A Hierarchical Model of Internal Attentional State over Time

Jaeah Kim (jaeahk@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Shashank Singh (shashank.singh@tuebingen.mpg.de)
Max Planck Institute for Intelligent Systems, Tübingen, Germany

Dan Yurovsky (yurovsky@cmu.edu)
Department of Psychology, Carnegie Mellon University Pittsburgh, PA 15213 USA

Anna V. Fisher (fisher49@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University Pittsburgh, PA 15213 USA

Erik D. Thiessen (thiessen@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University Pittsburgh, PA 15213 USA

Abstract

A key feature of attention is that it moves over time, guided both exogenously by changing external circumstances and endogenously by internal cognitive states. However, the endogenous mechanisms guiding the movement of attention in the absence of external cues remain poorly understood. This paper develops and validates a computational model of how internal attentional states, motivated by the Adaptive Gain Theory of the locus coeruleus-norepinephrine (LC-NE) system, can guide the movement of visual attention over time. By fitting our model to young children's gaze data as they perform a visual object tracking task, we investigate developmental changes in higher-order patterns of attending behavior between 3.5-6 years of age, and we hypothesize about how the LC-NE system might mediate these changes.

Keywords: sustained attention development; locus coeruleus-norepinephrine system; hierarchical hidden Markov model

While navigating life requires the ability to continuously sustain attention on a given locus (e.g., to complete a simple task), it also crucially requires switching attention between loci in response to changing circumstances, new information, or reprioritization of goals. To obtain a more complete picture of how attention operates over time, it is thus necessary to investigate not only how individuals maintain their attention on a particular locus, but also how they guide the movement of their attention between different loci. While extensive work beginning with Posner's cueing paradigm (Posner, Nissen, & Ogden, 1978; Posner, 1980; Posner & Petersen, 1990) has investigated how external cues can trigger the movement of attention, the internal mechanisms determining when and where to move attention, independent of specific external cues, remain poorly understood. Much research has also focused on characterizing sustained attention, the sustained allocation of attention to a single "target" locus (Langner & Eickhoff, 2013), but research on characterizing the movement of attention to, from, and between other loci outside of a target locus has been comparatively limited.

Much attentional behavior may transpire outside of the designated locus that is not captured by traditional experimental paradigms; for example, an individual may balance other goals alongside immediate task performance by engaging in activities such as preparing for future tasks, reflecting on past experiences, or exploring the task environment, even when these directly conflict with optimally performing the prescribed task (Andrews-Hanna, Smallwood, & Spreng, 2014). Although off-task behaviors have been studied qualitatively in mind-wandering research (Smallwood & Schooler, 2006, 2015), the movement of attention between loci has rarely been studied quantitatively.

Kim, Singh, Thiessen, and Fisher (2020) recently developed a quantitative approach to modeling the movement of visual attention between loci in participants performing TrackIt, a visual object tracking task (Fisher, Thiessen, Godwin, Kloos, & Dickerson, 2013). Their approach utilizes a hidden Markov model (HMM), in which hidden states correspond to possible objects of attention (among the objects being displayed by TrackIt) and observations correspond to participants' gaze, which is continuously recorded as they perform the task. Given the gaze data, this HMM allows researchers to infer the object of a participants' attention as it moves over time, and to identify transitions between objects.

An HMM is a natural choice for a simple model of human visual attention; at each time point t, the participant attends to something $S(t)$ (the hidden state), and we observe eye-tracking data that is primarily a function of $S(t)$ and random noise. Because human attention moves slowly relative to the frequency at which eye-tracking data is collected, the state $S(t)$ is strongly related to the preceding and successive states ($S(t-1)$ and $S(t+1)$). Unlike simpler models that consider data at each time point independently (Zelinsky & Neider, 2008), the HMM uses this short-term dependence to mitigate noise and handle complex scenarios such as object collisions (when multiple objects briefly occupy the same space), without sacrificing the fine temporal resolution of eye-tracking data. However, the model of Kim et al. (2020)does not make any substantial assertions about the cognition *underlying* attending behavior over time. In this paper, we develop a computational model, built on top of the HMM of Kim et al. (2020), of how higher-order patterns of attending behavior, which we refer to as *attentional modes*, can guide the movement of visual attention over time. Specifically, we describe a hierarchical extension of the HMM that incorporates longer time dependencies through attentional modes motivated by those outlined in the Adaptive Gain Theory of the Locus Coeruleus-Norepinephrine system (Aston-Jones & Cohen, 2005, described below). By building on the HMM of Kim et al. (2020), we are able to fit our model's parameters using participants' gaze data collected as they perform TrackIt. We then use our model to investigate development of young children's attending behavior.

