Cognitive Computation (2018) 10:703–717
https://doi.org/10.1007/s12559-018-9558-9

A Generative Model of Cognitive State from Task and Eye Movements

W. Joseph MacInnes[1] · Amelia R. Hunt[2] · Alasdair D. F. Clarke[3] · Michael D. Dodd[4]

Received: 19 November 2016 / Accepted: 29 April 2018 / Published online: 9 May 2018

Abstract

The early eye tracking studies of Yarbus provided descriptive evidence that an observer's task influences patterns of eye movements, leading to the tantalizing prospect that an observer's intentions could be inferred from their saccade behavior. We investigate the predictive value of task and eye movement properties by creating a computational cognitive model of saccade selection based on instructed task and internal cognitive state using a Dynamic Bayesian Network (DBN). Understanding how humans generate saccades under different conditions and cognitive sets links recent work on salience models of low-level vision with higher level cognitive goals. This model provides a Bayesian, cognitive approach to top-down transitions in attentional set in pre-frontal areas along with vector-based saccade generation from the superior colliculus. Our approach is to begin with eye movement data that has previously been shown to differ across task. We first present an analysis of the extent to which individual saccadic features are diagnostic of an observer's task. Second, we use those features to infer an underlying cognitive state that potentially differs from the instructed task. Finally, we demonstrate how changes of cognitive state over time can be incorporated into a generative model of eye movement vectors without resorting to an external decision homunculus. Internal cognitive state frees the model from the assumption that instructed task is the only factor influencing observers' saccadic behavior. While the inclusion of hidden temporal state does not improve the classification accuracy of the model, it does allow accurate prediction of saccadic sequence results observed in search paradigms. Given the generative nature of this model, it is capable of saccadic simulation in real time. We demonstrated that the properties from its generated saccadic vectors closely match those of human observers given a particular task and cognitive state. Many current models of vision focus entirely on bottom-up salience to produce estimates of spatial "areas of interest" within a visual scene. While a few recent models do add top-down knowledge and task information, we believe our contribution is important in three key ways. First, we incorporate task as learned attentional sets that are capable of self-transition given only information available to the visual system. This matches influential theories of bias signals by (Miller and Cohen Annu Rev Neurosci 24:167–202, 2001) and implements selection of state without simply shifting the decision to an external homunculus. Second, our model is generative and capable of predicting sequence artifacts in saccade generation like those found in visual search. Third, our model generates relative saccadic vector information as opposed to absolute spatial coordinates. This matches more closely the internal saccadic representations as they are generated in the superior colliculus.

Keywords Eye movements · Task · Cognitive state · Dynamic Bayesian network · Temporal model · Cognitive model

✉ W. Joseph MacInnes
jmacinnes@hse.ru

Amelia R. Hunt
a.hunt@abdn.ak.uk

Alasdair D. F. Clarke
a.clarke@essex.ac.uk

Michael D. Dodd
mdodd2@unl.edu

[1] School of Psychology, National Research University Higher School of Economics, Moscow, Russian Federation

[2] School of Psychology, University of Aberdeen, Aberdeen, UK

[3] School of Psychology, University of Essex, Colchester, UK

[4] University of Nebraska-Lincoln, Lincoln, NE, USA

Introduction

The goal of many psychologists and neuroscientists who study vision is to "reverse engineer" the human visual and oculomotor system: that is, to analyze an end product (e.g., a sequence of eye movements) to understand the system that produced it. To this end, researchers often use two different complementary approaches: decoding and simulation. Decoding underlying cognitive function has always been a goal of experimental psychology, but the surging popularity of Brain-Computer-Interfaces (BCI) [1] has led to an increased interest in this approach, especially as it relates to the use of classifiers on neural (e.g., [2]) and behavioral [3] data. Simulation, on the other hand, uses generative algorithms to understand cognitive processes by re-creating human-like behavior to determine the underlying cause. Simulations are also prevalent in robotics and computer vision applications [4]. We begin by first decoding human goals and tasks using data from high-speed eye tracking, and then second, we simulate relative eye movement properties using a generative Bayesian Model.

The human retina has a variable distribution of photoreceptors, with the highest resolution in the central fovea. To bring various parts of a scene or image to this high-resolution zone, we move our eyes frequently with ballistic eye movements called saccades. Fixations are periods of relative stability between saccades, typically lasting between 200 and 300 ms and allowing efficient sampling of selected locations. The generation of eye movements involves a robust neural network [5] and is influenced by bottom-up image salience [6, 7], expectation [8], motion [9], top-down control [10], biases [11, 12], and midlevel attention [13, 14].

One way of predicting human fixations is by finding areas of interest in natural images. These salience maps are of interest to both psychology and computer vision, and a popular way of measuring the success of these algorithms is by comparing the predictions to actual human fixations (The MIT Salience benchmark, 15). The most successful algorithms at this benchmark have similarities to theories in human visual processing. For example, the classic Itti and Koch [6] salience model is based on Feature Integration Theory [15], and the more recent and accurate deep learning models mimic layered feature extraction in the early visual cortex [16]. Models of this type have been used in applications such as image classification [17], object recognition [18], object segmentation [19], and reducing false alarms in motion detection [20]. Models that combine information from multiple sources have also been implemented such as [21] who combined bottom-up, top-down, and mid-level visual processing. Multi-model cognitive fusion [22, 23] has also been used to combine information from multiple modalities.

These models typically treat the viewed scene coordinates as an invariate map, or *spatiotopic coordinates*. While this is an accurate representation of the viewed scene, the native internal visual representation for humans is *retinotopic* [24], meaning that the representation of visual information shifts with every saccade. The superior colliculus (SC) is essential for saccade and fixation generation, and superficial layers receive retinotopic input directly from the retina as well as other areas [25]. Deeper layers integrate visual information with other modalities and coordinate motor responses including oculomotor responses deeper in the brain stem. Saccade generation in the SC is based on a retintopic map with neural activation on this map triggering a saccade in the matching retinal vector.

There have been attempts to incorporate specific retinal properties into these salience maps and models. Adoubib [26], for example, created a model of visual processing in the human ventral visual pathway by including information such as viewer distance and retinal sampling. The model maintains the same attention selection mechanism as Itti and Koch [6]—namely a winner take all fixation selection process combined with temporal inhibition (IOR, 27)—but uses a point cloud distribution to allow for non-rectangular representation and modifies this further with known retinal and angular artifacts. Similarly, Curtsurdis [4] has created more all-encompassing models of the full visual pathways, including an aspect of cognitive control. These models build on the classic salience model [6] and add object maps, goals, saccadic motor execution, and an overseer to control selection. The goal module enhances or inhibits appropriate lower level signals, while the Overseer module focuses on reward of appropriate actions. Collectively, these models do approximate the spatial distribution of fixations on a given image; however, these models do not (a) predict other saccadic properties (velocity, for example), (b) provide insight into the cognitive state of the viewer, or (c) try to capture known patterns of fixation sequences. Additionally, we know that fixation locations are not independent and can be influenced not only by IOR (mentioned above) but by the visual system programming saccades in parallel [28, 29]. In terms of vector sequences, for example, we know that repeat vectors are more common in visual search with an additional peak at reverse vectors [30, 31].

Eye movements provide an overt (but imperfect) measure of attention, so it is tempting to suggest that eye movement patterns can provide insight into internal cognitive states. Fecteau and colleagues [32] discovered that both saccadic reaction times and neural firing rates in the superior colliculus were modified by the degree to which the visual stimulus in their receptive field was relevant to the task. This has led to the proposal of an attentional network which is tuned to goals and priorities as much as to visual salience [33]. Performance in cuing and search has also been shown to be influenced by the current attentional set of the observer [34].

