that arise from using well-developed quantitative techniques to study abstract theoretical constructs.
© 2015 The Authors. Published by Elsevier Ltd. on behalf of IBRO. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Key words: mind wandering, task-unrelated thoughts, model-based cognitive neuroscience, sequential sampling model, mixture model, single-trial regression.

Contents

Introduction 290
Computational models of cognitive processes 292
Sequential sampling models of decision-making 292
Contrasting qualitative and quantitative models of cognition 294
Latent mixture model frameworks for classifying mind wandering into task-related and task-unrelated cognitive states 295
Using behavioral data to inform latent mixture models 296
Using behavioral data to inform HMMs 297
Using neural data to inform machine learning approaches 298
Limitations of mixture model approaches to classifying mind wandering 299
Regression frameworks for dynamically tracking transitions between task-related and task-unrelated cognitive states 300
Using neural data as single-trial regressors on the parameters of cognitive process models 300
General discussion 301
Using model-based methods to study mind wandering in a broader context 301
Acknowledgments 302
References 302

INTRODUCTION

People often ''mind wander'' during everyday tasks, temporarily losing track of time, place or current task goals. Some estimates suggest that mind wandering might occupy anywhere between 30% and 50% of our everyday life (Killingsworth and Gilbert, 2010). Prominent theories of mind wandering suggest that monotonous tasks cause people to drift between various cognitive states (e.g., Cheyne et al., 2009; Schooler et al., 2011; Smallwood and Schooler, 2015). Such states can be classified as *on task*, reflecting an external focus on the present stimulus environment, and *off task*, characterized by internally directed cognitions that are largely decoupled from the external perceptual environment (e.g., Bastian and Sackur, 2013; Mittner et al., 2014). Mind

*Corresponding author. Tel: +31-20-525-5648.
E-mail address: guy.e.hawkins@gmail.com (G. E. Hawkins).
Abbreviations: ADHD, attention deficit hyperactivity disorder; EEG, electroencephalography; fMRI, functional magnetic resonance imaging; HMMs, hidden Markov models; MEG, magnetoencephalography; RTCV, coefficient of variation of response times; SART, sustained attention to respond task; STN, subthalamic nucleus.

http://dx.doi.org/10.1016/j.neuroscience.2015.09.053
0306-4522/© 2015 The Authors. Published by Elsevier Ltd. on behalf of IBRO.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

wandering has received increased interest over the past decade from both behavioral and neural angles (Weissman et al., 2006; Mason et al., 2007; Christoff et al., 2009; Killingsworth and Gilbert, 2010; Wilson et al., 2014; Smallwood and Schooler, 2015). Nevertheless, there have been few attempts to combine the behavioral and neuroscience approaches within a unified model-based neuroscience framework, in order to achieve a deeper and more coherent account of mind wandering (for an exception, see Mittner et al., 2014).

In the laboratory, mind wandering is often studied in the context of simple cognitive tasks, such as sustained attention tasks, where task performance is measured in terms of simple behavioral variables, such as choice accuracy or response time (Smallwood and Schooler, 2006). Throughout the task, participants are occasionally interrupted with 'thought probes' that ask the participant to make an introspective judgment whether they were on task or off task in the preceding trial or trials (e.g., Giambra, 1995; Smallwood et al., 2004; Seli et al., 2015a). Responses to thought probes have been used in various ways to classify experimental trials into on-task and off-task states, and those classified states are then related back to task performance (e.g., Christoff et al., 2009; Stawarczyk et al., 2011b; Mittner et al., 2014). Using this approach and similar variants, mind wandering has been related to performance decrements in the ongoing primary task in behavioral variables – such as higher error rates and response time variability (e.g., Cheyne et al., 2009; Bastian and Sackur, 2013) – and changes in neural recordings – such as increased activity in the default mode (or task negative) network (e.g., Christoff et al., 2009; Andrews-Hanna et al., 2010; Stawarczyk et al., 2011b; Smallwood et al., 2013).

Such empirical associations provide descriptive accounts of mind wandering – *how* it affects task performance – but do not provide true explanatory accounts – *why* it affects task performance, since extant theories of mind wandering do not provide generative accounts of cognition. Rather, empirical mind-wandering phenomena are explained at the level of verbal theorizing (we expand on this point in 'Contrasting qualitative and quantitative models of cognition' section). Verbal theories can easily lead to imprecise predictions and consequent difficulties in discriminating between competing theories of the processes of interest. For decades the psychological literature has explored why task performance changes as a function of experimental manipulations using quantitative cognitive process models. Cognitive models decompose observed behavioral variables in experimental tasks, such as choices and/or response times, into latent components of processing that are typically of greater interest for theory development, such as efficiency of processing and response caution. Thus, quantitative cognitive models have the potential to bridge the gap between abstract high-level theories and observed data, which moves toward mechanistic accounts of mind wandering.

In this article we outline mind wandering as conceived within a model-based cognitive neuroscience framework (Forstmann and Wagenmakers, 2015; Forstmann et al., in press), where cognitive process models bridge the explanatory divide between neural and behavioral data. In particular, we consider mind wandering as a neural state or process that affects the parameters of cognitive models, which in turn affect observed behavioral performance. We argue that adopting a quantitative cognitive modeling framework can provide a fresh perspective on various measurement issues and theoretical proposals from the mind wandering literature. We do not aim to provide a comprehensive review of the mind wandering literature (for excellent reviews, see Smallwood and Schooler, 2006, 2015), but rather provide a perspective on this novel approach. At the conclusion of the article we provide some broader perspectives on the application of the model-based cognitive neuroscience framework to the study of mind wandering, including applications to neuropsychological patients and psychopathology.

The frameworks we discuss aim to identify the occurrence and predictors of mind wandering during performance in discrete events (experimental trials) over an extended period (the course of an experiment). In this sense, we focus on identifying *when* people mind wander on a trial-by-trial basis; our approaches are agnostic about the content of task-unrelated thoughts. Our goal is to use cognitive models to move toward mechanistic accounts that explain what happens to ongoing task performance when the mind begins to wander. Our general approach, however, is more broadly applicable than just to the specific study of mind wandering itself. To the experimental psychologist studying another topic for example, task-unrelated thoughts are *contaminants* – trials influenced by a process that is not relevant to the cognitive process of interest. Even if one has no interest in studying mind wandering per se, the quantitative frameworks we discuss can be considered principled methods for removing contaminant trials from data sets.

Furthermore, we focus on alternatives to the routine adoption of introspective thought probe methods to identify mind wandering. Although thought sampling has furthered our current understanding of mind wandering, it suffers from a number of potential issues. For example, thought sampling may be subject to situational factors, such as social desirability biases or the observer effect, or limits on metacognitive abilities, such as the level of insight participants have into their current state (Smallwood and Schooler, 2006; Schooler et al., 2011; Seli et al., 2015a). Even if introspective judgments can reliably report on underlying states, we argue that they do not provide insight into the mechanisms that influence mind wandering or its effect on ongoing task performance. In the final sections of this review, we outline methods that conceptualize thought probes as an *outcome measure*, not the *identifier*, of mind wandering. That is, we argue that thought probes represent another source of data – just like choices and response times, or neural measures – that can constrain model predictions (i.e., be treated as a dependent variable) rather than the

indicator from which to examine other data (i.e., used as an independent variable to classify choices or response times).

COMPUTATIONAL MODELS OF COGNITIVE PROCESSES

Cognitive process models are quantitative implementations of theories about the processes involved in a range of cognitions – memories, attention, decisions, and so on. They permit precise quantitative tests of the potential cognitive mechanisms and processes that generate behavioral data. In particular, they allow the researcher to hypothesize and stringently test empirically the effect of experimental manipulations on cognitive processes, and to quantify the evidence for competing formal accounts of the processes under investigation. Here we provide a very brief overview of the advantages of cognitive modeling. For a comprehensive introduction we refer the reader to Lewandowsky and Farrell (2011) or Forstmann and Wagenmakers (2015).

Cognitive models decompose the distribution of observed variables, such as choice proportions and/or response times, into latent components of processing. These components, often referred to as model parameters, are of greater interest to theorizing than raw behavioral measures. For instance, the study of recognition memory often measures various response proportions across a number of experimental conditions. A model such as signal detection theory takes the raw response proportions – hit rates and false alarm rates, which can be ambiguous to interpret in isolation – and transforms them through the quantitative machinery of the model into constructs of greater theoretical interest, such as memory strength and response bias (Green and Swets, 1966).

In addition to a more coherent theoretical outlook, cognitive models can provide behavioral insights that cannot be obtained from analysis of raw behavioral data. The well-known speed-accuracy tradeoff, for example, describes how one can make faster decisions at the expense of accuracy, and vice versa (Reed, 1973; Pachella, 1974; Wickelgren, 1977). The tradeoff is defined by the relationship between the choices people make and the time taken to make them. Conventional analysis of choice or response time data – where each dependent variable is treated independently – cannot discriminate between accounts based on a speed-accuracy tradeoff or changes in the efficiency of information processing. The class of cognitive models known as sequential sampling models has been used to study the speed-accuracy tradeoff in great detail (e.g., Ratcliff and Rouder, 1998; Forstmann et al., 2008; Rae et al., 2014), as well as many other decision-related phenomena, and is the class of models we discuss in this review as they deal with similar issues as signal detection theory but generalized beyond choices to also account for response times. We note that, although we illustrate ideas as applied to sequential sampling models, the methods we highlight in this manuscript generalize to other classes of computational cognitive models.