Adaptive Gain Theory of the Locus Coeruleus-Norepinephrine System

The locus coeruleus (LC) is a small group of about 32,000 noradrenergic neurons in the pons that provides the bulk of norepinephrine (NE) in the brain. The LC-NE system has historically been thought to play a crucial role in attention, initially found to regulate basic arousal on the sleep-wake spectrum (Berridge & Waterhouse, 2003), and later found to also directly influence behavioral performance in a capacity beyond just general arousal regulation, through two different types of LC activity: phasic and tonic. *"Phasic activity"* describes bursts of LC activity that are typically associated with focused perception of task-relevant information, while *"tonic activity"* describes overall background LC activity and is associated with overall levels of arousal.

The Adaptive Gain Theory (AGT) of the LC-NE system, proposed by Aston-Jones and Cohen (2005), describes three modes of attending behavior over time: a high tonic mode, characterized by overall high baseline activity in LC neurons and distracted attending, a phasic mode, characterized by overall lower baseline activity in LC neurons with phasic spikes of activity temporally corresponding to focused attending and task-relevant responding, and a low tonic mode corresponding to low baseline LC activity, no phasic spikes, and behavioral disengagement from the visual environment.

We propose to incorporate these three modes into the HMM using a hierarchical extension of the HMM (a Hierarchical Hidden Markov Model, HHMM). In particular, we propose to add a second hidden layer, illustrated in Fig. 1, with three latent states:

1. A *Distractible* mode (based on the AGT's high tonic mode), in which attention is not selective to the Target and transitions often between objects.
2. An *Optimally Engaged* mode (based on AGT's phasic mode), in which attention remains on Target.
3. A *Disengaged* mode (based on AGT's low tonic mode), in which attention is not allocated to any of the displayed objects.

(Aston-Jones & Cohen, 2005) proposed that these different modes of LC-NE activity play roles in navigating the *exploration-exploitation trade-off* of attention, the competition between attending to task-relevant sources of information and exploring new sources of information. Direct evidence supporting this hypothesis has recently been provided by Dubois, Habicht, et al. (2020), who showed that administering a norepinephrine blocker reduces participants' exploratory search behavior. Further indirect evidence, based on measurement of pupil diameter, which has been shown to be related to LC activity (Mathôt, 2018; Rajkowski, 1993; Gilzenrat, Cohen, Rajkowski, & Aston-Jones, 2003; Alnæs et al., 2014) suggests that both high tonic LC activity (Smallwood et al., 2011; Unsworth & Robison, 2016; Konishi, Brown, Battaglini, & Smallwood, 2017) and low tonic LC activity (Grandchamp, Braboszcz, & Delorme, 2014; Mittner et al., 2014; Unsworth & Robison, 2016; Konishi et al., 2017) are related to reduced processing of task-relevant stimuli and poorer performance on sustained attention tasks, suggesting that optimal engagement lies in balancing these two.

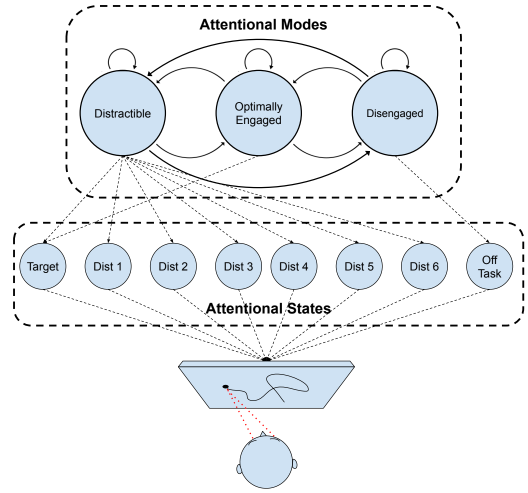

Figure 1: Schematic of the proposed HHMM. The three second-order hidden states corresponding to modes of LC-NE functioning are illustrated in the top layer. Below that are 8 first-order hidden states, corresponding to the 7 TrackIt objects and an "Off-Task" state. The participant's gaze at each time point depends on which of these 8 states they are in.