In a seminal experiment, Yarbus [35] demonstrated that different instructions could produce different patterns of eye

movements for a given image. A single observer was shown the painting "The unexpected visitor" by Ilya Repin and given various instructions including the following: estimate the material circumstances of the family, give the ages of the people in the portrait, remember the clothing worn by the people, and estimate how long the visitor had been absent from the family. Patterns of eye movements were shown to be different depending on instruction (see Fig. 1), and this result has been replicated for different tasks [36–38]. Task has also been shown to influence lower level saccadic properties such as number of fixations, gaze duration [39], and fixation duration [40]. Recently, Kardan and colleagues [41] showed that task instructions not only influenced saccadic amplitude and fixation duration, but also modulated how these performance features were influenced by low-level scene features. This suggests that while instructions can influence how we move our eyes, the link between task and eye movements may not be direct.

When we provide explicit instructions to our observers regarding the nature of an experimental task, instructions are probably one of many factors that influence an observer's internal attentional state. Attention to a task varies over time as measured by behavior, self-report, and Alpha channel activity [42] and could be modulated in a given task by the locus coeruleus-norepinephrine (LC-NE) system to promote either highly engaged (phasic) or disengaged (tonic) behavior [43]. To model how task influences saccadic selection, we propose a model where current task or instruction is only one influence to observers' internal cognitive state, and this hidden internal state is a driver of saccadic selection. This differs from Kardan [41] in that we do not model scene features, but we do include an intermediate mechanism—a hidden cognitive state—that could account for the way that task mediates gaze control. We also introduce an explicit temporal component that allows this cognitive state to change over time within any given instructed task. We also explore a variety of saccadic and fixation features to determine which may be more diagnostic of task.

Inferring a category, such as task, from a set of observations is called a *classification problem* in machine learning [44]. A number of recent studies have explored the saccadic-task correlation using a classification approach; that is, given only an observed set of eye movements, can

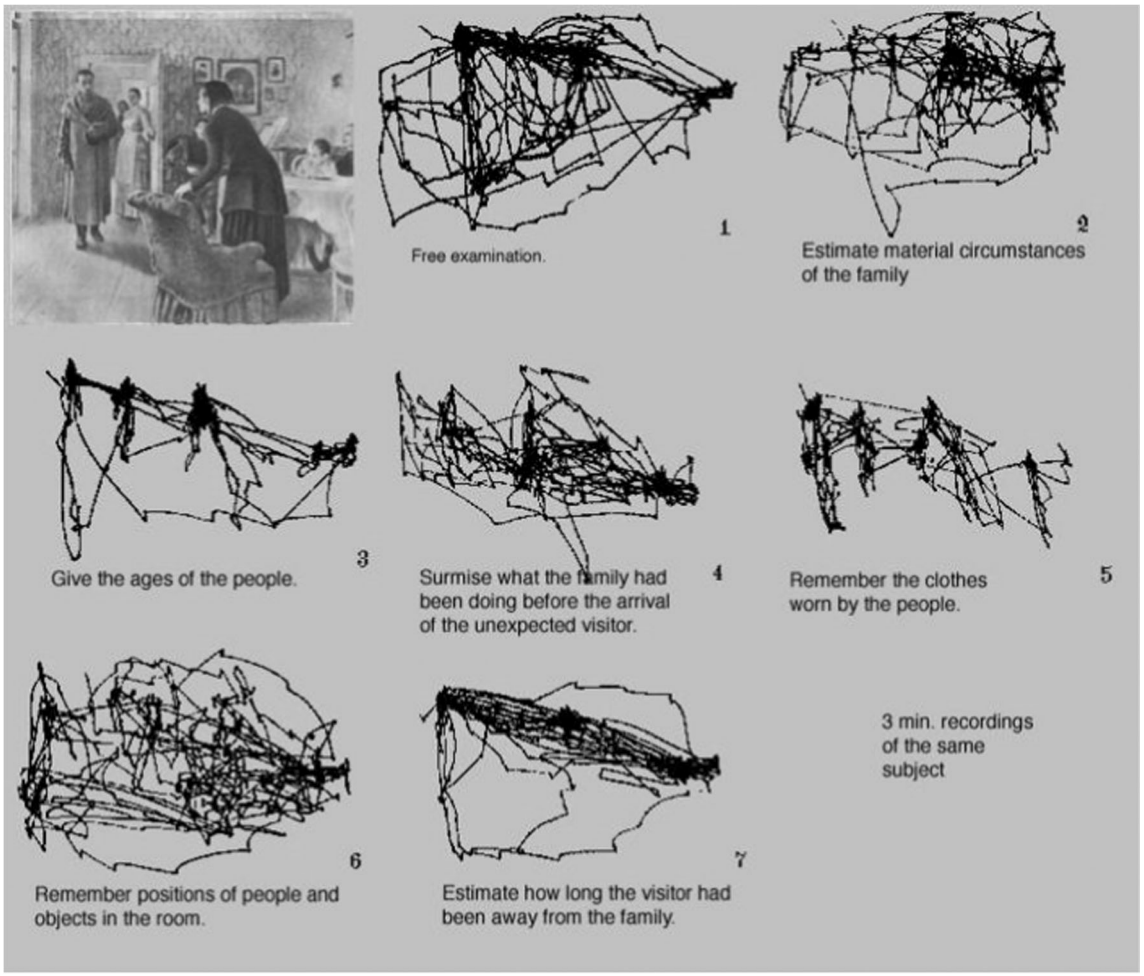

Fig. 1 Scene and instructions from Yarbus, (1967), Fig. 107. Reprinted with permission, 2017 (Springer)

we accurately determine the task that the subject was instructed to perform? Greene [45] demonstrated that a simple set of saccadic features such as number of saccades, mean saccadic amplitude, and mean fixation duration could successfully predict the observer or the image being viewed. However, their method was not able to predict task. Recently, researchers using different datasets (tasks and images) and different feature sets (learning algorithms and decision rules) have had more success. Henderson et al. [46] were able to classify Search and Reading tasks at accuracies of up to 75%, which was well above chance for the given tasks. Borji and Itti [3] were able to classify the original data from Greene et al. [45] (34% accuracy, chance was 25%) and also all seven of Yarbus' original tasks at above chance levels. Borji also went further by identifying saccade metrics that are particularly influenced by image-level properties, and therefore less useful for classifying the instruction set irrespective of these properties. Specifically, they showed that the position of fixations in an image only informed classifier accuracy for trials within that same image (though they did not account for spatial patterns like central bias [12, 47] or symmetry [48]).

In the present paper, we create a model of cognitive state that is capable of generating saccadic properties based on instructed task. We will start with a data set that has already shown behavioral differences across task and (a) determine which saccadic features are diagnostic of task, (b) infer how cognitive state changes over time by clustering eye movement properties, and (c) create a generative model of eye movement vectors based on shifts of cognitive states for given tasks. We use data from Dodd et al. [49], who asked observers to perform one of four tasks, either search for a specified target, remember a scene, rate its pleasantness, or view the scene without any particular instructions. They observed Inhibition of Return (IOR, which in this case was defined as slower saccadic responses to probes presented in a location that had recently been fixated) only during the search task, and not during the others. Mills et al. [40] further showed that these instructions generate differing spatial and temporal saccadic properties, and further, these saccadic properties are sufficient for human observers to infer another's task [50] or search objective [51]. Given that their observers followed the instructions well enough for this difference to emerge, we believe that eye movements for these tasks should diverge in other ways that could be discovered by our model. We began with an exploration of various saccadic features to connect our model to the recent literature and also to determine exactly which saccadic and fixation properties were important for accurately predicting instructed task. We included saccadic latency (fixation duration), saccade duration, amplitude, peak velocity, pupil size, and absolute saccadic angle based on screen direction. Since it is currently unclear how to effectively use spatial information across images [3], we chose not to include region of interest or absolute coordinate salience map information as input to the classifiers or model. While many models of visual salience predict fixation locations in screen or image coordinates, our prediction of saccadic vector properties more closely matches their internal representation as they are generated in the Superior Colliculus [23]. While we do not propose to create a neural model of saccade generation, we do propose a cognitive model of late-stage saccadic generation.