Sequential sampling models of decision-making

We focus on mind wandering during simple yet attention-demanding tasks through the lens of sequential sampling models; well-developed cognitive process models that have provided great insight to the mechanisms underlying speeded decision-making in the psychology and neuroscience literatures (e.g., Busemeyer and Townsend, 1993; Usher and McClelland, 2001; Smith and Ratcliff, 2004; Brown and Heathcote, 2008; Ratcliff and McKoon, 2008). Sequential sampling models assume that simple decisions – such as whether a string of letters represents a word, or whether a motion stimulus moves in one direction or another – are made through a process of gradually accumulating sensory information to a threshold. Throughout the paper we discuss the diffusion model as an exemplar of the family of sequential sampling models (for reviews, see Ratcliff and McKoon, 2008; Voss et al., 2013; Forstmann et al., in press). We note that all methods outlined in this paper can equally well be used with other sequential sampling models, and debates about particular models and their assumptions are peripheral to our main thesis.

Fig. 1 provides a schematic overview of the decision process in the diffusion model. The model assumes that noisy information is gradually sampled from the stimulus. The information is accumulated in a decision variable that tracks support for one response option over another. The process continues until it reaches one of two decision boundaries, triggering a response. In Fig. 1, the decision is whether the stimulus is moving to the left or right of a display, so the boundaries correspond to a response of 'left' or 'right'. The predicted response corresponds to the boundary that was crossed, and the predicted response time is the time it took for the decision variable to reach the boundary plus an offset time that accounts for peripheral processes such as encoding the stimulus display and executing a physical response (such as a button press).

The parameters of the diffusion model, and sequential sampling models in general, relate to constructs that are relevant to our understanding of mind wandering. For example, the average rate of information accumulation – the drift rate – indexes the efficiency of information processing; the distance between the boundaries indexes the level of response caution; the starting point relative to the response boundaries indexes response biases, because the decision variable can start closer to one boundary than another; and the time taken for the aspects of response time not accounted for by the decision itself, known as non-decision time. Modern implementations of sequential sampling models also consider variability in model parameters from one trial to the next (Ratcliff and Tuerlinckx, 2002). For example, trial-to-trial variability in drift rates reflects the assumption that the efficiency of processing is variable over time (Ratcliff, 1978).

The utility of cognitive models such as the diffusion model relies critically on the validity of its latent constructs – the model parameters. One approach to ascertain the validity of the model parameters is through tests of selective influence; a priori hypotheses about

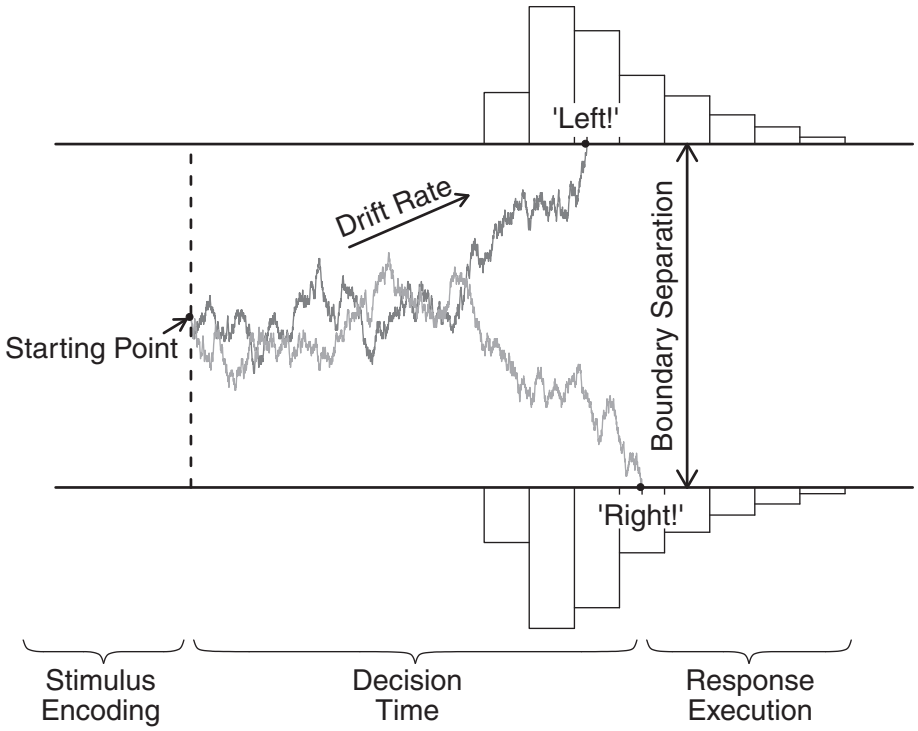

Fig. 1. Schematic illustration of the drift diffusion model of decision-making. We thank Don Van Ravenzwaaij for supplying the code to draw this figure.

the effect of experimental manipulations on particular latent constructs (for a detailed introduction, see Heathcote et al., 2015a). For example, Ratcliff and Rouder (1998) showed that manipulating task difficulty led to changes in the drift rate parameter (processing efficiency) but not boundary separation (response caution), and instructions to emphasize fast or cautious decision-making led to changes in boundary separation but not drift rate, even when both factors were manipulated simultaneously. Furthermore, provision of greater reward for one response over another leads to a shift in the start-point of information accumulation toward the higher reward boundary, and the non-decision time parameter increases when the motor component of the response is more challenging to produce (Voss et al., 2004). These results indicate that the diffusion model parameters have well validated interpretations.

The parameters of sequential sampling models can be inferred from behavioral data. From each experimental condition, the probability of a correct response and the observed distribution of response times for correct and error responses are used to infer the values of the model parameters that were most likely to have generated the observed data. In this way, the model decomposes observed variables, choices and response times, into parameter values that allow researchers to draw deeper conclusions, such as whether a change in response time across conditions is better described as a change in the efficiency of processing (drift rate) or cautiousness of responding (boundary separation). For tutorials on parameter estimation in sequential sampling models, we refer the reader to Donkin et al. (2009), Vandekerckhove and Tuerlinckx (2008), and Voss and Voss (2007).

One of the great benefits of estimating the parameters of cognitive process models is that the quality of the model fit to data can be used to determine which model from a set of candidates embodying different theoretical accounts provides the best account of a phenomenon (e.g., mind wandering). This procedure, referred to as model evaluation, comparison or selection, consists of a range of techniques that quantify the empirical support for models with different parameterization or different architectures (for a detailed review, see Shiffrin et al., 2008). For example, one may wish to compare which of two diffusion models provides the best account of an experimental effect, a model that attributes the effect to the drift rate parameter (processing efficiency) or boundary separation (response caution). Model selection techniques can also be used to select between models with different architectures; for example, whether a particular data set is best explained by the diffusion model or another sequential sampling model such as the linear ballistic accumulator (Brown and Heathcote, 2008; e.g., Heathcote et al., 2015b, in press). In the context of mind wandering, by evaluating which model parameterization or architecture provides the most parsimonious account of behavioral data we gain insight to the nature of the processes underlying mind wandering, a point we return to throughout the manuscript. When used appropriately, quantitative model evaluation decisively selects between competing theories of psychological processes. It is not possible to garner the same level of empirical support from qualitative or 'verbal' models, which are the most

common form of models in the psychological literature (Lewandowsky and Farrell, 2011).

Finally, in addition to the theoretical advantages of using sequential sampling models to understand data, at a practical level existing models are at a stage of maturity where they can be used to understand data in experimental paradigms that are commonly used to study mind wandering. For example, there are well-developed sequential sampling models of the go/no-go task (e.g., Gomez et al., 2007), commonly studied in the mind-wandering literature as the sustained attention to respond task (SART; Robertson et al., 1997; Smallwood et al., 2004; Smallwood and Schooler, 2006; Smilek et al., 2010). There are also variants of sequential sampling models that account for less commonly used tasks in the mind wandering literature, such as the stop-signal task (Logan et al., 2014).

Contrasting qualitative and quantitative models of cognition

In all scientific endeavors, data can only be interpreted and understood through the lens of a theory. Cognitive process models, such as the diffusion model, can be thought of as *quantitative* instantiations of a theory: a set of input parameters is transformed through a series of functions – the formalization of the model – to generate quantitative behavioral predictions. The predictions can be rigorously tested against data – via parameter estimation and model selection – to determine whether a model provides a good account of patterns in data and interpretable theoretical conclusions.

In contrast, existing theories of mind wandering can be thought of as *qualitative* theories. Qualitative theories are described in verbal terms and are thus less strictly defined than quantitative theories, leading to less precise behavioral predictions. Take as an example the performance decrements in laboratory-based tasks such as the SART that have been interpreted through the lens of different theories of mind wandering. One key hypothesis states that executive resources are used to perform goal-directed tasks, and this finite pool of resources is depleted when the mind disengages from the task – since task-unrelated thoughts consume resources – thus leaving fewer resources for the ongoing task, resulting in suboptimal performance (Smallwood and Schooler, 2006). This resource model is consistent with data showing that increased mind wandering is associated with poorer performance on resource-demanding tasks (Mrazek et al., 2012) and activation of executive networks (Christoff et al., 2009). An alternative explanation proposes that people switch between states of perceptual coupling – when attentional processes are directed to sensory input – and a task-disengaged state of perceptual decoupling – when attention is diverted from sensory input to inner thoughts (for detailed reviews, see Schooler et al., 2011; Smallwood and Schooler, 2015).