Development of Sustained Attention

Attending behavior shows marked developments in children between the ages of 3.5 to 6 years (see (Fisher & Kloos, 2016) for a detailed review), the population targeted in our study. Many studies have documented improvements in selectively attending to and sustaining attention on task-relevant information in the presence of distracting task-irrelevant information (Diamond, 2006; Fisher et al., 2013; Ruff & Rothbart, 2001). Recent studies have shown that this increased selectivity of attention has costs in terms of reduced processing of task-irrelevant information (Blanco & Sloutsky, 2020a; Deng & Sloutsky, 2016; Dubois, Aislinn, et al., 2020; Plebanek & Sloutsky, 2017). This can be viewed as a developmental trend along the exploration-exploitation trade-off in the guidance of attention, with younger children exhibiting more exploratory attention to support longer term learning and older children exhibiting increasingly more exploitative attention to support shorter-term performance (Blanco &

Sloutsky, 2020a; Dubois, Aislinn, et al., 2020; Gopnik, 2020; Laureiro-Martínez, Brusoni, & Zollo, 2010; Mehlhorn et al., 2015). As described above, the LC-NE system has been implicated in the mediation of exploration and exploitation, suggesting that it may play a role in this developmental trend. While further work is needed to verify connections between behavior and the LC-NE system, the present study begins to evaluate the plausibility of this hypothesis by testing predictions about behavior across development from our cognitive model motivated by the LC-NE system. Specifically, we hypothesize that, over the course of development between 3.5-6 years of age, children will spend less time in the Distractible mode and more time in the Optimally Engaged mode.

Specific Contributions

In this paper, we investigate two sets of questions regarding the HHMM model. On the modeling side, we first demonstrate practical feasibility and face validity of the HHMM by fitting its parameters to data from real participants performing TrackIt, confirming that the fitted parameters satisfy basic expectations about the model, and evaluating how well the model predicts participants' task performance. We also perform an ablation study to evaluate the relative importance of each mode in our model. On the developmental side, we test the two hypotheses motivated above, namely that, over the course of development between the ages of 3.5-6 years, children will spend (a) less time in the Distractible mode and (b) more time in the Optimally Engaged mode.

Methods

TrackIt

TrackIt, illustrated in Figure 2, is a visual object-tracking task introduced by Fisher et al. (2013) to measure sustained attention in young children. Participants are instructed to track, using only their eyes, a single *Target* object moving about on a grid, among other moving *Distractor* objects. At the end of each trial, all objects vanish from the grid, and participants are asked to identify the grid cell the Target occupied immediately before vanishing. The accuracy of this final response, referred to as *Location Response*, is used as the main behavioral measure of task performance. This measure allows developmentally sensitive assessment of sustained attention over a range of ages, with children as young as 3 years old consistently completing the task and providing usable data (Fisher et al., 2013; Keebler, Kim, Stanley, Thiessen, & Fisher, 2020; Kim, Vande Velde, Thiessen, & Fisher, 2017).

Because TrackIt requires continuous overt attention to the Target, eye-tracking provides information about a participant's visual attention with high temporal resolution. Moreover, TrackIt explicitly provides task-irrelevant objects, alongside the Target, to which the participant can attend, allowing us to distinguish attentional lapses due to distraction by task-irrelevant stimuli from those caused by disengagement from the visual task. These features make TrackIt, together with eye-tracking, well suited to investigating the three attentional modes proposed by the AGT (Aston-Jones & Cohen, 2005), in contrast to other widely-used sustained attention tasks, such as the continuous performance test (CPT), which provide temporally sparse data and only allow for the distinction of on- and off-task behaviors (Fisher & Kloos, 2016; Rosvold, Mirsky, Sarason, Bransome Jr, & Beck, 1956; Riccio, Reynolds, Lowe, & Moore, 2002), motivating our use of continuous gaze data collected from children performing TrackIt to fit and evaluate our HHMM model.

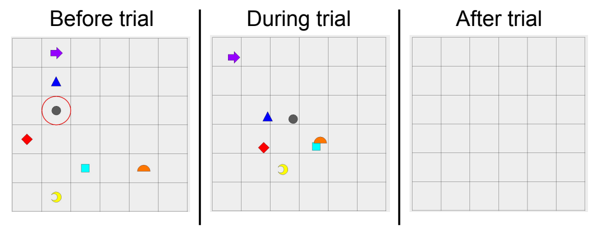

Figure 2: An example TrackIt trial. The Target object, in this case a grey circle, is indicated before the trial by a red circle. During the trial, all objects move in unpredictable piecewise-linear paths and disappear after a random duration. After the trial, the participant is asked to indicate the grid cell the Target occupied before disappearing. A video of an example TrackIt trial provided by Kim et al. (2020) can be found at `https://github.com/CMU-CDL/TrackIt/blob/main/endogenous_TrackIt_example.mp4?raw=true`.