Recent classifiers have been shown to be task-sensitive, using mean fixation and saccadic data collapsed across individual trials [3]. Since a model of saccadic generation would have to work on the level of individual saccades, however, we first looked at which saccadic properties, if any, could be diagnostic of task from single saccades as opposed to saccadic aggregates from the full trial.

In addition to eye movement patterns, discussions of cognitive state should also include the pupil. Pupillary dilation has been linked to degree of arousal [52], memory load [53], and attentional load [54]. Recent studies have shown correlations between pupil size and effects from the Stroop task [55] and Inhibition of Return [56]. Aston-Jones and Cohen [43] proposed in their adaptive gain theory that pupil size is regulated in part by the locus coeruleus—norepinephrine system (LC-NE). They propose two modes of LC neuronal activity: Phasic—reflecting focused performance on an attended task; and Tonic—which favors exploration over focus on a single task. Posner and Fan [57] have also suggested LC as a key structure in the "alerting" function of attention. Since pupil diameter closely correlates with LC neuronal firing frequency [58, 59], pupil size can serve as an additional measure of attentional focus in our model.

Properties included in classifier and model

- Latency
- Duration
- Amplitude
- Peak velocity
- Absolute saccadic angle
- Pupil size

Methods

Observers and Stimuli

Data used as input for the classifiers were first reported in Dodd et al. [49]. Over 17,000 saccades from 53 observers and 67 photographic images were coded as input to a set of classifiers. Observers were randomly assigned to a group and given one of the following four instructions: Search for the

letter Z or N in the scene, memorize the scene and prepare for a memory test at the end of the session (not actually tested), rate the pleasantness of the picture from 1 to 7, or no specific instruction was given and observers freely viewed the image. All tasks lasted for 8 s and are hereafter referred to as Search ($n = 14$), Memorization ($n = 13$), Preference ($n = 14$), and View ($n = 12$). The visual search task included a probe after 6 s of search on some trials, so only the first 6 s of eye movements were used from each task to equate conditions. Eye movements were measured using an SR Research Eyelink 2 eye tracker sampling at 500 Hz. Nine-point calibration was conducted for each observer, with average validation error of less than .5° visual angle.

Saccade attributes were extracted from each saccade including latency, duration, amplitude, peak velocity, and absolute saccadic angle as compared to the screen's horizontal plane. The relative angle (see Figs. 2 and 3) and relative amplitude of the current saccade compared to the previous one were also calculated.

Pupil size was normalized for all models to account for individual differences and potential luminosity changes across observers. The *Z*-score for mean pupil size during a fixation was calculated based on the mean pupil size for each subject and the *Z*-score of individual means accounts for individual differences in pupil size.

Clustering

Clustering of attentional states based on saccadic properties was performed using the MATLAB clustering and visualization toolbox (Abonyilab.com) with Dunn's Index used to select the optimal number of clusters. Dunn's Index is a score that reflects the cohesion within a cluster and the separateness between clusters [60] and was calculated for numbers of clusters from two to 14. We chose 10 clusters as optimal for this data given that fewer clusters and higher Dunn's Index scores were preferred (Fig. 5).

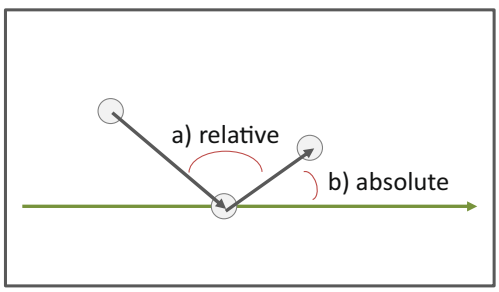

Fig. 2 Illustration of saccadic angle (a) relative to the vector of the previous saccade or (b) compared to the absolute horizontal vector of the screen

Discrete K-Means clustering with 10 centroids was performed on the mean saccade data for each trial to assign the trials to one of the 10 clusters to produce labels for comparisons. The value/location of the K centroids and assignment of each observation to a centroid were learned by attempting to minimize the within-cluster sum of squares of the error between each point in the cluster and cluster centroid.

$$arg\ min\ O \sum_{i=1}^{k} \sum_{x \in O_i} \|x - \mu_i\|^2$$

where O_i is the set of observations currently assigned to cluster I and μ_I is the mean of points in cluster i. The basic K-means algorithm requires the number of clusters to be set and uses Euclidean distance for the centroid calculation, but many options are available [61]. These cluster labels were then cross-tabulated with the original instructed tasks.

Dynamic Bayes Network

Bayesian networks are graphical models that treat evidence as observations of random variables and edges as directional dependencies between variables (see 62 for an overview and tutorial). Probability distribution tables are learned for each node and represent the likelihood of the random variable having a value given only its prior probability and the probability of its parents—variables it is directly dependent on. Our Bayesian networks were trained and tested using the Genie software package [63]. Learning the structure of the DBN graph used a Bayesian graph search, although the exact search algorithm is not reported in the package documentation. It certainly behaves as others in this class of algorithm by computing the posterior probability of potential graphs given the observed data, and maximizing the choice of graph given the observed data:

$$argmax_G P(G|D)$$

Parameter learning for all Bayesian Networks used the Expectation Maximization (EM) algorithm [64].

Continuous saccadic data (see Fig. 3 for initial distributions) were discretized into five bins with divisions chosen to ensure equal numbers of saccades in each bin. Bin sizes of three and seven were also tested to see if binning granularity was important, but classifier results were similar in each case. In previous research [45, 46], input data has been preprocessed so that training data represented the mean value of that participant for an entire trial. We replicated this approach for an initial classifier; however, for the second classifier and the DBN model, we chose to include all saccades from all trials. While this increased the overall number of training examples to the model, it also increased the variance introduced by individual saccades.

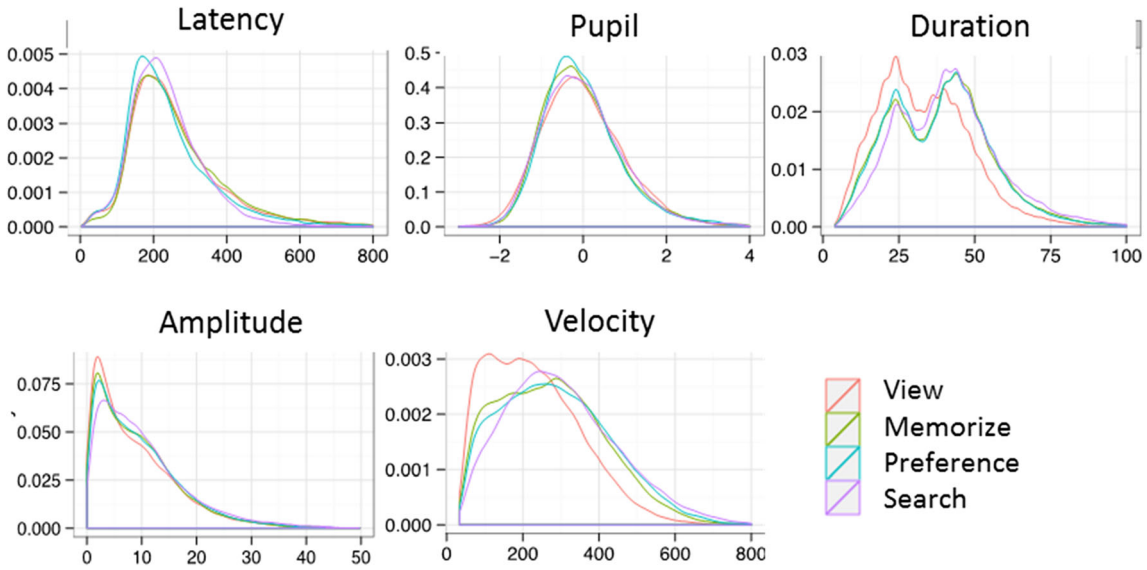

Fig. 3 Density plots for saccadic features split by task. Saccadic features show potential for being diagnostic of task, in particular duration, amplitude, and velocity