The resource and perceptual decoupling models provide intuitively appealing accounts of mind wandering, and both are able to explain general patterns in data. However, both theories posit different explanations for its occurrence, which raises the question of which model is best supported by data; when two (or more) verbal models predict qualitatively similar patterns in data it is not clear how to select between theories. A pertinent example is a related debate about whether mind wandering depletes executive resources or is the result of an executive failure (see McVay and Kane, 2010, 2012a,b; Smallwood, 2010). The very existence of this controversy highlights the difficulty of discriminating between theories formulated at an abstract, verbal level. Quantitative models differ in this respect: the functions that make up the model architecture quantitatively constrain model predictions, which provide greater ability to select between theories. To illustrate this point we borrow Lewandowsky and Farrell's (2011) example from the 'hard' sciences.

For centuries, it was believed that the sun and the planets orbited the earth according to Ptolemy's geocentric model of the solar system. Copernicus challenged the dominant geocentric model, proposing that the planets follow a circular orbit around the sun. The predictions of the Copernican heliocentric model provided an approximately equally good account of planetary motion as the Ptolemaic geocentric model. Without a metric to define a *good* account of the data – a quantitative modeling comparison – it would not have been possible to reach this conclusion. Since both models provided an equivalent fit to data, the more parsimonious Copernican model was eventually preferred over the Ptolemaic model. Later, Kepler proposed that planetary motion follows an elliptical rather than circular orbit. The Copernican and Keplerian models thus differed in a quantitative manner – circular versus elliptical orbits – but not at a qualitative level – both theories propose that planets orbit the sun. The Keplerian heliocentric model provided a more quantitatively precise account of the data than the Copernican model and was thus preferred. This example highlights a transition from theories that differ in a qualitative manner – a geocentric to a heliocentric model – to theories that differ in a quantitative manner – circular to elliptical orbits. Discriminating between the two heliocentric models was only possible through quantitative comparison of model predictions.

Returning to theories of mind wandering, if the resource and perceptual decoupling theories make similar verbal predictions – for example, that one experimental condition will make more errors than another – and this pattern is observed in data, it is difficult to discriminate between theories of the potential mechanisms or processes underlying mind wandering. Instantiating theories in a quantitative modeling framework more tightly constrains the prediction space; for example, two quantitative models of mind wandering might predict differences in the rate of increase in SART errors across conditions. Differences in predictions between models are thus more likely, because the predictions are more precise, leading to more decisive conclusions about theories. To borrow from physics once more, Einstein's description of the mass–energy equivalence ($E = mc^2$) would not be nearly as useful, or well-known, if it simply stated that mass can be

converted to energy; it is the formalization and precision of the stated equivalence that makes the theory so valuable.

The foregoing discussion is not intended to disregard theoretical advances in mind wandering. To the contrary, we argue that existing theories should be further developed to move them into the realm of quantitative models, to increase their explanatory power and provide greater insight to mind wandering. Moving from a qualitative to quantitative theory involves multiple steps. We illustrate the nature of some of the conceptual questions one must consider in this process using the resource model as an example. Relevant questions might involve, for example, how one formalizes (i.e., mathematically defines) 'resources'; whether people have a single pool of resources or multiple pools; whether the pool of resources is a fixed quantity or variable across time, contexts, and people; and how resources are functionally mapped and allocated to task performance. A major advantage of this approach is that each decision about the implementation of a particular theoretical assumption in the model can be quantitatively tested using corresponding models. Determination of which model provides the best account of the data discriminates among the underlying theoretical positions. We argue that theories of psychological constructs should be developed in such a quantitative, and not qualitative, framework, mirroring progress in other disciplines.

An illustrative example of this approach comes from the study of prospective memory, which involves remembering to perform an action or task at some time in the future, often while completing a primary task. Extant theories suggested that prospective memory tasks consume and re-direct resources from the primary task, which lead to decrements in ongoing task performance (e.g., Smith, 2003). In a sequential sampling model framework, a simple hypothesis would map remembering to do a prospective memory action, which reduces resources, to a reduction in drift rate on the ongoing task. A recent model-based analysis (Heathcote et al., 2015b) indicated this is not the case: people instead raised their level of response caution, consistent with the verbal theory that participants delayed their ongoing-task responses so that they do not preempt prospective memory responses (Loft and Remington, 2013). This was a counter-intuitive yet highly informative outcome that changes the course of theorizing about prospective memory. It is possible that a quantitative analysis of resource theories of mind wandering might lead to similar outcomes. Testing such hypotheses first requires those theories to be developed in a quantitative framework.

We do not pursue such extensive theoretical developments of quantitative models of mind wandering here. Rather, we use sequential sampling models as a vehicle to demonstrate how methods from model-based cognitive neuroscience can be used to advance the study of mind wandering and constrain future quantitative models of mind wandering. Likely the sequential sampling framework as it stands is too simple to provide a complete account of mind wandering. Nevertheless, we believe that it provides a useful starting point for the development of comprehensive yet quantitatively precise models of mind wandering.

We now present an outlook on general model-based frameworks that can be used to understand mind wandering as a mediator that drives the parameters of cognitive models, in particular sequential sampling models. The approaches differ in their psychological assumptions and the research questions they can address. Common across the frameworks, however, is the assumption that on-task and off-task states have different data-generating parameters, and these parameter differences mediate the observed behavioral effects. The proposals outlined below are presented as a sample of possible model-based frameworks to operationalize mind wandering and are by no means intended to provide an exhaustive overview of possible modeling approaches or conclusive theoretical insights regarding the information processing origins of mind wandering.

LATENT MIXTURE MODEL FRAMEWORKS FOR CLASSIFYING MIND WANDERING INTO TASK-RELATED AND TASK-UNRELATED COGNITIVE STATES

We use the term *mixture model* to refer to a class of methods that assumes the presence of discrete generating sources in the observed data. For example, one might assume that there are periods of high task engagement and periods of mind wandering, which we refer to as on-task and off-task states, respectively; this is similar to the perceptual decoupling theory of mind wandering (Smallwood and Schooler, 2015). In the mixture framework, the two states are assumed to be mutually exclusive and driven by different parameter values in a cognitive model. For example, the on-task state might be characterized by a larger drift rate than the off-task state, with all other parameter values remaining constant across the two states. Such a model would reflect the psychological assumption that mind wandering has a selective influence on the efficiency of task information processing. The aim of mixture modeling is to use the observed data to infer, separately for each decision trial, which of the latent or hidden (i.e., unobservable) states gave rise to the observed datum – the on-task process or the off-task process.

A coarse yet common approximation to identify mixtures in data is to assume that certain values – such as a response time faster or slower than a cutoff value – represent different processes. For example, while studying the effects of sleep deprivation, Ratcliff and Van Dongen (2011) used the convention of classifying responses slower than 500 ms in a psychomotor vigilance task as attentional lapses, which may indicate an increased propensity for mind wandering. This approach requires two explicit steps: researchers first select an appropriate cutoff value to categorize responses (e.g., 500 ms), sometimes after having observed the data, and then inference is performed on differences in the proportion of responses in the two categories. The problem with

this approach is that the inference in the second step assumes the cutoff value from the first step was determined a priori, and furthermore that it was the only way in which the data could have been divided (for further discussion of this issue, see Hawkins et al., in press-a). Even in situations when the value was derived from previous literature (as in Ratcliff and Van Dongen, 2011) it does not preclude the possibility that other sub-divisions of the data were possible.

Our review focuses on two methods for mixture modeling that overcome the aforementioned problems by identifying discrete, latent classes of responses: Bayesian latent mixture models and machine learning approaches that are informed by an independent stream of data.

Using behavioral data to inform latent mixture models

Although multiple approaches exist for estimating mixture models from data, we focus on Bayesian methods since they confer many benefits for cognitive modeling, including a one-step analysis for identifying mixture models, simultaneous estimation of participant- and group-level parameters via hierarchical modeling, and quantifying uncertainty in parameter estimates via posterior distributions over parameters. For an overview of the advantages of Bayesian parameter estimation methods and a practical guide to their implementation we refer the reader to Lee and Wagenmakers (2013).

Bayesian mixture models are conceptually straightforward and have been applied to a range of data analysis and cognitive modeling applications (e.g., Steyvers et al., 2009; Lee and Wagenmakers, 2013; Scheibehenne et al., 2013; Bartlema et al., 2014). We first assume that one or more discrete states or processes generated the observed data. As a toy example, we might assume that each observed datum from a set of continuous measurements (e.g., height) is drawn from one of two normal distributions (e.g., men and women). On average, the population mean of the male population is larger than the female population, but there is a considerable standard deviation such that some males are shorter than some females. Now consider the situation where we only have access to the measurements of height but not the sex of the person that provided each measurement. The computational problem is to use the observed distribution of height, and our assumption that both males and females contributed height measurements, to infer the properties of the two populations (i.e., the means and standard deviations of the male and female distributions, and the proportion of males to females) and to assign a probability to each datum that the person was male or female. The probability of classification to one class or the other is proportional to the prior probability of the two populations and the ratio of the density of the height measurement under the parameters of the respective population distributions.

It is conceptually simple to scale the Bayesian mixture model of the previous example to applications of psychological interest; for example, assuming that an observed data set is comprised of a mixture of on-task and off-task cognitive processes. Vandekerckhove et al. (2008) considered a similar case, by assuming that some trials in a perceptual decision-making experiment were *contaminants* – data points that were not generated by the (diffusion model) process of interest – and hence were not germane to the primary research question (for a related approach see Vandekerckhove and Tuerlinckx, 2007). Although Vandekerckhove et al. (2008) did not intend to study mind wandering, their approach indirectly modeled processes relevant to mind wandering: an experimental psychologist typically aims to identify contaminant trials to remove them from the analysis. The mind-wandering researcher aims to identify those same 'contaminants' and study the processes that generated them.