Hierarchical Hidden Markov Model

Our model consists of a two-level hidden Markov chain, with the higher level encoding 3 *modes* (Distractible, Optimally Engaged, or Disengaged) and the lower level encoding 8 attentional *states* (Target, 6 Distractors, or Off-Task). The higher-level Markov chain is parametrized by a time-invariant transition matrix $\Pi_{\text{Modes}} \in [0,1]^{3\times 3}$ between the 3 modes and an initial distribution $\pi_{\text{Modes}} \in [0,1]^3$ on the 3 modes. We impose no explicit assumptions on Π_{Modes} and π_{Modes}, although, because we do not expect transitions between modes to occur too frequently, we will expect the diagonal values of Π_{Modes} to be much larger than the off-diagonal values. The lower-level Markov chain over attentional states is parametrized by a transition matrix that varies over time depending on which of the three modes the participant is in. *Within each mode*, the transition matrices (denoted $\Pi_{\text{Distractible}}$, $\Pi_{\text{Optimally Engaged}}$, and $\Pi_{\text{Disengaged}}$) between attentional states are time-invariant, as described below. Similarly, the initial distribution of attentional states depends on the initial mode. Within the Distractible mode, the initial attentional state distribution is uniform over the 7 objects; within the Optimally Engaged mode, the initial attentional state is always the Target; in the Disengaged mode, the initial attentional state is always Off-Task.

In the Distractible mode, we do not expect the participant to preferentially attend to the Target. Since, the 7 objects are randomly sampled in each trial from the same set of possible objects, we thus expect transitions between the 7 objects to be uniformly likely on average; i.e., all diagonal entries of the matrix are constrained to be identical, and all off-diagonal entries of the matrix are also constrained to be identical:

$$\Pi_{\text{Distractible}} = \begin{bmatrix} c_1 & c_2 & c_2 & \cdots & c_2 & 0 \\ c_2 & c_1 & c_2 & \cdots & c_2 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ c_2 & c_2 & c_2 & \cdots & c_1 & 0 \\ \underbrace{1/7}_{\text{Target}} & \underbrace{1/7 \quad 1/7 \quad \cdots \quad 1/7}_{\text{6 Distractors}} & & & & \underbrace{0}_{\text{Off-Task}} \end{bmatrix} \in [0,1]^{8\times 8}. \tag{1}$$

In the Distractible mode, we assume, as in the original HMM, that the participant's gaze has a Gaussian distribution centered around the center of the object currently being attended. The Gaussian is assumed to have the same covariance for each object. Moreover, given the trajectories of objects in TrackIt are, on average, horizontally and vertically uncorrelated, the horizontal and vertical components of the gaze should also, on average, be uncorrelated, and so we assume the covariance matrix of this Gaussian is diagonal:

$$\Sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}. \tag{2}$$

In the Optimally Engaged mode, we assume the participant attends only to the Target. We acknowledge that this assumption is somewhat strong, and it is possible that participants' gaze occasionally moves towards other objects even when their attention is wholly on the Target (e.g., due to covert attention). However, to keep the number of model parameters small and to clearly distinguish the Distractible and Optimally Engaged modes, in this work, we maintain this assumption, so that the transition matrix is simply

$$\Pi_{\text{Optimally Engaged}} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix} \in [0,1]^{8\times 8}. \tag{3}$$

In the Optimally Engaged mode, as in the Distractible mode, we assume that the participant's gaze has a Gaussian distribution centered around the center of the object being attended, in this case the Target; also, the covariance of the Gaussian distribution is assumed to be identical to that in the Distractible mode (Eq. 2).

Finally, in the Disengaged mode, we assume the participant's attention is independent of the TrackIt objects, giving

$$\Pi_{\text{Disengaged}} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \in [0,1]^{8\times 8}. \tag{4}$$

Furthermore, in this mode, the participant's gaze is assumed to be uniformly distributed on the $1080\,\text{px} \times 1920\,\text{px}$ display.

Data Collection

We analyze a TrackIt and eye-tracking dataset originally collected by Kim et al. (2020), publicly available on OSF (`https://osf.io/u8jbs/`). Python code for reproducing our analyses is available on GitHub (`https://github.com/sss1/hmm`). Here, we briefly review the data collection process of Kim et al. (2020). See Kim et al. (2020) for further details regarding the experimental procedure, participant demographics, TrackIt parameters, and data preprocessing.

Kim et al. (2020) recruited 50 typically-developing children, aged 3.5-6 years ($M = 4.60$ years, $SD = 0.67$ years), 23 male and 27 female. Each child performed 10 TrackIt trials, excluding an initial practice trial during which the experimenter explained the task. Because child eye-tracking data contains many missing values, after linearly interpolating short (≤ 10 frames, ≈ 167 ms) intervals of missing gaze data, Kim et al. (2020) discarded all data from 8 children with $> 50\%$ of eye-tracking data missing in > 5 of trials. This left data from 42 children (420 total trials), aged 3.5-6 years ($M = 4.65$ yrs, $SD = 0.71$ yrs), 17 male and 25 female.