Results

Section 1: Classifier

Recent research has established that classifiers can predict task from aggregate trial data [3, 46, 65], so we divert briefly to demonstrate that our chosen features can replicate these earlier results (see Figs. 4 and 5). Ten unique training/testing sets were generated for each of the following analyses by randomly sampling the dataset into independent 90/10% splits. We performed this 90/10 sampling in three different ways to determine how the classifier would perform over the full dataset, across different images and across different subjects. The first analysis used 90/10 splits with any sample chosen from the full dataset, but we followed this with splits where the training samples were chosen from 90% of the images and tested on the remaining 10%. The final analysis split the sampling from 90/10% of subjects. The final two analyses were to test how well the classifier generalizes to new images or new subjects that were not included in the training set. All results were compared using non-parametric Wilcoxon signed-rank test (see 3). Saccadic features and pupil size were trained with an augmented Naïve Bayes network that was able to classify the full data set task at

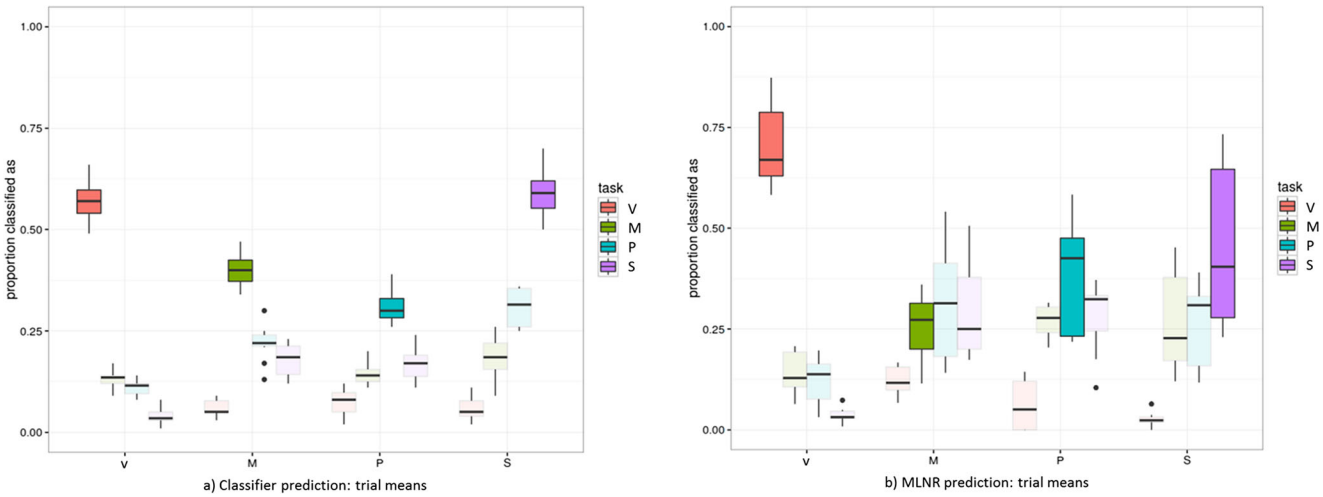

Fig. 4 Classifier accuracy for Augmented Naïve Bayes (**a**) Network and Multinomial Logistic Regression (**b**) predicting task from saccadic and pupil features. The Naïve Bayes outperformed the MNLR and was above chance predicting View (V), Memorize (M), and Search (S) tasks, though both classifiers had difficulty with the Preference (P) task

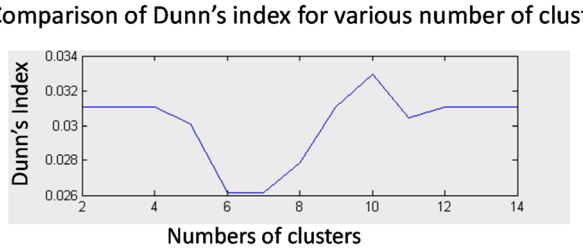

Fig. 5 Dunn's Index calculates the compactness and separateness of clusters. The highest index score suggests ten clusters as an optimal number of hidden cognitive states

53.9%, which is above chance (25%, $t(50) = 3.45$, $p < .001$) and better than a logistic regression (MNLR) using the same parameters (accuracy 45.8%; $t(50) = 3.31$, $p < .001$). Training and testing across images were successful (53.6%, $t(50) = 3.45$, $p < .001$), again with all tasks classified. Training and testing across subject performed well overall (41.2%, $t(50) = 3.45$, $p < .0014$), and all tasks except Preference were accurately classified.

This approach is similar to Greene [45] and Henderson et al. [46], in that we preprocessed the input data so that each example represented the mean value of that observer for an entire trial. We were able to classify task given the mean saccadic data for a trial, but to generate individual eye movements or sequences of eye movements within a trial our generative model should be able to infer task given the parameters of any individual saccade within a trial. The augmented naïve Bayes algorithm was, therefore, retrained with the full set of data with each individual eye movement used to train or test the classifier (Table 1).

Overall classifier accuracy to classify the task from just a single eye movement was above randomized chance ($t = 3.45$, $p < .001$) at 35.8% accuracy, though only Search (53%) and View (45.5%) tasks were accurately classified. The resulting classifier was also biased towards predicting the Search task, which may have exaggerated its accuracy at the cost of reduced Memorize (21.7%) and Preference (23.0%) task accuracy. This differs from the earlier classifier trained on aggregate or summary trial data which was able to predict the memorize task (see also 43) for successful memorize classification on aggregate trial data. Using aggregate data as classifier input and test examples likely removes saccadic outliers that are otherwise more difficult to classify. Classifiers built on each saccade as individual input and test samples, while less accurate on some tasks, are a more complete representation of saccadic behavior.

The overall classifier results demonstrate that our data set and algorithm can replicate other recent classifiers [3, 46, 65]. Our inability to correctly classify most trials in the Preference task across subject demonstrates that, unsurprisingly, not all tasks can generate distinct patterns of eye movements relative to all other tasks. When forming a preference, individual differences in eye movements may combine features of the other tasks, such as looking for details, remembering, and just passive looking. Thus, the accuracy of a classifier will depend critically on which tasks are included in the set [3, 45].

Table 1 Accuracy for all tasks for classifiers with training folds chosen using the full dataset, by Image and by Subject. Asterisks indicate above-chance classification

	Overall accuracy	View	Memorize	Preference	Search
Training folds					
All Data	53.9*	80.3*	44.3*	35.7*	55.0*
By Subject	53.6*	79.0*	42.0*	41.0*	51.0*
By Image	41.2*	67.8*	33.8*	24.2	38.8*
Individual saccade	35.8*	53.0*	21.7	23.0	45.5*

Section 2: Clustering Cognitive State

A computational cognitive model has a different primary goal than a machine learning classifier. While a classifier strives to produce the highest accuracy, a model strives to improve our understanding of a complex system through simulation. A classifier may make use of any algorithm that improves its accuracy regardless of whether it is biologically plausible and can, in fact, exceed human performance on some tasks [66]. While computational cognitive models still may use accurate predictions of experimental data as one measure of fit, they must also match and test our theoretical understanding of the cognitive processes involved, and an improved theoretical understanding may initially come with reduced classification accuracy for a single dataset. Though clusters of gaze location have been used to highlight salient features in video [9], our approach is the first we are aware of to cluster saccadic features so as to infer underlying cognitive state. Our current goal is to maintain as much of our classifier accuracy as possible while moving from a classifier to a more formal model of top-down influences in eye movement generation. Obviously, the function of the cognitive system we are trying to model here is not to classify eye movements, but to generate them. Therefore, the final stage after creating the model is to test the model's ability to generate realistic saccade sequences.