Vandekerckhove et al. (2008) defined a Bayesian latent mixture model that classified trials into one of three categories: decision trials generated from the diffusion model, guesses, and delayed startups. Here, we focus on Vandekerckhove et al.'s (2008) hypothesis that some trials have a delayed startup. Under this hypothesis there are two discrete types of diffusion process: the first determines performance when the participant is focused on the task at hand, and the second determines performance when the participant's focus is elsewhere – a contaminant, in Vandekerckhove et al.'s (2008) terminology. The on-task and off-task diffusion processes were assumed to have the same data-generating parameters except for a 'delayed startup' in the off-task process, implemented as a larger value of the non-decision time parameter than the on-task process.

Because quantitative models generate precise predictions, the assumption that the off-task process differed to the on-task process in only one parameter (non-decision time) still leads to dissociable predictions as compared to changes in another model parameter (say, drift rate). The delayed startup theory of mind wandering predicts that off-task trials lead to slower responses than on-task trials, on average. This occurs since an increase in the non-decision time parameter leads to a slower onset of the decision process, and therefore a global upward shift in the distribution of response times and longer decision latencies, on average. This account is plausible when a participant may be engaged in task-unrelated thoughts when the decision stimulus appears on screen. Once the stimulus appears it takes some amount of time for the participant to re-orient to the task at hand. Once re-focused, the participant makes a decision in an otherwise similar manner to trials where they were focused on the task at stimulus onset (i.e., with the same starting point, drift rate, and boundary separation as the on-task diffusion process). As this example illustrates, since the non-decision time parameter has no bearing on the diffusion process itself, the delayed startup theory predicts that the on-task and off-task processes do not differ in decision outcome (correct, error) or variability in response times.

Clearly, however, task-unrelated thoughts have been linked to not only changes in response times but also to decreases in decision accuracy (e.g., Cheyne et al., 2009; Stawarczyk et al., 2011a; McVay and Kane,

2012a; Bastian and Sackur, 2013; Seli et al., 2013). A strict interpretation of the delayed startup theory is, therefore, easily falsified. However, it is straightforward to augment the delayed startup process with an additional change in, say, drift rate. Such a change might be interpreted as a degraded diffusion process that took longer to start (larger non-decision time) and was less efficient (lower drift rate). This augmented model predicts longer response times, on average, as before, but the reduction in drift rate predicts decreased decision accuracy and increased variability in the distribution of response times. This change in model parameterization may then account for the empirically observed pattern of longer response times and an increased proportion of errors in off-task relative to on-task behavior. The key point is that a range of models assuming a mixture of discrete data-generating processes can be proposed and quantitatively tested against data to determine their appropriateness, and modified as necessary.

The delayed startup hypothesis and related proposals using Bayesian latent mixture modeling aim to classify trials into one of a number of mutually exclusive categories. A critical assumption of this framework is that each decision is considered an independent random sample from one of the generating distributions (i.e., on-task diffusion process, off-task or 'delayed startup' diffusion process), which means the approach assumes no sequential structure. This is at odds with our intuitions about mind wandering: one would expect an increased chance of an off-task trial following an off-task (versus on-task) trial. Indeed, the mind wandering literature has produced empirical results that are consistent with this intuition. For example, mind wandering-related variability in response times can follow phasic increases and decreases over the course of perceptual experiments (e.g., Bastian and Sackur, 2013; Bompas et al., 2015), and alterations in neural activity can precede performance deficits including erroneous responses up to 30 s before the behavioral outcome is observed (e.g., Eichele et al., 2008; O'Connell et al., 2009; Macdonald et al., 2011). These findings suggest that a more precise model of mind wandering ought to account for the temporal correlation of switching between on-task and off-task states, which can be obtained with hidden Markov models (HMMs).

Using behavioral data to inform HMMs

As in latent mixture modeling, HMMs, also known as dependent mixture models, use the observed output of a process (response times, decision accuracy) to infer the 'hidden' state or process that generated the data (on-task, off-task). However, and crucially, HMMs generalize latent mixture models by assuming the discrete generating states are related over time through a Markov process rather than independently distributed. This allows HMMs to estimate a critically informative piece of information for the study of mind wandering – *transition probabilities*; for any given trial, the probability of switching from an on-task state to an off-task state, or from an off-task state to an on-task state.

Reliable estimation of the transition probabilities in a HMM requires a signature or regularity in the data that is related to the discrete states of interest. In the mind wandering literature, for example, periods of increased response time variability have been associated with greater propensity for task-unrelated thoughts (e.g., Cheyne et al., 2009; Stawarczyk et al., 2011a; Bastian and Sackur, 2013; Seli et al., 2015a). Bastian and Sackur (2013) noted phasic increases and decreases in the coefficient of variation of response times (RTCV), a standardized measure of response variability, in the commonly studied SART, where self-reported ratings of task-unrelated thoughts were associated with larger RTCV. Bastian and Sackur (2013) used patterns in the observed RTCV to inform a HMM that inferred *go* responses tended to occur in 'runs' of on-task, and then off-task, states. Specifically, the authors estimated that the probability of switching from on-task to off-task from one trial to the next was lower than the probability of the reverse state change, switching from off-task to on-task (.11 versus .18). This result raises two important issues. First, the off-task or mind-wandering state was more volatile than the task-focused state (i.e., larger transition probability). Second, the transition probabilities were not complementary (sum to one) which implies that trials are not independently distributed according to a particular base rate of on-task versus off-task states; neighboring time points are more likely to be related than distant time points. The transition probabilities also allow one to derive the expected duration of runs of on-task and off-task states: the mean duration of an on-task episode was $1/.11 = 9.09$ units of experimental time, which translates to 18.2 s under Bastian and Sackur's (2013) division of experimental time (i.e., experimental sessions were split into many units each of 2 s duration), while off-task episodes were shorter, on average – 11.1 s ($1/.18 \times 2$).

In this context, HMMs provided great insight to the transition from task-engaged states to mind wandering and back again, including the intriguing proposition that the off-task state is more volatile than the on-task state. However, Bastian and Sackur's (2013) approach was restricted to a purely descriptive model of response time distributions (the ex-Gaussian model). Such descriptive models provide precise fits to data but their parameters lack an interpretation in terms of cognitive processes (Matzke and Wagenmakers, 2009), and thus provide limited insight into the cognitions driving task performance in the two discrete states. Although more research is required to rigorously link HMMs in mind wandering to cognitive process models, we can derive some predictions from Bastian and Sackur's (2013) results. In particular, the HMM used the variability of observed response times as a signal that discriminated between the two discrete states. Specifically, the distribution of off-task responses was more positively skewed than on-task responses, and hence contained more variable response times (i.e., larger RTCV). In sequential sampling models, the variance in predicted response time distributions can increase when (1) the drift rate decreases, indicating reduced efficiency of information processing, (2) there is larger trial-to-trial variability in drift rate, indicating greater

across-trial noise in the processing of ostensibly similar stimuli, or (3) there is greater boundary separation, indicating more cautious responding. Any of these potential outcomes could provide a neat mapping to theorizing that mind wandering increases the variability of responding as well as effects on the overall speed of responding.

Using neural data to inform machine learning approaches

The latent mixture model and HMM approaches assume there is a mixture of on-task and off-task states in the structure of the model, and then reverses the generative process to estimate the proportion of on-task versus off-task trials in the data. An alternative to the problem of inferring a mixture in data is to use data-driven methods where one stream of data (e.g., neural recordings) is used to classify another (e.g., behavioral data). There are varying levels of complexity in how neural measures can be used to classify trials as belonging to a particular state. We first provide a brief overview of a simpler and relatively common (non-machine learning) approach to using neural data to classify trials as belonging to a particular state, with a hypothesized example. We then outline what is, to our knowledge, the only application of machine learning approaches in the mind wandering literature (Mittner et al., 2014).

In the mind-wandering literature it has been hypothesized that increased power of pre-stimulus alpha activity is related to attentional lapses and the propensity to engage in task-unrelated thoughts (O'Connell et al., 2009; Macdonald et al., 2011; Bompas et al., 2015; though see also Braboszcz and Delorme, 2011). Alpha waves are neural oscillations in the 8–12 Hz frequency range measurable via electroencephalography (EEG) and magnetoencephalography (MEG). Although alpha activity has only recently become a focus of study in the mind wandering literature, it has been studied in depth in the attention literature, where it is generally found that alpha activity increases during wakeful rest and is thought to index disengagement from the external visual environment (e.g., Cooper et al., 2003; Ergenoglu et al., 2004; Van Dijk et al., 2008; Mathewson et al., 2009; Romei et al., 2010). Specifically, alpha oscillations are thought to reflect cortical inhibition of task irrelevant areas, a top-down control process that prevents irrelevant brain regions from interfering with task performance (Klimesch et al., 2007). One hypothesis is that state changes involved in mind wandering – transitioning from on-task to off-task – are associated with changes in the localization of alpha oscillations, such that cortical inhibition processes shift from task irrelevant areas to task relevant areas. This would lead to decrements in performance while simultaneously freeing the mind to wander.