Model Implementation & Fitting

We implemented the model in Python using TensorFlow 2.5 (Abadi et al., 2016), specifically using the Hidden Markov Model class provided by the TensorFlow Distributions library (Dillon et al., 2017). The hierarchical model was implemented as a "flattened" HMM with 9 states (one for each of the 7 objects in the Distractible mode, one for the Optimally Engaged mode, and one for the Disengaged mode) corresponding to possible mode-state pairs in the HHMM.

We trained the model in a fully unsupervised way by maximizing the likelihood of participant's data. Specifically, attentional mode and state sequences were fit jointly with all 11 model hyperparameters (6 free parameters in Π_{Modes}, 2 free parameters in π_{Modes}, 1 free parameter in $\Pi_{\text{Distractible}}$, and 2 free parameters in the Gaussian emission distribution), independently for each of the 42 participants. After specifying the model in TensorFlow, the model parameters were fit to a participant by maximizing the log-likelihood of all 10 non-practice trials of data collected from that participant. Optimization was performed using 10^3 iterations of the Adam optimizer (Kingma & Ba, 2014) with TensorFlow's default learning rate (10^{-2}). The following initial values were used for the optimization procedure. Π_{Modes} was initialized with all off-diagonal values 0.005 (corresponding to an average mode switch every $3.\overline{33}$ s). π_{Modes} was initialized with uniform initial probability $1/3$ for each mode. $\Pi_{\text{Distractible}}$ was initialized with off-diagonal values $c_2 = .05$ (corresponding to an average object switch every $0.\overline{33}$ s). The Gaussian emission distribution was initialized with variances $\sigma_x = \sigma_y = 100$ px. After training the model, each participant's most likely sequence of attentional modes and states in each trial was computed by the Viterbi algorithm (Forney, 1973).

Results

Fitted HHMM Parameters

We first present values of the parameters of the HHMM fitted to participants' eye-tracking and TrackIt data. Table 1 gives descriptive statistics for fitted HHMM parameters across the

Table 1: Univariate statistics for distributions (across 42 participants) of each parameter of the HHMM learned from participants' data. DT, OE, and DE denote Distractible, Optimally Engaged, and Disengaged modes, respectively.

Measure	Mean	Std. Dev.	Min	Max
π_{DT}	.27	.23	1×10^{-3}	.89
π_{OE}	.64	.24	5×10^{-4}	.995
π_{DE}	.09	.12	9×10^{-4}	.57
$\Pi_{DT \to DT}$	.98	.009	.96	.997
$\Pi_{DT \to OE}$	.01	.01	3×10^{-5}	.04
$\Pi_{DT \to DE}$	.004	.003	4×10^{-5}	.02
$\Pi_{OE \to DT}$	.003	.001	1×10^{-4}	.009
$\Pi_{OE \to OE}$	.99	.008	.95	.998
$\Pi_{OE \to DE}$	.004	.008	3×10^{-5}	.04
$\Pi_{DE \to DT}$	.005	.003	7×10^{-6}	.01
$\Pi_{DE \to OE}$	.01	.03	7×10^{-6}	.22
$\Pi_{DE \to DE}$	.98	.03	.78	.998
c_1	.995	.001	.992	.999
c_2	.005	.001	.001	.008
σ_x	86.98	12.98	72.38	131.13
σ_y	93.18	13.71	75.01	126.12

42 participants. Broadly speaking, fitted values were consistent with expectations for participants' behavior. For example, for all participants, for transitions both across modes (Π_{Modes}) and between objects within the Distractible mode ($\Pi_{\text{Distractible}}$, given in terms of c_1 and c_2 in Eq. 1), the probability of staying within the same mode or state was much greater than that of transitioning. Also, although there was significant variation between participants, participants tended to begin trials in the Optimally Engaged mode ($\pi_{OE} = 64\%$), although they occasionally began trials Distractible ($\pi_{DT} = 27\%$) and rarely began trials Disengaged ($\pi_{DE} = 9\%$).

Optimally Engaged Mode and TrackIt Performance

We next tested the hypothesis that the Optimally Engaged mode supports TrackIt task performance, as measured by Location Accuracy. Consistent with this hypothesis, the proportion of frames a participant spent in the Optimally Engaged mode (according to the HHMM) was strongly correlated with their Location Accuracy ($r = .84$, 95% CI $(.72, .91)$, $p < .001$ for the null hypothesis of 0 correlation; Student (1908)). This correlation was stronger than the correlation of .71 between Location Accuracy and the proportion of frames in which the participant was classified as attending to Target, though the difference between these correlations was not significant.