Data training steps for the Dynamic Bayesian Network

1. Verify usefulness of saccadic features
2. Calculate the optimal number of hidden attentional states using Dunn's index
3. Cluster saccadic and pupil data
   (a) Test that clusters of attentional state correlate with instructed task
4. Construct DBN model with new nodes for observed pupil size and hidden attentional states
5. Test model against original data, and against expected saccadic sequences

In the previous section, we assumed that the internal state of each observer reflected which one of the four instruction sets they had been given, which in turn generated saccades from separate—albeit potentially overlapping—distributions. We were able to associate individual eye movements with a particular task instruction, but can this help us formulate a generative model of the control settings that drive eye movement selection? If each distinct control setting generated saccades with different characteristics, then we should be able to discover these distributions through their saccadic behavior. It is likely that each task would not correspond to a single cognitive state, so we will not begin with that assumption. We begin by trying to determine the number of hidden control settings used by observers to generate saccades. Second, we will use a Dynamic Bayesian Network (DBN) [67] to show how these hidden control settings could be modeled as bias signals of state change in a Dynamic Bayesian Network (DBN).

In Bayesian terms, cognitive state is a "hidden" node, meaning that we cannot observe it directly with the present data. We begin by determining the optimal number of distribution clusters for cognitive state based on their statistical cohesion. If attentional or cognitive task control settings generate different distributions of saccades as suggested in "Section 1: Classifier", the number of these distributions should be derivable from the data. We can also test whether these clusters roughly correspond to the classification accuracy seen in "Section 1: Classifier". For example, given the accurate classification of the Search task, we would expect to discover a single cluster which includes most of these eye movements and/or trials. The Preference task, however, could simply reflect a frequent switch between clusters that otherwise reflect *searching*, or *inspecting*. To differentiate the instructed task from the inferred internal state, we will continue using Search, Memorize, Preference, and View as the tasks from "Section 1: Classifier" but refer to the assumed internal states as *searching*, *memorizing*, *judging*, and *inspecting*.

The cross tabulation (Fig. 6a) demonstrates significant overlap between instruction task and internal state (Chi2 [68] = 183, $p < .001$). For example, Cluster 5 shows strong affiliation to the Viewing task, while Clusters 1, 8, and 10 are under-represented in Viewing. Likewise, Search has high overlap with Cluster 1. While Cluster 5 is strongly represented in all four tasks, this is not diagnostic of task and likely represents a default saccadic generation state, such as *inspecting,* shared by many tasks and internal states. Given that the best fitting clusters for the Preference task (clusters 1 and 5) also fit with other tasks, it is not surprising that the classifier had the most difficulty with this task. This could be evidence that forming a preference simply alternates other tasks such as *searching* and *memorizing*.

Not only might tasks contain multiple internal states, but the transitions between states might differ. Even though state nine is not common in any task (Fig. 6a), once an observer enters that state, they are likely to continue (see Table 2) regardless of state. States one and two are well represented in all tasks; however, the transitions between these states differ. For example, in the Viewing task, transitions to state two are more likely from either state one or two, while Memory and Preference tasks are more likely to transition to state one. Search, which was one of the easier tasks to classify, was more likely to maintain state one or two once the observer entered that state. These transitional differences are also highlighted in Fig. 6b–e which show the relative likelihood of transitions compared to average transition performance across all tasks. Given the temporal nature of these state transition differences, we propose a model of cognitive state that is sensitive to changes in state over time, even within a given task.

Section 3: A Model of Task and Internal States

Given that mean saccadic properties on trials can be formulated as distinct clusters (See Fig. 5) and that these clusters are related to task instruction, we present a generative model of eye movements where the internal state is represented by a Dynamic Bayesian Network (DBN). We propose that internal state of the visual system can be represented by a Markov chain with saccades selected from a distribution influenced primarily by the current internal state of the model. Choice of state and state transition is handled by the DBN in such a way as to avoid a selection "homunculus" through a state transition process which depends entirely on the current cognitive state and experimenter instructions. This mimics the bias signals for cognitive control as suggested by Miller and Cohen [69] where the state is self-selecting as an integral part of the DBN itself. An analogy for these bias signals used by the authors was a "self-switching railroad track" and is comparable to the way a DBN switches temporal states using only information that is internally available to the model. As seen with the cluster/task cross tabulation, there is not an exact overlap between instructed task and grouping of saccadic behavior suggesting that instruction alone is insufficient to determine state. State transition analyses also suggest that patterns of internal state change differently according to which instructed task was given. Our DBN will learn these hidden state transitions in order to improve the model beyond instruction task alone.

Saccades were generated by the model through a random selection from saccadic property distributions as determined by the cognitive state associated with the current state of the model (Fig. 7a). The only input into state selection at time t is the previous state (time $t-1$) and input knowledge of the instructed task (Figs. 7b and 8). Miller and Cohen's [69] bias signal is modeled as a random likelihood transition from one internal state to another given the current state and task. While instructed task will influence the successive states, it will do so