To test the hypothesis that alpha activity is related to task-unrelated thoughts, one could use a classification approach that first sorts trial-level data on pre-stimulus alpha power recorded from task-relevant regions. The sorted data are partitioned into 'low' (i.e., more on-task) and 'high' (i.e., more off-task) alpha sets and a cognitive model is fit to the two sets of behavioral data (for a similar approach using multivariate pattern analysis in a related domain, see Ratcliff et al., 2009). Quantitative model comparison is used to determine the most parsimonious account of the data: which parameters should be estimated separately across the on-task and off-task trials and which should remain fixed. To the extent that pre-stimulus alpha activity is related to mind wandering, parameter differences across the low and high alpha sets can be attributed to the processes that differ between the two cognitive states. Although potentially insightful, there is at least one major drawback of the split-half approach: it imposes an artificial categorization ('low' versus 'high' pre-stimulus alpha activity) on a continuous measure (alpha power). This forced categorization is not necessarily meaningful since borderline trials are forced into one group or another (i.e., higher alpha power in the 'low' group, and vice versa). One can circumvent this problem by removing a middle segment of data, such as removing the middle third of trials with intermediate alpha power and only comparing the lowest third to the highest third; however such classification schemes use data inefficiently.

An alternative to median-split segmentation rules are data-driven, machine learning algorithms that are trained to classify trials as on-task or off-task on the basis of an observed variable or variables. For example, Mittner et al. (2014) had participants perform a stop-signal task, a common measure of response inhibition, while recording functional magnetic resonance imaging (fMRI) activity and pupil diameter (see Fig. 2). A thought sampling method was used where, at pseudo-random times throughout the behavioral task, the participant was asked to indicate whether their focus was on-task or off-task in the previous trial. The self-report ratings were used as labels for the neural data to train a classification algorithm to learn distinct patterns of neural activity that were predictive of the self-reported on- (or off-) task rating. Once trained, the algorithm probabilistically classified individual trials as on-task or off-task for the unlabeled (majority of) trials. Mittner et al. (2014) fit various independent stochastic accumulator models (variants of the diffusion model) to the data classified as on-task and off-task to determine which parameters provided the best account of the difference in performance between the on-task and off-task states. Quantitative model comparison was used to determine the most parsimonious account of the data, which was a model that indicated on-task relative to off-task trials were more likely to have larger drift rates for the *go* and *stop* processes (implicating more efficient stimulus processing), and a larger response threshold (implicating more cautious responding). Bode et al. (2012) provided a related example with applications to attentional processing that used multivariate pattern classification of EEG data.

The machine learning approach to trial classification has three primary advantages. Firstly, it provides a means for (probabilistically) classifying the on-task versus off-task status for *all* experimental trials. This overcomes a distinct downside to the common approach in thought sampling on-task versus off-task behavior that can only classify the trials that immediately preceded the presentation of a thought probe. Machine-learning

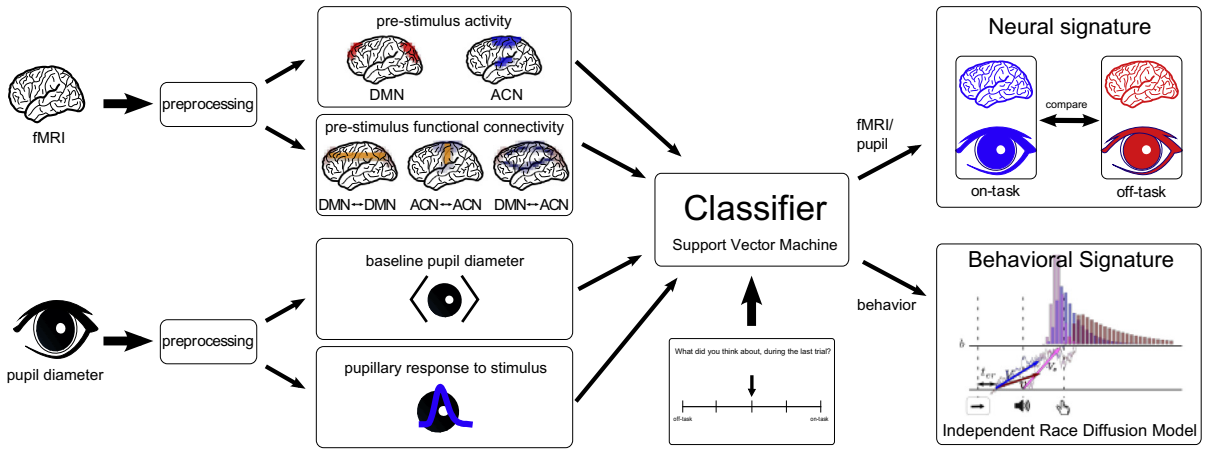

Fig. 2. Summary of Mittner et al.'s (2014) analysis procedure. fMRI and pupil diameter recordings were preprocessed and theoretically derived features were extracted and fed into a machine learning classifier. Self-reported mind wandering scores were used as labels and the classifier was trained to predict them. After training, all trials were classified and the neural and behavioral signature of on-task and off-task behavior was analyzed. Reproduced with permission from Mittner et al. (2014).

classification provides access to a greater range of data from which to understand the neural and behavioral outcomes of mind wandering. Secondly, although it does not explicitly model the temporal structure of the task like HMMs, temporal relationships are implicit in the (typically autocorrelated) neural signal used to classify trials. Since there are neural signatures that reliably predict mind wandering, and mind wandering occurs in temporally structured phases, then neural activity can be used to implicitly classify temporally related periods of on-task and off-task behavior. Finally, the specification of data-driven models with many regressors predicts the occurrence of mind wandering with greater accuracy than is generally possible with single regressors.

Limitations of mixture model approaches to classifying mind wandering

Although the mixture modeling approaches outlined here are appealing, they are not without their drawbacks. Most importantly, mixture models assume a latent mixture of on-task and off-task states and do not consider the possibility that mind wandering exists along a continuum that drifts between periods of greater task engagement through to task-unrelated thoughts. At a theoretical level, this may or may not be consistent with one's views on the occurrence of mind wandering. Mixture models also raise practical issues with parameter estimation. Cognitive models tend to have high dimensionality (i.e., require estimation of multiple parameters from data), such as the drift rate, start point, boundary separation, and non-decision time parameters of the diffusion model. In general, as model dimensionality increases there is a requirement for larger and more informative data sets to ensure reliable parameter estimation. The problem is even greater in mixture models because they aim to infer discrete data-generating sources, which necessitates estimation of more model parameters (i.e., separate model parameters for each discrete state).

There can also be problems with estimating Bayesian latent mixture models and HMMs from behavioral data alone. It is probable that the data generating states we seek to explore in the study of mind wandering – on-task and off-task performance – do not predict highly dissociable effects on observable behavior; at least not to the extent that they predict qualitatively different response outcomes in data. Unlike the example problem with the height of men and women, which are well separated on average, it is challenging to unambiguously classify decision and response time data to a single discrete data-generating source when the discrete states differ only in model parameterization (and not model structure). For example, Vandekerckhove et al.'s (2008) delayed startup model classified only .6% of a representative participant's trials as contaminants (approximately 43 trials from a total of 8000 trials). The mind wandering literature, however, suggests that task-unrelated thoughts are far more common, occupying anywhere between 30% and 50% of our time in everyday life (Killingsworth and Gilbert, 2010). Although Vandekerckhove et al. (2008) were not proposing a model of mind wandering, adoption of their model as a candidate account of mind wandering requires one to accept that (1) mind wandering is far less common than has been reported, (2) the behavioral signature of mind wandering is not recovered by Vandekerckhove et al.'s (2008) model, or, more likely, (3) behavioral data alone have limited ability to constrain mixture parameters in cognitive models of mind wandering (i.e., there is limited diversity in behavioral data to identify mind wandering and hence inform the parameters of models with multiple processes). Supporting the latter proposal does not cast doubt on the latent mixture modeling approach in general. Rather, it suggests that accurately identifying mixture parameters and transition probabilities between latent classes requires richer sources of data to inform parameter estimation, such as behavioral data complemented by simultaneous neural recordings during task performance.

However, even incorporating neural measurements to the classification problem is not without its challenges.

For example, one potential problem with the machine learning approaches is that it can be difficult to identify the effect size of the unique contribution of specific neural signals when using non-linear classifiers. The final approach we discuss overcomes some problems of the mixture model approaches by incorporating explicit neural measures into a hypothesis-driven, regression-based framework of mind wandering.

REGRESSION FRAMEWORKS FOR DYNAMICALLY TRACKING TRANSITIONS BETWEEN TASK-RELATED AND TASK-UNRELATED COGNITIVE STATES

In this section we discuss flexible regression-based modeling approaches. Regression approaches are part of a more general class known as generative models that can allow, for example, simultaneous modeling of multiple streams of data, such as behavior and neural responses, and how those streams may interact to generate the observed data. Although we do not discuss such detailed possibilities here, this flexibility and ability to simultaneously model multiple streams of data opens exciting possibilities for future research.

Using neural data as single-trial regressors on the parameters of cognitive process models

Here, we restrict our focus to a particular regression approach that specifies covariates on the parameters of cognitive models in the form of single-trial regressors. The values of the regressors are derived from each trial of an experiment and could involve stimulus-related properties such as brightness (Vandekerckhove et al., 2011) or item similarity (Hawkins et al., in press-b), or neural measures such as single-trial fMRI and/or EEG activity (e.g., Cavanagh et al., 2011; Borst and Anderson, 2015; Frank et al., 2015; Nunez et al., 2015; Turner et al., 2015). The single-trial measures of stimulus-related or neural activity are then regressed onto structured trial-by-trial variation in the parameters of cognitive process models. This approach is powerful because it defines functional roles of stimulus properties or neural activity as *causes,* not *correlates*, of observed behavior, via their hypothesized influence on parameters of cognitive models. Regression approaches also overcome the restrictive assumption of the mixture models that assume participants are in mutually exclusive on-task or off-task states. Finally, regression approaches maintain the benefits of the machine learning and HMM approach because they implicitly model temporal correlations in task performance, to the extent that temporal information is present in regressors such as neural activity.