Ablation Study

To evaluate the importance of each of the three HHMM modes, we ran an ablation study, in which we compared the fit of the full model to the fit of each of the three submodels in which one of the modes is removed. To do this, for each participant, we performed 10-fold cross-validation, splitting trials into 9 training trials and 1 test trial, fitting the model to the training trials, and then computing the likelihood of the test trial's gaze data. Since the duration of TrackIt trials varied randomly and the log-likelihood of a trial decreases roughly linearly with the duration of that trial, we normalized each trial's log-likelihood by dividing by the number of frames in that trial. The mean normalized test likelihoods of the full model and each submodel are reported in Table 2.

Table 2: Results of ablation study. Means and standard errors (across 42 participants) of log-likelihood on held-out test trial (averaged across cross-validation folds and normalized by trial duration), for each model. Higher (less negative) values indicate greater likelihood of observing the held-out test data under each model, after fitting to the same participant's training data. Bold values indicate log-likelihoods statistically indistinguishable from those of the best model.

Model	log-Likelihood ($\pm$ std. err)
Full Model	$\mathbf{-9.799 \pm 0.030}$
No Distractible Mode	-9.983 ± 0.028
No Optimally Engaged Mode	-9.932 ± 0.028
No Disengaged Mode	$\mathbf{-9.803 \pm 0.029}$

The results indicate that the full model and submodel without the Disengaged mode both fit significantly better than submodels without the Distractible or Optimally Engaged modes, suggesting that the Distractible and Optimally Engaged modes both explain significant proportions of participants' behavior. However, fits of the full model and the submodel without the Disengaged mode were not statistically distinguishable. This is likely due to the fact, discussed below, that participants spent far less time in the Disengaged mode than in Distractible or Optimally Engaged modes.

Developmental Results

We next used the HHMM to investigate how the proportion of time participants spend in each mode changes with age. Linear regressions, illustrated in Figure 3, indicated significant effects of Age on the proportions of time spent in the Distractible mode ($t(40) = -4.81$, $p < .001$, $R^2 = .41$) and in the Optimally Engaged mode ($t(40) = 4.21$, $p < .001$, $R^2 = .37$), but not on the proportion of time spent in the Disengaged mode ($t(40) = 0.49$, $p = .60$, $R^2 = .02$). These results, specifically the effects of Age on the proportions of time spent in the Distractible and Optimally Engaged modes, were consistent with the hypotheses described in the Introduction.

Discussion

In the context of the TrackIt task, we implemented a hierarchical HMM model of attentional modes over time and fit the model's parameters using participants' gaze data. We found a strong correlation between the model's Optimally Engaged mode and participants' task performance, measured by Location Accuracy. Our ablation study suggested that the Distractible and Optimally Engaged modes both played significant roles in explaining participant behavior, in terms of the model's ability to predict a participant's gaze behavior on held-out trials using data from the participant's other trials.

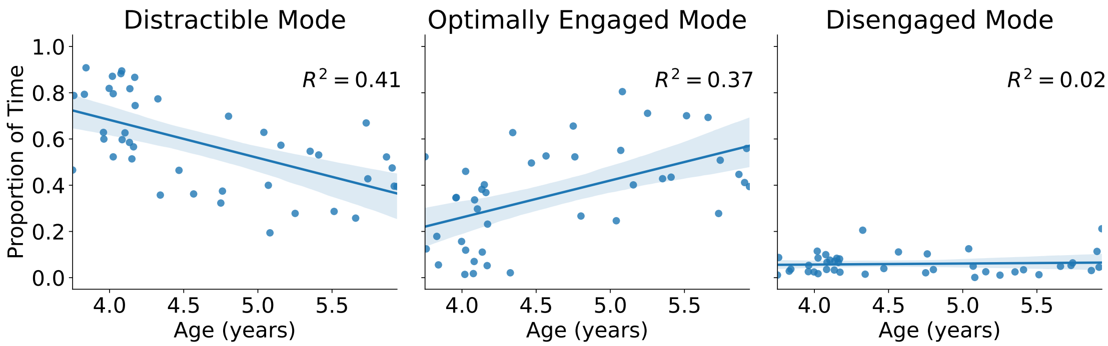

Figure 3: Proportion of time spent in each mode, according to the HHMM, as a function of participant age. Shaded regions indicate bootstrapped 95% confidence bands.