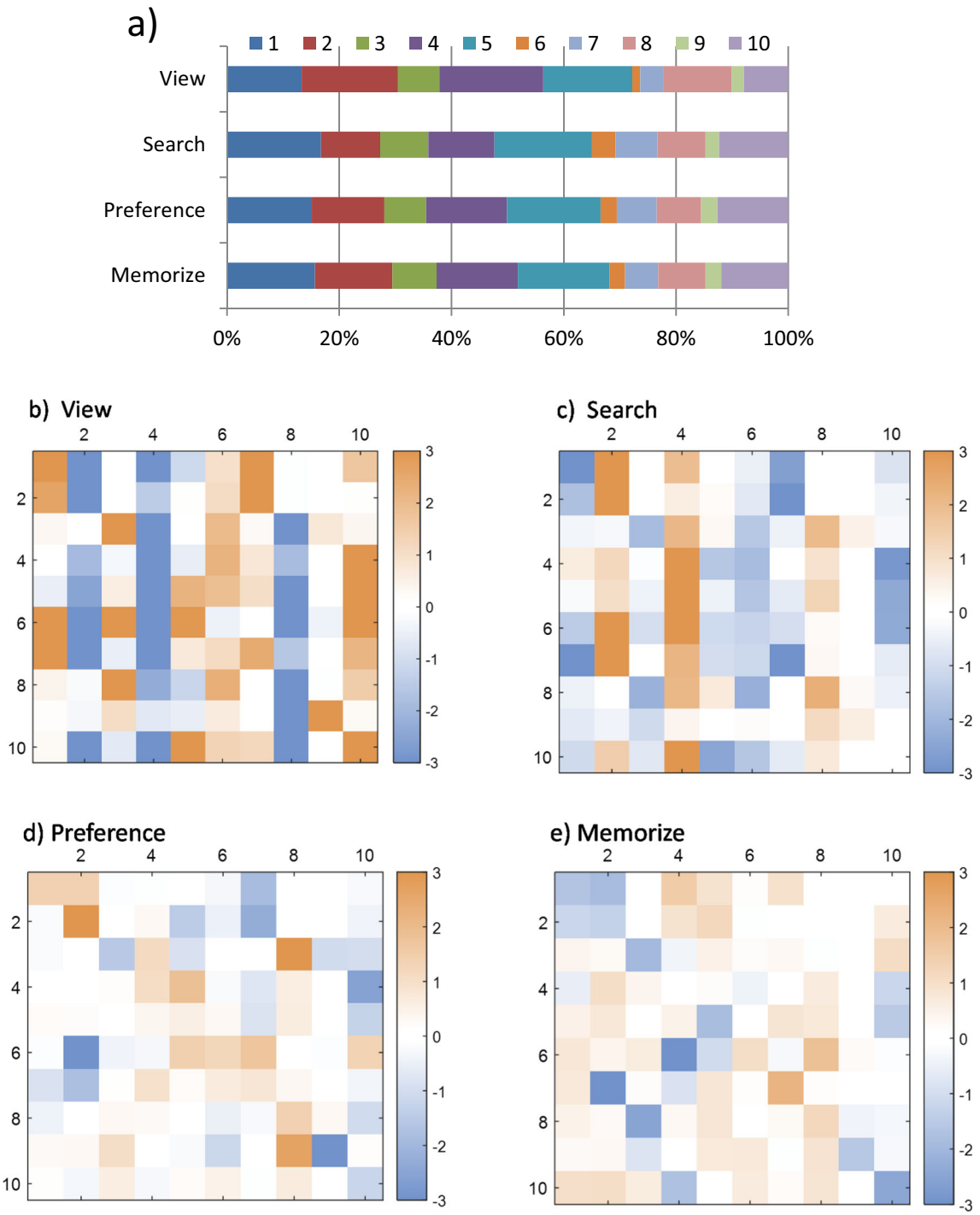

Fig. 6 **a** Cross tabulation mosaic of task condition and absolute likelihood of a clustered internal state. Blocks represent proportion of clustered saccades for each task/cluster combination. Since clusters are based on individual saccades instead of mean trial values, the result echoes the saccade classifier with View and Search being more highly distinguishable. Transition heat maps for **b** View, **c** Search, **d** Preference, and **e** Memorize represent the *difference* in likelihood from the overall data that a task will observe a transition from state A (vertical-axis) to state B (horizontal-axis)

in conjunction with internal state likelihood and according to the learned joint probability distribution for the hidden state transitions given a particular task. To learn these transitions, we will use the individual saccade data (see classifier C from previous section) which, while a less accurate classifier than the mean trial data, will allow us to generate individual eye movements with our model for a given instructed task. Although "Section 1: Classifier" and "Section 2: Clustering Cognitive

Table 2 Table of most likely state transitions in absolute values. Given the task and previous state. "transitions" that repeat the same cluster mean that internal state is most likely to remain the same over time for that task

Previous state	View	Mem.	Pref.	Search
1	2	1	1	1
2	2	1	1	2
3	8	8	3	3
4	4	4	4	5
5	4	5	5	5
6	4	5	5	5
7	2	2	2	1
8	8	3	8	8
9	9	9	9	9
10	4	5	5	5

State" helped justify the choice of parameters to include in the final model, neither section had a direct influence on learning the structure or probability matrices of the model. We maintained the choice of saccadic variables and the learned optimal number of states, but the probability distribution tables correlating task with cognitive state were learned as part of the model using same cross validation scheme outlined below.

Prediction accuracy was again calculated for instructed task given the eye movement properties of an individual saccade. With the internal state as an intermediate, hidden state separating the saccadic data from the instructed task, the new model was still able to predict the task with 36.4% accuracy and was better than chance ($t = 3.45$, $p < .001$) as measured by the Wilcox signed rank test. While improving the theoretical basis of the model, we were still able to classify individual saccades with the same accuracy ($t < 1$) as the original classifier (35.8%). Task prediction was also less biased than the original classifier with preference (23%) and memorize (25.4%) roughly at chance, though they were still likely to be misclassified as Search.

Although cross validation should prevent overfitting of the probability distribution tables, we wanted to check our model for overfitting from the variables themselves, and to determine if all features are diagnostic in the scope of the original classifier. We removed saccadic and pupil features one at a time and compared resulting models by their Aikake Information Criterion (AIC) [70]. AIC is a measure of model fit which compares likelihood scores penalized by model size and is defined by the formula, where $\ln(M)$ is the log likelihood of the trained model and P is the number of parameters:

$$AIC = -2\ln(M) + 2p$$

Although the DBN shifts focus from the classifier to the generative model, it still performs better as a task classifier (smaller AIC) than the Augmented Naïve Bayes despite the additional parameter of the cognitive hidden state (DBN AIC = 1,536,532; classifier AIC = 1,632,743). Since AIC allows comparison of non-nested models, we also compared the full DBN model with each possible reduced model to keep the one with the lowest scored AIC. This process was repeated iteratively until further reductions in model parameters did not improve AIC score. Reducing the initial model (AIC = 1,536,532) by the first parameter showed an improvement (smaller AIC) regardless of which parameter was removed, with the exception of saccadic angle (AIC = 1,537,500). The lowest score was for the model with pupil size removed (AIC = 1,493,033). Removing additional parameters from this model did not, however, result in lower scores (all AICs > 150,000). Removing most single parameters from the full model improved the fit in the first stage, but improvements did not extend to removing multiple parameters. This suggests that many parameters contained redundant information regarding the instructed task. This is consistent with Kardan [65] who found that classifiers performed better if they accounted for feature dependencies.

Generative models can be tested in ways other than classification. Since Dynamic Bayesian models are generative, meaning that they are capable of generating new instances of observations given a trained model, we used the completed model to generate simulated parameters for 10,000 saccades. These saccades were sampled from all tasks and attentional sets and compared to the original saccades from observers' data. If the model accurately reflects the generative process of saccade selection given a specific task and state, then the observers' and model's data should be comparable. We performed a linear mixed effects model of human vs Bayesian model for each saccade parameter given the subject and task as fixed factors and cognitive state as a random factor. There was no significant difference in the human and Bayesian data sets (all Fs < 1) suggesting that the model was able to accurately capture these parameters.

Finally, since the DBN should be able to capture temporal saccadic dependencies in observers' data, we wanted to test whether the model was also capable of reproducing temporal patterns. One such pattern is the large increase in forward saccades and the smaller increase in return saccades when considering the current saccadic angular vector compared to the previous vector. Observers' data for the current tasks (Fig. 9b) is comparable to saccadic analyses from similar research [13, 14] in that saccadic angle at given time $t(x)$ is dependent on the angle of saccade at time $t(x-1)$. Saccades generated from the classifier in "Section 1: Classifier" do not code these temporal dependencies and simply choose from the distribution of typical absolute angles. These absolute angles have an overall horizontal bias in absolute angle, resulting in a relative angle bias of repeating this direction. The DBN does code temporal dependencies, however, and the relative saccadic angle of saccades generated from the DBN shows the tendency to repeat