We argue that regression approaches that use single-trial regressors are an excellent example of the explanatory power that can be obtained when operating within a model-based cognitive neuroscience framework. Analyzing neural and behavioral data in a single framework provides greater insight into both streams of data than is possible by considering either stream in isolation. To our knowledge, there have been no attempts to model the neural and cognitive processes underlying mind wandering in regression frameworks with single-trial regressors. We first provide an example of the framework in a related domain followed by a hypothesis for the study of mind wandering.

Frank et al. (2015) simultaneously recorded fMRI and EEG activity while participants completed a reinforcement-learning task, and then regressed single-trial neural activity onto parameters of the diffusion model. They tested the hypothesis that mediofrontal theta band activity in the EEG signal and the BOLD response in the fMRI signal of the subthalamic nucleus (STN) and pre-supplementary motor area modulated the response boundary, by estimating linear regression coefficients for the effect of the neural measures on the value of the model parameter. A positive regression coefficient for STN activity, for example, indicates that trials with increased STN activity lead to greater boundary separation, and vice versa.

Although there has been no attempt to apply single-trial regression to the study of mind wandering, we see two main advantages that may follow from such applications: enhanced understanding of the neural and behavioral consequences of mind wandering, and potential for more reliable measures of mind wandering. To continue the illustrative example from the previous section, one could test hypotheses about a particular neural measure (such as pre-stimulus alpha power) and its relation to task performance as indexed by a model parameter (e.g., drift rate) in mind wandering. Specifically, one could calculate a normalized measure of pre-stimulus alpha power at electrodes over task-relevant regions for each trial for use as a single-trial regressor on the drift rate parameter. In this way, alpha power is hypothesized to modulate drift rate at the individual trial-level: a positive regression coefficient for a particular subject indicates that an increase in that subject's pre-stimulus alpha activity causes a corresponding increase in the drift rate on their subsequent trial, and the reverse interpretation for a negative regression coefficient. The size of the regression coefficient relating pre-stimulus alpha power to drift rate gives a measure of the effect size of alpha power on the efficiency of information processing, separately for each subject. A hierarchical modeling approach, which simultaneously models subject-and group-level parameters, provides a principled approach to aggregate the value of the regression coefficient over subjects, allowing for clear hypothesis tests (e.g., is the regression coefficient different to zero). The single-trial regression approach is both flexible and powerful, because it allows precise hypothesis tests of the effect of any neural measure of interest and its relation to any parameter of a cognitive model. Furthermore, the approach is not restricted to single-trial dynamics. For example, it has been shown that decreased deactivation of the default mode network can precede an erroneous response up to 30 s before the error occurs (e.g., Eichele et al., 2008). Such hypotheses about long-range neural

dynamics can be implemented as regressors in an analogous manner to single-trial regressors.

We see one final promise of single-trial regression for the study of mind wandering: the ability to conceptualize thought probes as an outcome measure, not the identifier, of mind wandering. The regression approach described here allows one to reverse the direction of inference compared to other methods of analysis, such as the machine learning approaches. There, responses to thought probes are a pre-requisite for the analysis as they serve as the labels to train a classifier. Here, however, thought probes are not required in the model fitting procedure at all. Rather, a model is implemented with a single-trial neural regressor thought to drive on-task performance, such as pre-stimulus alpha power. The value of the neural measure on thought probe trials, combined with its regression coefficient, can be used to generate predictions for the observed probe responses – on-task or off-task. The extent to which the model predicts responses to thought probes indicates the extent to which the neural measure reliably indexes mind wandering. In this way, we move toward developing models that predict task-related and task-unrelated behavior and the frequency of task-unrelated thoughts.

GENERAL DISCUSSION

The past decade has seen much progress in our understanding of the behavioral and neural consequences of mind wandering. Mind wandering also has great lay interest and important ramifications in everyday functioning. Nevertheless, the abstract nature of mind wandering and task-unrelated thoughts make them difficult to reliably measure, and consequently difficult to develop a complete theory.

In this perspectives piece we have argued that cognitive process models have the potential to illuminate the mechanisms and processes that underlie the task-unrelated thoughts that occur when people lose focus from a primary task. The approaches we outlined are applicable not only to commonly used experimental paradigms in the mind wandering literature, such as the SART, but they generalize to a range of other paradigms. These include tasks used to measure response inhibition, thought to be impaired during mind wandering, such as the stop-signal and Erikson flanker tasks, which have been modeled with sequential sampling models (e.g., White et al., 2012; Logan et al., 2014). The use of a model-based framework also frees one from the use of highly simplified tasks that are used to induce boredom, and, by extension, mind wandering. For example, one can use the frameworks we discussed in paradigms that are less commonly used to study mind wandering, such as standard perceptual decision-making tasks that typically involve experimental manipulation of factors such as choice difficulty and the relative emphasis placed on the speed or accuracy of responding. Finally, although we have only outlined relatively simple laboratory-based tasks, there is no in-principle reason why the models and methods we have described cannot be used to account for performance and mind wandering in more complex tasks (e.g., Eidels et al., 2010). We believe that considering a broader range of experimental paradigms within a model-based framework will lead to more general conclusions about the influence of mind wandering on completion of everyday tasks.

The model-based frameworks we outlined provide different perspectives from which to consider the study of mind wandering. Each approach has advantages and disadvantages and the method one adopts ought to be guided by the research question of interest. For example, the mixture model approaches assume mutually exclusive data generating states – such as an on-task state and an off-task state. In contrast, the single-trial regression approach allows one to explore mind wandering along a continuous dimension that may gradually transition between on-task and off-task poles. Regardless of the chosen framework, the routine incorporation of cognitive models to the study of mind wandering will lead to a deeper understanding of the mechanisms underlying task-focused and task-unrelated thoughts and their relationship with neural activity and behavioral performance. The mixture models that incorporate neural data and the single-trial regression frameworks both allow one to determine which neural measures – connectivity measures, pupil dilation, increased blood flow in specific cortical or subcortical regions, oscillatory activity, and so on – are the strongest predictors of mind wandering. In this sense, one could develop a range of models that differ in whether they assume discrete on-task and off-task states or a single-trial regression approach, and the set of neural predictors employed. Once applied, we can use model selection techniques to determine which model provides the most parsimonious account of the behavioral data.

In addition to enhancing our understanding of mind wandering, the predictive models that follow from the model-based frameworks outlined here can aid identification of contaminants in experimental tasks. One way to approach the contaminant problem could be through the development of 'automated' model fitting routines. These routines would be supplied with neural and behavioral data to produce output of trials classified as more likely to be on task or off task, and in turn the total percentage of time spent mind wandering in an experiment. Such analyses can only follow from a model-based approach because whether a trial is likely to have been on task or off task is assessed relative to the model assumed for the data.

Using model-based methods to study mind wandering in a broader context

In this article we presented perspectives on the use of methods from model-based cognitive neuroscience to understand mind wandering in simple tasks in normal populations. Once these methods have been developed and validated, there is no reason they cannot be extended to research questions of interest to real-life contexts. One application may be in further understanding ''goal neglect", where a person can

understand and describe the requirements of a task but fails to act on those requirements (Duncan et al., 1996). Goal neglect is well known to occur at higher rates in patients with frontal lobe damage (e.g., Luria, 1966), but has also been observed in normal adult populations – where it is correlated with working memory capacity (Kane and Engle, 2003), fluid intelligence and instructional complexity of the task (Duncan et al., 2008, 2012; Bhandari and Duncan, 2014) – and in normally developing children (Roberts and Anderson, 2014). Goal neglect shares similarities to mind wandering – for example, it could be conceived as "zoning out", or a particularly profound attentional lapse – and it is feasible that the model-based methods presented here could be used to further understand goal neglect. For example, it is known that people with lower working memory capacity tend to mind wander more often during attention demanding tasks (e.g., Kane et al., 2007; McVay and Kane, 2009; Kane and McVay, 2012). Resource models of mind wandering propose that the frequency of task-unrelated thoughts increases when there are fewer resources available for the primary task. It is, therefore, possible that people with lower working memory capacity have fewer executive resources to apply to goal-directed tasks. Such hypotheses could be tested within an existing cognitive modeling framework – similar to Heathcote et al.'s (2015b) study of resource theories in prospective memory – or in newly developed quantitative models that formalize the role of 'resources' in task performance. Initial work suggests that lower working memory capacity and higher rates of mind wandering are related to greater variability in drift rate throughout the experiment, which matches observed extreme response times and higher error rates in mind wandering (McVay and Kane, 2012a). Given such developments, one could use cognitive modeling to understand differences in mind wandering, and possibly goal neglect, in normal and patient populations.

Mind wandering is also intimately linked with psychopathology. The most prominent example is attention deficit hyperactivity disorder (ADHD), where clinical diagnoses are related to greater rates of spontaneous mind wandering (Shaw and Giambra, 1993; Seli et al., 2015b). However, a similar relationship is observed in non-clinical samples between greater ADHD symptomatology and spontaneous but not deliberate mind wandering (Seli et al., 2015b), which also transfers to real-life contexts (Franklin et al., in press). Mind wandering has also been implicated in a range of other psychopathologies including depression and depressive symptomatology (Watts and Sharrock, 1985; Deng et al., 2014), dysphoria (Smallwood et al., 2007) and negative mood (Smallwood et al., 2009), anxiety in social contexts (Mrazek et al., 2011), and schizophrenia (Shin et al., 2015). In a separate line of research, these clinical disorders, and others, have been examined through the lens of sequential sampling models, which have led to detailed understanding of empirical phenomena in clinical domains (e.g., Heathcote et al., in press; for an overview, see White et al., 2010). We believe that the independent study of psychopathology and mind wandering, and psychopathology and cognitive modeling, can be integrated within the model-based neuroscience frameworks we have proposed in this article.