In contrast, the Disengaged mode did not provide significant explanatory power, and predictions from our model suggested that participants spend a very small proportion of time ($< 5\%$ of frames) in the Disengaged mode. A simple explanation for this is that the TrackIt task was short and highly engaging, even for young children. However, based on manual analysis of the same data by human coders, (Kim et al., 2020) suggested that participants spent significantly more time (14% of frames) in an "Off Task" state, in which gaze was decoupled from the trajectories of any TrackIt objects. Anecdotally, from visualizations of children's gaze behavior, we observed that children sometimes exhibited a "zoning out" behavior, in which their gaze remained fixated at a single point on the screen, independent of moving TrackIt objects; this was not captured by the Distractible mode in our model, which typically captured only when participants' gaze moved far away from any of the TrackIt objects (e.g., to the edge of the display, away from the TrackIt grid). This suggests the HHMM might underestimate how much time participants spend in the Disengaged mode. A more refined model of gaze behavior in the Disengaged state (e.g., incorporating gaze velocity to better distinguish tracking an object from Disengaged gaze that, by chance, falls near a object) might help address this. Overall, more work is needed to understand whether children enter the Disengaged mode while performing TrackIt, and, if so, how they behave in this mode.

Our developmental analyses supported both of our hypotheses: time spent in the Distractible mode decreased with age, and time spent in the Optimally Engaged mode increased with age. Since the HHMM modes were motivated by the Adaptive Gain Theory (AGT) of LC-NE function (Aston-Jones & Cohen, 2005), which asserts that the Distractible mode serves the purpose of promoting exploration, these findings are consistent with the possibility that documented decreases in exploratory behavior with development over the course of early childhood (Mehlhorn et al., 2015; Blanco & Sloutsky, 2020b; Gopnik, 2020) may be explained by functional changes in the LC-NE system. Since anatomy of the LC is believed to mature during infancy, much earlier than the age range studied in this paper (McLean & Shipley, 1991; Marshall, Christie, Finlayson, & Williams, 1991; Nakamura & Sakaguchi, 1990), we hypothesize that such changes in LC function may stem from changes in higher-order brain regions, such as prefrontal cortex, that both undergo development in this age range (Casey, Giedd, & Thomas, 2000; Diamond, Briand, Fossella, & Gehlbach, 2004; Posner & Rothbart, 2007) and modulate LC activity (Jodoj, Chiang, & Aston-Jones, 1998; Aston-Jones & Cohen, 2005).

Further work is needed to strengthen the connection we hypothesize between modes of attending behavior in TrackIt and modes of LC activity as proposed by the AGT. Since the LC is small and deep within the brain, most studies directly relating LC activity to behavior have relied on invasive electrophysiological recordings in non-human primates (Rajkowski, 1993). However, small fluctuations in pupil diameter have also been shown to be tightly coupled to LC activity (Mathôt, 2018), and so a feasible approach to investigating this in humans may be through measurement of pupil diameter as participants perform TrackIt; if the modes of the HHMM correspond well to those of the AGT, then we would expect to see distinct patterns in participant's pupil dilation corresponding to their mode as identified by the HHMM.

Conclusion

Although much of attention research has characterized sustained attention through the degree of engagement with a given task, attention is not merely a mechanism for focusing intently on a single task, but also a continuously operating process by which humans can balance multiple competing priorities by interweaving them over time. Attentional modes, whose role may be unclear within the context of performance on a single task, may play a central role in guiding attention to adaptively subserve behavior over time. This paper presented a computational model allowing for the identification and measurement of certain attentional modes, motivated by the Adaptive Gain Theory of the LC-NE system, and provided evidence that changes in the employment of these modes might explain changes in the allocation of attention over the course of young children's development.

Acknowledgments

This work was supported by the US National Science Foundation under Grants No. BCS-1451706 and DGE-1252522, and by the German Federal Ministry of Education and Research (BMBF) through the Tübingen AI Center (FKZ: 01IS18039B).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... others (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI) 16)* (pp. 265–283).

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of vision*, *14*(4), 1–1.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*(1), 29.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403–450.

Berridge, C. W., & Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain research reviews*, *42*(1), 33–84.

Blanco, N. J., & Sloutsky, V. M. (2020a). Attentional mechanisms drive systematic exploration in young children. *Cognition*, *202*, 104327.

Blanco, N. J., & Sloutsky, V. M. (2020b). Systematic exploration and uncertainty dominate young children's choices. *Developmental Science*, e13026.

Casey, B., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological psychology*, *54*(1-3), 241–257.

Deng, S., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62.

Diamond, A. (2006). *The early development of executive functions*. Oxford University Press.

