Mind-wandering may—depending on circumstances—be either adaptive or non-adaptive. While mind-wandering during a task that requires continuous cognitive control may be problematic, mind-wandering during a task that does not require continuous attention may in fact contribute to improved problem-solving and creativity, since it frequently involves prospective memory (Baird et al., 2012). At present, there are several theories of mind-wandering that have emphasized different aspects of this process. The executive failure theory states that mind-wandering occurs out of a failure to focus attention on relevant information (McVay & Kane, 2009), while the perceptual decoupling theory states that mind-wandering is primarily a process of decoupling from the external environment, such that it can be devoted to internal processes (Smallwood, Beach, Schooler, & Handy, 2008; Smallwood et al., 2011). Evidence for perceptual decoupling comes from studies that have found that the amplitude of evoked potentials is reduced during states of task-unrelated thought (Smallwood et al., 2008). In addition, the pupil responds less to presented stimuli during states of mind-wandering (Smallwood et al., 2011). Instead of processing perceptual stimuli, the brain appears to be engaged in episodic processing during the periods of distraction (Andrews-Hanna, Smallwood, & Spreng, 2014). Executive failure theory is based on studies that relate cognitive control abilities to the ability to resist mind-wandering (Kane & McVay, 2012). Alternatively, it has been suggested that mind-wandering results from failures in meta-cognition, the ability to observe one’s thoughts (Fox & Christoff, 2014).

None of the above-mentioned theories has been formalized in computational models. Closest related to studying mind-wandering come models of distraction and fatigue. For example, Gunzelmann, Gross, Gluck, and Dinges (2009) investigated the effects of fatigue on performance on a monotonous psychomotor vigilance task. According to his model, fatigue impacted a parameter used to compute the utility of particular task strategies (this parameter has been associated with motivation). This parameter change made random key presses more likely as the participant became more tired. In addition, they modelled lapses in behavior by having productions that had a sufficiently low activation that they would only be performed as a result of random fluctuations in their activation parameter. Note how their model does not model distraction through mind-wandering as an explicit process, but instead assumes that the cognitive system is not functioning during distraction. Similarly, Gonzalez, Best, Healy, Kole,
and Bourne Jr. (2011) modelled the effects of fatigue on a data entry task as a reduction in motivation in combination with a reduction in attentional control. This attentional control parameter affects the activation of different pieces of information, and the larger this parameter is, the better these pieces of information can be distinguished.

A previous ACT-R model of a sustained attention task that is often used in mind-wandering studies (Peebles & Bothell, 2010) focused primarily on explaining response times decrease just preceding an attentional lapse (as reflected in an error). They produced this phenomenon by a competition between two response strategies: one strategy is responding whenever a stimulus is detected, which is very fast, while an alternative strategy first checks the stimulus before responding. When the fast strategy fails then ACT-R will switch to the most costly slow strategy. Note that this model does not implement an explicit cognitive mechanism for what happens during distraction. Here we intend to build on that previous model by implementing a competition between a “distracted” and an “attentive” model, where the distracted model makes the mind-wandering process explicit.

**Model**

Our model of distraction (Figure 1) consists primarily of a competition between a sub-model for paying attention to the task and a sub-model for mind-wandering. The model was implemented in the Adaptive Control of Thought-Rational (ACT-R) cognitive architecture (Anderson, 2007). Tasks are implemented in this cognitive architecture by specifying a set of if-then statements (production rules) that describe how different cognitive resources interact. Two ACT-R mechanisms are of crucial importance for our model. First, ACT-R has a memory store, where the activation of each memory chunk determines its use and its retrieval time. The activation in turn is determined by how often a chunk is retrieved, its activation at baseline, and how much activation spreads from other, related memory chunks. The second mechanism that determines what happens in the model at a particular moment is the utility associated with each production rule. When production rules help to generate rewards, their utility goes up, leading them to be used more frequently. However, given that in mind-wandering there are no external reward processes that guide the process, we will not make use of this second mechanism in our model.

In this application, the model starts out by focusing its attention on the stimulus on the screen. When there is a stimulus, it will process the stimulus and perform the appropriate action. When there is no stimulus, it will continually run a production which checks what the most active goal (“paying attention” or “distraction”) is in declarative memory (“check whether attending” in Figure 1). The activations of the goals in declarative memory are governed by rules from episodic memory decay (Altmann & Gray, 2008). This means that items that are retrieved in activation, but over time the activation decays. At the start of the task, the “paying attention” goal is activated because it has been retrieved from episodic memory. This goal then decays over time, and at some point the “distraction” goal becomes stronger (Figure 1). When the “distracted” goal is retrieved by this checking production, the mind-wandering model commences.

Mind-wandering consists of a continuous retrieval of declarative memories. The retrieval process keeps continuing until at some point a memory that says “remember to attend” is retrieved. At that point, the model returns to paying attention and the whole cycle can start again. There is spreading activation between memories, which ensures that—as in real life—memories that are of the same valence (positive, negative, or neutral) tend to be recalled in sequence (van Vugt, Hitchcock, Shahar, & Britton, 2012).