Finally, model-based approaches that use neural measures to predict mind wandering have potential for practical applications, depending on the ease of acquiring the neural measure. As a simplified example, we might find that a relatively easily acquired measure – such as pupil diameter – is predictive of reductions in the efficiency of information processing (i.e., lower drift rate) during mind wandering (cf. Mittner et al., 2014). This finding could be incorporated in workplaces where lapses of attention can have large consequences, such as air traffic control (cf. Casner and Schooler, 2015). While the user is completing their ongoing task, pupil diameter could be monitored online. When the system observes patterns of pupil dilation known to be predictive of mind wandering (such as decreased pupillary response to stimuli) it could alert the operator to waning attention, potentially before the off-task state has reduced the efficiency of information processing and its subsequent effect on behavior.

We have provided only a brief overview of a few stylized examples of the theoretical and applied benefits that may follow from adopting a model-based cognitive neuroscience of mind wandering. We believe that routine investigation of mind wandering that combines neural and behavioral data with cognitive process models will continue to grow as a topic of study in its own right, and lead to a more complete understanding of task-related and task-unrelated thoughts on brain and behavior.

Acknowledgments—This research was supported by a Netherlands Organisation for Scientific Research (NWO) Vidi grant to Birte Forstmann and an Australian Research Council (ARC) Professorial Fellowship to Andrew Heathcote.

REFERENCES

Andrews-Hanna JR, Reidler JS, Huang C, Buckner RL (2010) Evidence for the default network's role in spontaneous cognition. J Neurophysiol 104:322–335.

Bartlema A, Lee M, Wetzels R, Vanpaemel W (2014) A Bayesian hierarchical mixture approach to individual differences: case studies in selective attention and representation in category learning. J Math Psychol 59:132–150.

Bastian M, Sackur J (2013) Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. Front Psychol. http://dx.doi.org/10.3389/fpsyg.2013.00573.

Bhandari A, Duncan J (2014) Goal neglect and knowledge chunking in the construction of novel behavior. Cognition 130:11–30.

Bode S, Sewell DK, Lilburn S, Forte JD, Smith PL, Stahl J (2012) Predicting perceptual decision biases from early brain activity. J Neurosci 32:12488–12498.

Bompas A, Sumner P, Muthumumaraswamy SD, Singh KD, Gilchrist ID (2015) The contribution of pre-stimulus neural oscillatory activity to spontaneous response time variability. Neuroimage 107:34–45.

Borst JP, Anderson JR (2015) The discovery of processing stages: analyzing EEG data with hidden semi-Markov models. Neuroimage 108:60–73.

Braboszcz C, Delorme A (2011) Lost in thoughts: neural markers of low alertness during mind wandering. Neuroimage 54:3040–3047.

Brown SD, Heathcote A (2008) The simplest complete model of choice reaction time: linear ballistic accumulation. Cogn Psychol 57:153–178.
Busemeyer JR, Townsend JT (1993) Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol Rev 100:432–459.
Casner SM, Schooler JW (2015) Vigilance impossible: diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. Conscious Cogn 35:33–41.
Cavanagh JF, Wiecki TV, Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ (2011) Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat Neurosci 14:1462–1467.
Cheyne JA, Solman GJ, Carriere JS, Smilek D (2009) Anatomy of an error: a bidirectional state model of task engagement/disengagement and attention-related errors. Cognition 111:98–113.
Christoff K, Gordon AM, Smallwood J, Smith R, Schooler JW (2009) Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proc Natl Acad Sci U S A 106:8719–8724.
Cooper NR, Croft RJ, Dominey SJ, Burgess AP, Gruzelier JH (2003) Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. Int J Psychophysiol 47:65–74.
Deng Y-Q, Li S, Tang Y-Y (2014) The relationship between wandering mind, depression and mindfulness. Mindfulness 5:124–128.
Donkin C, Averell L, Brown SD, Heathcote A (2009) Getting more from accuracy and response time data: methods for fitting the linear ballistic accumulator model. Behav Res Methods 41:1095–1110.
Duncan J, Emslie H, Williams P, Johnson R, Freer C (1996) Intelligence and the frontal lobe: the organization of goal-directed behavior. Cogn Psychol 30:257–303.
Duncan J, Parr A, Woolgar A, Thompson R, Bright P, Cox S, Bishop S, Nimmo-Smith I (2008) Goal neglect and Spearman's g: competing parts of a complex task. J Exp Psychol Gen 137:131–148.
Duncan J, Schramm M, Thompson R, Dumontheil I (2012) Task rules, working memory, and fluid intelligence. Psychon Bull Rev 19:864–870.
Eichele T, Debener S, Calhoun VD, Specht K, Engel AK, Hugdahl K, Von Cramen D, Ullsperger M (2008) Prediction of human errors by maladaptive changes in event-related brain networks. Proc Natl Acad Sci U S A 105:6173–6178.
Eidels A, Donkin C, Brown SD, Heathcote A (2010) Converging measures of workload capacity. Psychon Bull Rev 17:763–771.
Ergenoglu T, Demiralp T, Bayraktaroglu Z, Ergen M, Beydagi H, Uresin Y (2004) Alpha rhythm of the EEG modulates visual detection performance in humans. Cogn Brain Res 20:376–383.
Forstmann BU, Wagenmakers E-J, editors. An introduction to model-based cognitive neuroscience. Springer.
Forstmann BU, Dutilh G, Brown S, Neumann J, Von Cramon DY, Ridderinkhof KR, Wagenmakers E-J (2008) Striatum and pre-SMA facilitate decision-making under time pressure. Proc Natl Acad Sci U S A 105:17538–17542.
Forstmann BU, Ratcliff R, Wagenmakers E-J (in press) Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. Annu Rev Psychol. http://dx.doi.org/10.1146/annurev-psych-122414-033645.
Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, Badre D (2015) FMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J Neurosci 35:485–494.
Franklin MS, Mrazek MD, Anderson CL, Johnston C, Smallwood J, Kingstone A, Schooler JW (in press) Tracking distraction: the relationship between mind-wandering, meta-awareness and ADHD symptomatology. J Atten Disord. http://dx.doi.org/10.1177/1087054714543494.
Giambra LM (1995) A laboratory method for investigating influences on switching attention to task-unrelated imagery and thought. Conscious Cogn 4:1–21.
Gomez P, Ratcliff R, Perea M (2007) A model of the go/no-go task. J Exp Psychol Gen 136:389–413.
Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.
Hawkins GE, Hayes BK, Donkin C, Pasqualino M, Newell BR (in press-a) A Bayesian latent-mixture model analysis shows that informative samples reduce base-rate neglect. Decision. http://dx.doi.org/10.1037/dec0000024.
Hawkins GE, Hayes BK, Heit E (in press-b) A dynamic model of reasoning and memory. J Exp Psychol Gen.
Heathcote A, Brown SD, Wagenmakers E-J (2015a) An introduction to good practices in cognitive modeling. In: Forstmann BU, Wagenmakers E-J, editors. An introduction to model-based cognitive neuroscience. New York: Springer. p. 25–48.
Heathcote A, Loft S, Remington RW (2015b) Slow down and remember to remember! A delay theory of prospective memory costs. Psychol Rev 122:367–410.
Heathcote A, Suraev A, Curley S, Love J, Michie P (in press) Decision processes and the slowing of simple choices in schizophrenia. J Abnorm Psychol.
Kane MJ, Engle RW (2003) Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. J Exp Psychol Gen 132:47–70.
Kane MJ, McVay JC (2012) What mind wandering reveals about executive-control abilities and failures. Curr Dir Psychol Sci 21:348–354.
Kane MJ, Brown LH, McVay JC, Silvia PJ, Myin-Germeys I, Kwapil TR (2007) For whom the mind wanders, and when: an experience-sampling study of working memory and executive control in daily life. Psychol Sci 18:614–621.
Killingsworth MA, Gilbert DT (2010) A wandering mind is an unhappy mind. Science 330:932.
Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53:63–88.
Lee MD, Wagenmakers E-J (2013) Bayesian cognitive modeling: a practical course. New York: Cambridge University Press.
Lewandowsky S, Farrell S (2011) Computational modeling in cognition. California: Sage.
Loft S, Remington RW (2013) Brief delays in responding reduce focality effects in event-based prospective memory. Q J Exp Psychol 66:1432–1447.
Logan GD, Van Zandt T, Verbruggen F, Wagenmakers E-J (2014) On the ability to inhibit thought and action: general and special theories of an act of control. Psychol Rev 121:66–95.
Luria AR (1966) Higher cortical functions in man. London: Tavistock.
Macdonald JS, Mathan S, Yeung N (2011) Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. Front Psychol. http://dx.doi.org/10.3389/fpsyg.2011.00082.
Mason MF, Norton MI, Van Horn JD, Wegner DM, Grafton ST, Macrae CN (2007) Wandering minds: the default network and stimulus-independent thought. Science 315:393–395.
Mathewson KE, Gratton G, Fabiani M, Beck DM, Ro T (2009) To see or not to see: pre-stimulus α phase predicts visual awareness. J Neurosci 29:2725–2732.
Matzke D, Wagenmakers E-J (2009) Psychological interpretation of the ex-Gaussian and shifted Wald parameters: a diffusion model analysis. Psychon Bull Rev 16:798–817.
McVay JC, Kane MJ (2009) Conducting the train of thought: working memory capacity, goal neglect, and mind wandering in an executive-control task. J Exp Psychol Learn Mem Cogn 35:196–204.
McVay JC, Kane MJ (2010) Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). Psychol Bull 136:188–207.
McVay JC, Kane MJ (2012a) Drifting from slow to "D'oh!": working memory capacity and mind wandering predict extreme reaction