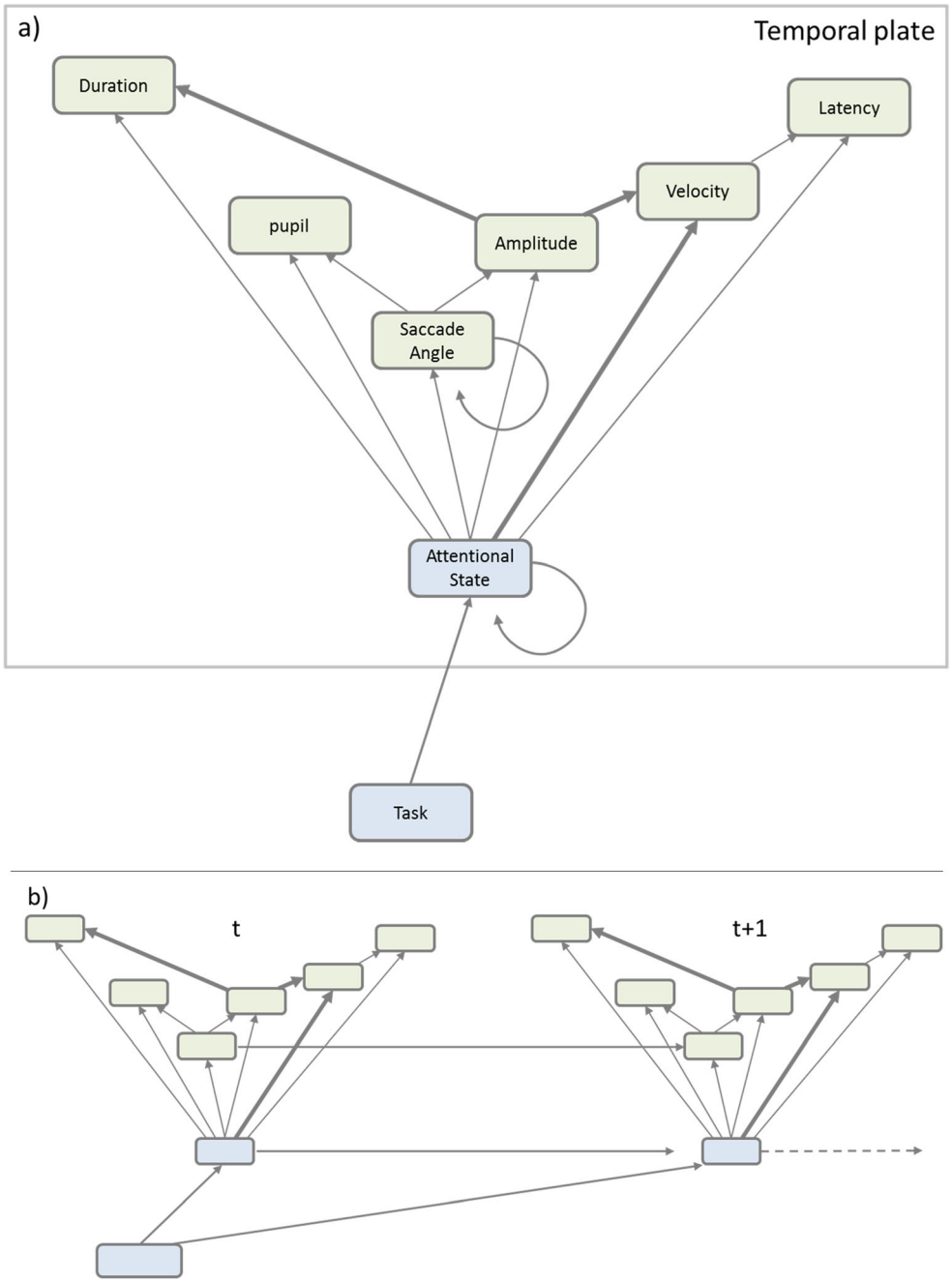

Fig. 7 Dynamic Bayesian Network **a** with hidden cognitive state of the observer influenced only by instructed task and state at previous time. State influences the selection of saccadic properties while that state is active. Nodes inside the temporal plate are free to change with each time unit (saccade), while Task is held steady throughout the trial. The circular gray arrow represents temporal dependency and in this model is restricted to the hidden cognitive state and saccade angle. The same model **b** with time "unrolled" to better show the temporal dependency. Task is only set once for the entire sequence, while cognitive state and saccadic direction have the potential to update on every time unit based on the original task and the previous state. Other eye movement and pupil properties also change every time unit but only based on the current cognitive state

saccadic direction *and* the smaller tendency to reverse direction found in other search studies [13, 14].

General Discussion

While a Bayesian representation is not a neural level model, we believe that our DBN is grounded as a plausible cognitive description of neural task states and saccadic selection. The prefrontal cortex (PFC) is associated with executive control, including an "attentional set" or cognitive state that provides a framework for selecting task-relevant information [71, 72]. Different regions of the PFC activate depending on the nature of this information and processing resources needed for a given attentional set [72]. Switching between these states could be implemented in the PFC through a control system that biases activity to the appropriate network given any combination of sensory input, current state, and desired consequences [69] or possibly directed by the measured distance between current state and subjective goal [73]. Once selected, the appropriate state would guide top-down selection of eye movements through connections to the frontal eye fields [74, 75].

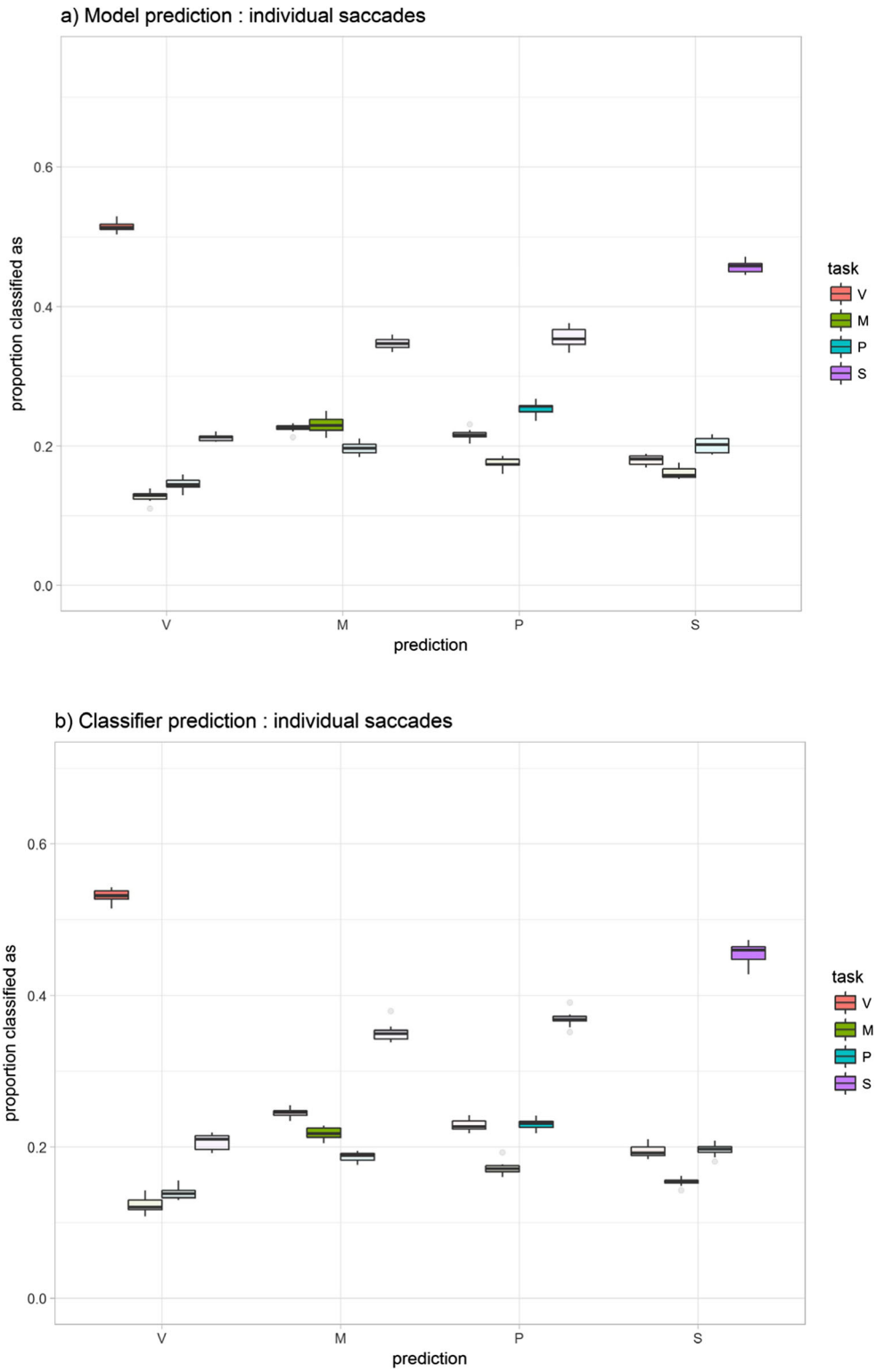

Fig. 8 Boxplot of the confusion matrix for the Dynamic Bayesian model (**a**). Saturated boxes are proportion of saccades correctly matching the generating task, while faded boxes are proportions that are misclassified as one of the other three tasks. For comparison, the classifier for individual saccades from "Section 1: Classifier" is shown (b)