Our main goal in this paper is to find out whether the hypothesized mind-wandering model can in fact describe empirical mind-wandering data. Studies have experimentally studied mind-wandering by giving participants a very boring task, in which participants are likely to drift off. Here, we will model data from two experiments: Mrazek, Smallwood, and Schooler (2012) (Experiment 1) and Bastian and Sackur (2013) (Experiment 2). Both experiments are variants of the sustained attention to response task (SART), in which participants are requested to press a button as quickly as possible every time a target is presented, but to withhold a button press to a more rarely presented non-target (Cheyne et al., 2009; Smallwood et al., 2004).

When the distraction model is inserted in a model of the SART task, we assume performance is determined by the following mechanisms, building on Peebles and Bothell (2010)’s model. When task stimuli are presented while the model is in paying attention mode, the model will look at the stimuli and retrieve the relevant stimulus-response mapping from episodic memory. Conversely, when the model is distracted, it will not retrieve the stimulus-response mapping from episodic memory but instead respond with the habitual response. However, responding may take a little while, because the model will only be able to respond when it is not busy retrieving a memory in its mind-wandering train. This
potential delay before responding is responsible for creating
the increase in response time variability that is typically ob-
served in mind-wandering studies (Bastian & Sackur, 2013;
Mrazek et al., 2012). In Experiment 2, thought probes may
also be presented. Whenever a thought probe occurs, the
model will press the “on-task” button whenever it is in paying
attention mode, while it will press the “off-task” button when
it is busy retrieving memories from episodic memory during
distraction. The models can be retrieved from http://www.ai.rug.nl/~mkvanvugt/mindwanderingModels.zip. A
flow chart of the model is shown in Figure 1.

Model testing

Experiment 1

We first used our model to simulate the average data pub-
lished by Mrazek et al. (2012). In this experiment, the targets
consisted of the letter “O”, and non-targets consisted of the
letter “Q.” Stimuli were presented for 2 s with an interstimu-
lus interval of 2500 ms. There were in total 240 stimuli; 216
targets and 24 non-targets.

Figure 2 shows that the simulated performance of the
model reproduces both the observed number of SART errors
and the coefficient of variation of the response time. Errors
are produced whenever the model is mind-wandering. The
coefficient of variation results from variability in memory re-
trieval time.

Having established the model can produce behavior simi-
lar to human participants, it becomes possible to examine how
the model produces this behavior. Figure 6 shows that
according to our model, the frequency of distractions shows
a U-shape: initially, there are quite a few distractions, which
reduces in the middle of the task, and increases at the end.

While there is evidence for an increase in the frequency of
distraction towards the end of the task (Bastian & Sackur,
2013), it is not clear whether the distraction at the begin-
ing of the task is plausible. Future studies that have better mea-
sures of the frequency of distraction (e.g., Katidioti et al., sub-
mitted) should clarify this issue.

An important question is how crucial the proposed mech-
anism is for terminating mind-wandering. A simpler mecha-
nism for achieving this goal may be a direct competition be-
tween the “distraction” and “paying attention” goals. In other
words, at any moment during mind-wandering, a production
could fire that reflects the end of the mind-wandering pro-
cess. Figure 3 shows that this alternative mechanism makes
too few mistakes because the episodes of mind-wandering are

---

Figure 1: Model time line. Each box corre-
spends to a production (some less im-
portant productions have been left out).
The model starts on the left top with
retrieving its current goal, correspond-
ing to goal checking. Initially, the “at-
tending” goal has the highest activation
(see dashed blue box), but over time the
attending goal declines in activation to
become similar to the distracted goal.
When this “distracted” goal is retrieved,
the model switches to retrieving memo-
ries from declarative memory, represent-
ning mind-wandering. Mind-wandering
(cyan) continues until “remember to at-
tend” (purple) is retrieved. At that time,
the model goes back to monitoring goals.
When a stimulus is presented (pink line),
then the model identifies it and retrieves
the stimulus-response mapping in case it
is attending. When it is distracted, it fin-
ishes retrieving the current distraction and
then presses the default response.

Figure 3: Behavior produced by a model in which the thought
pump is ended with a production “end-thought-pump” rather
than a specific memory retrieval.
The duration of mind-wandering is determined by the episodic memory retrievals that make up the mind-wandering process. When the pool of to-be-retrieved memories is larger, then distractions will tend to persist longer, because the chance that the distraction-ending memory is retrieved is smaller. A larger number of retrievable memories corresponds to something akin to the number of retrieval cues. In some contexts, people may be able to think of many different things, while in other contexts they can only retrieve a limited number of items. A further determinant of distraction duration is the association structure of the distracting memories. When memories spread activation to the memory that ends the distraction, this will decrease distraction duration; when they spread activation to other memories, this increases distraction duration. These factors could potentially be manipulated to account for individual differences in distractability.