times and executive control errors. J Exp Psychol Learn Mem Cogn 38:525–549.
McVay JC, Kane MJ (2012b) Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. J Exp Psychol Gen 141:302–320.
Mittner M, Boekel W, Tucker A, Turner BM, Heathcote A, Forstmann BU (2014) When the brain takes a break: a model-based analysis of mind wandering. J Neurosci 34:16286–16295.
Mrazek MD, Chin JM, Schmader T, Hartson KA, Smallwood J, Schooler JW (2011) Threatened to distraction: mind-wandering as a consequence of stereotype threat. J Exp Soc Psychol 47:1243–1248.
Mrazek MD, Smallwood J, Franklin MS, Chin JM, Baird B, Schooler JW (2012) The role of mind-wandering in measurements of general aptitude. J Exp Psychol Gen 141:788–798.
Nunez MD, Srinivasan R, Vandekerckhove J (2015) Individual differences in attention influence perceptual decision making. Front Psychol. http://dx.doi.org/10.3389/fpsyg.2015.00018.
O'Connell RG, Dockree PM, Robertson IH, Bellgrove MA, Foxe JJ, Kelly SP (2009) Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. J Neurosci 29:8604–8611.
Pachella RG (1974) The interpretation of reaction time in information-processing research. In: Kantowitz BH, editor. Hum Inf Process: Tutor Perform Cogn. Hillsdale, NJ: Lawrence Erlbaum. p. 41–82.
Rae B, Heathcote A, Donkin C, Averell L, Brown S (2014) The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. J Exp Psychol Learn Mem Cogn 40:1226–1243.
Ratcliff R (1978) A theory of memory retrieval. Psychol Rev 85:59–108.
Ratcliff R, McKoon G (2008) The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput 20:873–922.
Ratcliff R, Rouder JN (1998) Modeling response times for two-choice decisions. Psychol Sci 9:347–356.
Ratcliff R, Tuerlinckx F (2002) Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. Psychon Bull Rev 9:438–481.
Ratcliff R, Van Dongen HPA (2011) Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. Proc Natl Acad Sci U S A 108:11285–11290.
Ratcliff R, Philiastides MG, Sajda P (2009) Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci U S A 106:6539–6544.
Reed AV (1973) Speed-accuracy trade-off in recognition memory. Science 181:574–576.
Roberts G, Anderson M (2014) Task structure complexity and goal neglect in typically developing children. J Exp Child Psychol 120:59–72.
Robertson IH, Manly T, Andrade J, Baddeley BT, Yiend J (1997) 'Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia 35:747–758.
Romei V, Gross J, Thut G (2010) On the role of pre-stimulus alpha rhythms over occipito-parietal areas in visual input recognition: correlation or causation? J Neurosci 30:8692–8697.
Scheibehenne B, Rieskamp J, Wagenmakers E-J (2013) Testing adaptive toolbox models: a Bayesian hierarchical approach. Psychol Rev 120:39–64.
Schooler JW, Smallwood J, Christoff K, Handy TC, Reichle ED, Sayette MA (2011) Meta-awareness, perceptual decoupling and the wandering mind. Trends Cogn Sci 15:319–326.
Seli P, Cheyne JA, Smilek D (2013) Wandering minds and wavering rhythms: linking mind wandering and behavioral variability. J Exp Psychol Hum Percept Perform 39:1–5.
Seli P, Jonker TR, Cheyne JA, Cortes K, Smilek D (2015a) Can research participants comment authoritatively on the validity of their self-reports of mind wandering and task engagement? J Exp Psychol Hum Percept Perform 41:703–709.
Seli P, Smallwood J, Cheyne JA, Smilek D (2015b) On the relation of mind wandering and ADHD symptomatology. Psychon Bull Rev 22:629–636.
Shaw GA, Giambra L (1993) Task-unrelated thoughts of college students diagnosed as hyperactive in childhood. Dev Neuropsychol 9:17–30.
Shiffrin RM, Lee MD, Kim W, Wagenmakers E-J (2008) A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. Cogn Sci 32:1248–1284.
Shin DJ, Lee TY, Jung WH, Kim SN, Jang JH, Kwon JS (2015) Away from home: the brain of the wandering mind as a model for schizophrenia. Schizophr Res 165:83–89.
Smallwood J (2010) Why the global availability of mind wandering necessitates resource competition: reply to McVay and Kane (2010). Psychol Bull 136:202–207.
Smallwood J, Schooler JW (2006) The restless mind. Psychol Bull 132:946–958.
Smallwood J, Schooler JW (2015) The science of mind wandering: empirically navigating the stream of consciousness. Annu Rev Psychol 66:487–518.
Smallwood J, Davies JB, Heim D, Finnigan F, Sudberry M, O'Connor R, Obonsawin M (2004) Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. Conscious Cogn 13:657–690.
Smallwood J, O'Connor RC, Sudbery MV, Obonsawin M (2007) Mind-wandering and dysphoria. Cogn Emot 21:816–842.
Smallwood J, Fitzgerald A, Miles LK, Phillips LH (2009) Shifting moods, wandering minds: negative moods lead the mind to wander. Emotion 9:271–276.
Smallwood J, Tipper C, Brown K, Baird B, Engen H, Michaels JR, Grafton S, Schooler JW (2013) Escaping the here and now: evidence for a role of the default mode network in perceptually decoupled thought. Neuroimage 69:120–125.
Smilek D, Carriere JS, Cheyne JA (2010) Failures of sustained attention in life, lab, and brain: ecological validity of the SART. Neuropsychologia 48:2564–2570.
Smith RE (2003) The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. J Exp Psychol Learn Mem Cogn 29:347–361.
Smith PL, Ratcliff R (2004) The psychology and neurobiology of simple decisions. Trends Neurosci 27:161–168.
Stawarczyk D, Majerus S, Maj M, Van der Linden M, D'Argembeau A (2011a) Mind-wandering: phenomenology and function as assessed with a novel experience sampling method. Acta Psychol 136:370–381.
Stawarczyk D, Majerus S, Maquet P, D'Argembeau A (2011b) Neural correlates of ongoing conscious experience: both task-unrelatedness and stimulus-independence are related to default network activity. PLoS One 6. http://dx.doi.org/10.1371/journal.pone.0016997.
Steyvers M, Lee MD, Wagenmakers E-J (2009) A Bayesian analysis of human decision-making on bandit problems. J Math Psychol 53:168–179.
Turner BM, Van Maanen L, Forstmann BU (2015) Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. Psychol Rev 122:312–336.
Usher M, McClelland JL (2001) On the time course of perceptual choice: the leaky competing accumulator model. Psychol Rev 108:550–592.
Van Dijk H, Schoffelen J-M, Oostenveld R, Jensen O (2008) Pre-stimulus oscillatory activity in the alpha band predicts visual discrimination ability. J Neurosci 28:1816–1823.
Vandekerckhove J, Tuerlinckx F (2007) Fitting the Ratcliff diffusion model to experimental data. Psychon Bull Rev 14:1011–1026.
Vandekerckhove J, Tuerlinckx F (2008) Diffusion model analysis with MATLAB: a DMAT primer. Behav Res Methods 40:61–72.
Vandekerckhove J, Tuerlinckx F, Lee MD (2008) A Bayesian approach to diffusion process models of decision-making. In: Sloutsky VM, Love BC, McRae K, editors. Proc 30th Annu Conf Cogn Sci Soc. Austin, TX: Cogn Sci Soc. p. 1429–1434.

Vandekerckhove J, Tuerlinckx F, Lee MD (2011) Hierarchical diffusion models for two-choice response times. Psychol Methods 16:44–62.
Voss A, Voss J (2007) Fast-dm: a free program for efficient diffusion model analysis. Behav Res Methods 39:767–775.
Voss A, Rothermund K, Voss J (2004) Interpreting the parameters of the diffusion model: an empirical validation. Mem Cogn 32:1206–1220.
Voss A, Nagler M, Voss V (2013) Diffusion models in experimental psychology: a practical introduction. Exp Psychol 60(385):402.
Watts FN, Sharrock R (1985) Description and measurement of concentration problems in depressed patients. Psychol Med 15:317–326.
Weissman DH, Roberts KC, Visscher KM, Woldorff MG (2006) The neural bases of momentary lapses in attention. Nat Neurosci 9:971–978.
White CN, Ratcliff R, Vasey MW, McKoon G (2010) Using diffusion models to understand clinical disorders. J Math Psychol 54:39–52.
White CN, Brown S, Ratcliff R (2012) A test of Bayesian observer models of processing in the Erikson flanker task. J Exp Psychol Hum Percept Perform 38:489–497.
Wickelgren WA (1977) Speed-accuracy tradeoff and information processing dynamics. Acta Psychol 41:67–85.
Wilson TD, Reinhard DA, Westgate EC, Gilbert DT, Ellerbeck N, Hahn C, Brown CL, Shaked A (2014) Just think: the challenges of the disengaged mind. Science 345:75–77.

(Accepted 16 September 2015)
(Available online 28 September 2015)