Diamond, A., Briand, L., Fossella, J., & Gehlbach, L. (2004). Genetic and neurochemical modulation of prefrontal cognitive functions in children. *American Journal of Psychiatry*, *161*(1), 125–132.

Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., ... Saurous, R. A. (2017). Tensorflow distributions. *arXiv preprint arXiv:1711.10604*.

Dubois, M., Aislinn, B., Moses-Payne, M. E., Habicht, J., Steinbeis, N., & Hauser, T. U. (2020). Tabula-rasa exploration decreases during youth and is linked to adhd symptoms. *BioRxiv*.

Dubois, M., Habicht, J., Michely, J., Moran, R., Dolan, R., & Hauser, T. (2020). Noradrenaline modulates tabula-rasa exploration. *BioRxiv*.

Fisher, A. V., & Kloos, H. (2016). Development of selective sustained attention: The role of executive functions. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (p. 215-237). Washington, DC, US: American Psychological Association.

Fisher, A. V., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3-to 5-year-old children: Evidence from a new paradigm. *J of Experimental Child Psychology*, *114*(2), 275–294.

Forney, G. D. (1973). The Viterbi algorithm. *Proceedings of the IEEE*, *61*(3), 268–278.

Gilzenrat, M. S., Cohen, J., Rajkowski, J., & Aston-Jones, G. (2003). Pupil dynamics predict changes in task engagement mediated by locus coeruleus. In *Society for neuroscience abstracts* (Vol. 515, p. 19).

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B*, *375*(1803), 20190502.

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in psychology*, *5*, 31.

Jodoj, E., Chiang, C., & Aston-Jones, G. (1998). Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. *Neuroscience*, *83*(1), 63–79.

Keebler, E., Kim, J., Stanley, O., Thiessen, E. D., & Fisher, A. V. (2020). Using the TrackIt task to measure the development of selective sustained attention in children ages 2-7. In *Cogsci*.

Kim, J., Singh, S., Thiessen, E. D., & Fisher, A. V. (2020). A hidden markov model for analyzing eye-tracking of moving objects. *Behavior research methods*, *52*(3), 1225–1243.

Kim, J., Vande Velde, A., Thiessen, E. D., & Fisher, A. V. (2017). Variables involved in selective sustained attention development: Advances in measurement. In *Proceedings of the 39th annual conf. of the Cognitive Science Society.*

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition*, *168*, 16–26.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. *Psychological bulletin*, *139*(4), 870.

Laureiro-Martínez, D., Brusoni, S., & Zollo, M. (2010). The neuroscientific foundations of the exploration-exploitation dilemma. *Journal of Neuroscience, Psychology, and Economics*, *3*(2), 95.

Marshall, K., Christie, M., Finlayson, P., & Williams, J. (1991). Developmental aspects of the locus coeruleus-noradrenaline system. *Progress in Brain Research*, *88*, 173–185.

Mathôt, S. (2018). Pupillometry: psychology, physiology, and function. *Journal of Cognition*, *1*(1).

McLean, J. H., & Shipley, M. T. (1991). Postnatal development of the noradrenergic projection from locus coeruleus to the olfactory bulb in the rat. *Journal of Comparative Neurology*, *304*(3), 467–477.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., … Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the brain takes a break: a model-based analysis of mind wandering. *Journal of Neuroscience*, *34*(49), 16286–16295.

Nakamura, S., & Sakaguchi, T. (1990). Development and plasticity of the locus coeruleus: a review of recent physiological and pharmacological experimentation. *Progress in neurobiology*, *34*(6), 505–526.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of selective attention: When children notice what adults miss. *Psychological science*, *28*(6), 723–732.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. *Modes of perceiving and processing information*, *137*(158), 2.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25–42.

Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain*. APA Books.

Rajkowski, J. (1993). Correlations between locus coeruleus (lc) neural activity, pupil diameter and behavior in monkey support a role of lc in attention. *Soc. Neurosc., Abstract, Washington, DC, 1993*.

Riccio, C. A., Reynolds, C. R., Lowe, P., & Moore, J. J. (2002). The continuous performance test: a window on the neural substrates for attention? *Archives of clinical neuropsychology*, *17*(3), 235–272.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of consulting psychology*, *20*(5).

Ruff, H. A., & Rothbart, M. K. (2001). *Attention in early development: Themes and variations*. Oxford University Press.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PloS one*, *6*(3), e18298.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological bulletin*, *132*(6), 946.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual review of psychology*, *66*, 487–518.

Student. (1908). Probable error of a correlation coefficient. *Biometrika*, 302–310.

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 601–615.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*(5), 553–566.