Together, these results show that it is possible to use our model of mind-wandering to simulate performance on a task. However, the results are fairly weak since we only fit two average numbers: the number of errors and response time variability. More data are needed to adequately constrain our cognitive model. We therefore use the complete dataset collected by Bastian and Sackur (2013) to further test the model, which allows us to examine more behavioral measures. An additional advantage of that dataset is that the task was interspersed with thought probes that asked the participant to report on the content of their thoughts. The responses to thought probes are another constraining factor for our model. Moreover, it highlights an important advantage of modeling mind-wandering explicitly, as we did here. When a model has no explicit process description of mind-wandering, it cannot predict responses to thought probes.

**Experiment 2**

In Experiment 2, participants performed a very similar task as in Experiment 1, although the timing was a little bit different. Importantly, we did not change the model parameters at all to predict performance in this task. In this experiment, the non-target consisted of the digit 3, and the target consisted of all other digits. The digits were presented for 500 ms with an interstimulus interval of 1500 ms. There were in total 888 stimuli; 811 targets and 77 non-targets. In addition, 24 thought probes that were randomly interspersed in the task. These thought probes asked a series of four questions about task performance. First, participants were asked “How focused were you on the task? 0: on-task, 1: task-related thought, 2: distraction, 3: mind wandering.” Second, “Did you know that you were in the just-reported mental state or did you only notice it when asked? 0=aware, 1=unaware.” The third question concerned the phenomenology/type of the thoughts, while the fourth question assessed the temporal orientation of the thoughts (past, present, future, or no particular time). In this paper, we will only model the question about whether the participant is on-task.

Figure 4 shows that task performance could be modelled accurately with the model for Experiment 1, although in this
case, the model is performing slightly too well for the participants. Potentially, model fits could be improved by adjusting parameters.

In addition to average responses, it is also important to consider the entire response time distribution (e.g., Ratcliff, 2002). Figure 5 shows that the modeled and observed response time distributions for task performance overlay considerably, although the response time variability predicted by the model is too small.

Finally, our explicit model of mind-wandering allows us to model the responses to questions about the contents of thoughts. At random moments in the task, the participant is asked whether they were on-task or off-task. Figure 7 shows that the model over-estimates the proportion of being on-task relative to human participants, which is consistent with the model’s overperformance evident in Figure 4. Another notable feature visible in Figure 7 is that participants require about 3–4 seconds to formulate their response to the question “Were you on-task?” This response time is much longer than those observed for cognitive tasks, and our model is not able to predict it. Two potential mechanisms that could be involved in generating this time are (1) the conversion of a pre-verbal into a verbal response (Teasdale & Chaskalson, 2011) or (2) mental time travel to several moments before the thought probe appeared to retrieve the memories that occurred at that time (Howard & Kahana, 2002; Tulving, 2002). Future modeling efforts should investigate these ideas.

**Discussion**

We proposed a model that describes mind-wandering mechanistically. We showed how it could account for task performance in two experiments featuring the Sustained Attention to Performance task (without changing model parameters between the two). While previous models only treat distraction abstractly as noise in the cognitive system (VandeKerkhove & Tuerlinckx, 2007) or an absence of cognitive activity (Gunzelmann et al., 2009), we made an explicit model of the mind-wandering process. This allowed us to not only model task performance, but also responses to thought probes. Our model provides a potential implementation of the executive failure theory, where in our case executive failure is implemented as a failure to keep checking what the current goal is. It is also related to perceptual decoupling in that perceived stimuli are not further analyzed, but it places the constraints at a higher level than initial stimulus processing.

In the future, our explicit model of mind-wandering could allow us to examine the effect of different types of mind-wandering on cognitive processing. For example, depressive rumination impairs task performance, and our model can make predictions about exactly how it does so. By describing the thought process from moment to moment, we will be able to investigate how not only cognitive control factors affect task performance, but also the content of thought. For this to be done, it will be important to populate the model with memories that reflect the distribution of memories that is observed in actual participants. Such data can be obtained from studies that report the content of thoughts in thought probes (Bastian & Sackur, 2013).

While our model makes a promising start with modeling task performance, there is still some work to be done. Our model predicts better performance in Experiment 2 than is produced by the participants (Figure 4). In addition, the frequency of mind-wandering episodes (Figure 6) shows a U-shape, rather than the previously reported increase Bastian and Sackur (2013).

The most dramatic discrepancy is in the response times to thought probes, which are much faster in our model than in real participants. A future iteration of our model may need to include a mechanism by which the participant converts the content of thoughts into a verbal report.

At the same time, we have relatively few datapoints and cannot make strong inferences about our model. One possible future direction may be predicting the frequency of distractions. We have recently started to measure those by means of eye movements to an ambient video monitor (Katidioti et al., submitted). Our model could potentially describe how the frequency of distraction changes over time, and depends on different factors such as task difficulty. This is particularly important because it is often thought that introspective judgments are unreliable (Larson,Perlstein, Stigge-Kaufman, Kelly, & Dotson, 2006).

In short, we have developed a mechanistic model of mind-wandering. This model can in the future be used to disentangle different types of mind-wandering. In addition, future experiments should elucidate the neural correlates of distraction and mind-wandering, such that those measures can be used to track distraction online (Bengson, Mangun, & Mazaheri, 2012).

**Acknowledgments**

MvV developed the basic model during a stay as Mind & Life visiting scholar at Amherst College.
References