While our model only considers top-down influences of saccade generation, it could be extended to include bottom-up generation as has been done with the Superior Colliculus [76], with input from a salience map [2] or priority map [21, 33]. For example, Corbetta and Shulman [77] suggest distinct but overlapping networks drive attention, with a temporo-parietal network driving bottom-up attention, which can interrupt activity in the frontal-parietal network associated with top-down attentional control, via a trigger in inferior frontal cortex. In the context of our DBN model, states could be

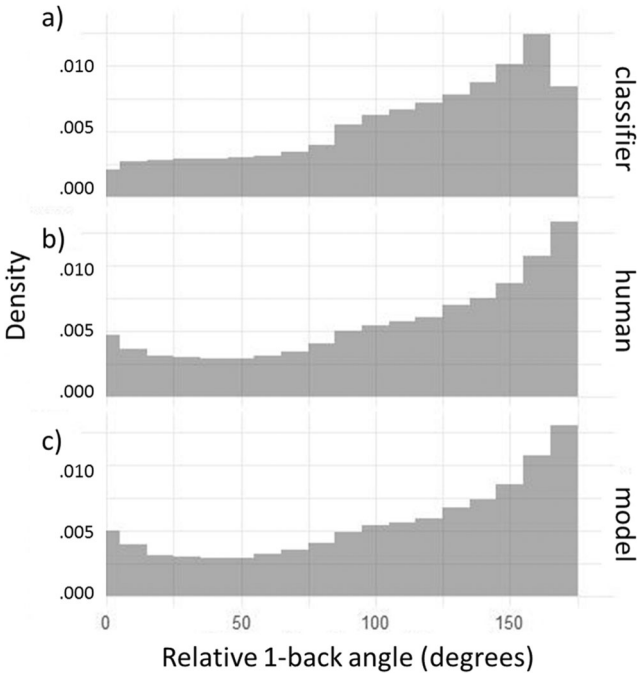

Fig. 9 Densities of relative saccadic direction for the **a** individual saccade classifier from "Section 1: Classifier", **b** observer data from the experiment, and **c** Dynamic Bayesian model from "Section 3: A Model of Task and Internal States." The model and classifier were trained only on absolute saccadic angle as a feature, yet the temporal aspect of the DBN captured sequence/pair information in the likelihood that each given saccadic angle at $t(x)$ would be followed by a saccadic angle at $t(x + 1)$

implementations of top-down attentional control, and low level information could provide information in the decision to switch states. It should be noted that attentional control and salience information might not be simply additive, as shown in Kardan [41].

Although Adaptive Gain Theory (AGT) [78] suggests that pupil size should correlate with internal cognitive state [79], our model was able to maintain an accurate representation of task and saccadic features without the use of pupil size as an additional variable. While models of foraging are improved by including pupil size and LG-NE [80], our pupil size was modeled as being influenced by cognitive state in the model to a similar, yet independent, extent as saccadic properties. It is possible that these saccadic properties contained redundant information making pupil size unnecessary. Our cognitive state did differ from AGT in that it consisted of five discrete stages of pupil size as opposed to a binary split of tonic and phasic mode [43, 78]. The original data for our study [49] were also not conducted under ideal conditions for detecting pupil size differences, so it is possible that pupil data could contribute more reliable state information when collected under conditions with better light control. Our final model was able to maintain the same classification accuracy from the data of a single saccade as the augmented Bayes Classifier in "Section 1: Classifier" while improving the overall information criteria for the full network and more accurately accounting for what we know of attentional state and cognitive control. It is also possible that pupil responses would best be modeled as a separate network from saccadic generation, although the network learning in the section-one classifier optimized pupil size as integrated with saccadic features.

We were also able to train a Bayesian classifier to recognize the instructed task given only saccadic attributes as input data. Two tasks, Search and View, were classified consistently above chance even when only given input from a single eye movement. The accuracy of predicting the Search task was expected since the original data set showed behavioral differences in generating observable Inhibition of Return [49] and other saccade properties [40]. Saccadic tendencies to continue forward and the increase in return saccades in particular become pronounced in the temporal DBN model where intersaccadic dependencies could be learned. Memorize was also classified above chance but only when mean eye movement data from the entire trial was included. The Preference task was only classified above chance when pooling data from all images and trials and did not generalize across subjects.

Given the generative nature of this model, it is capable of saccadic simulation in real time. We demonstrated that the properties from its generated saccades closely match those of human observers given a particular task and cognitive state. Future work is planned to include bottom-up influences as well as individual differences to separate the contribution of top-down cognitive state and the LC-NE system.

Funding Amelia Hunt has received research grants from the James S. MacDonnell foundation. Michael Dodd has received funding by NIH/NEI Grant 1R01EY022974.

Compliance with Ethical Standards

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Conflict of Interest All other authors declare that they have no conflict of interest.

References

1. İşcan Z, Özkaya Ö, & Dokur Z. Classification of EEG in a steady state visual evoked potential based brain computer interface experiment. In Proceedings of the 10th international conference on Adaptive and natural computing algorithms-Volume Part II (pp. 81–88). Springer-Verlag; 2011.
2. Carlson TA, Schrater P, He S. Patterns of activity in the categorical representations of objects. J Cogn Neurosci. 2003;15(5):704–17.

3. Borji A, Itti L. Defending Yarbus: Eye movements reveal observers' task. J Vis. 2014;14(3)
4. Cutsuridis V, Taylor JG. A cognitive control architecture for the perception–action cycle in robots and agents. Cogn Comput. 2013;5(3):383–95.
5. Schiller PH. The neural control of visually guided eye movements. In: Cognitive neuroscience of attention, ed Richards JE (Erlbaum, Mahway, NJ); 1998. p. 3–50.
6. Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci. 2001;2:194–203. https://doi.org/10.1038/35058500.
7. Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci. 2004;5:1–7.
8. Tatler BW, Hayhoe MM, Land MF, Ballard DH. Eye guidance in natural vision: reinterpreting salience. J Vis. 2011;11(5):5. https://doi.org/10.1167/11.5.5.
9. Mital PK, Smith TJ, Hill RL, Henderson JM. Clustering of gaze during dynamic scene viewing is predicted by motion. Cogn Comput. 2011;3(1):5–24.
10. Siebold A, van Zoest W, Donk M. Oculomotor evidence for top-down control following the initial saccade. PLoS One. 2011;6(9):e23552.
11. Tatler BW, Vincent BT. Systematic tendencies in scene viewing. J Eye Mov Res. 2008;2(2)
12. Clarke A, Tatler B. Deriving an appropriate baseline for describing fixation behavior. Vis Res. 2014;102:41–51, ISSN 0042-6989. https://doi.org/10.1016/j.visres.2014.06.016.
13. MacInnes W, Hunt A, Hilchey M, Klein R. Driving forces in free visual search: an ethology. Attention, Perception and Psychophysics. 2014;76(2):280–95.
14. Smith TJ, Henderson JM. Does oculomotor inhibition of return influence fixation probability during scene search? Attention, Perception, & Psychophysics. 2011;73(8):2384–98.
15. Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol. 1980;12(1):97–136.
16. Hinton GE. Learning multiple layers of representation. Trends Cogn Sci. 2007;11(10):428–34.
17. Krizhevsky A, Sutskever I, & Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097–1105); 2012.
18. Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci. 2007;104(15):6424–9.
19. Tu Z, Abel A, Zhang L, Luo B, Hussain A. A new spatio-temporal saliency-based video object segmentation. Cogn Comput. 2016;8(4):629–47.
20. Pang Y, Ye L, Li X, and Pan J, Moving object detection in video using saliency map and subspace learning, IEEE Transactions on Circuits Systems for Video Technology, https://doi.org/10.1109/TCSVT20162630731, 2016 (also arXiv:1509.09089).
21. Wischnewski M, Belardinelli A, Schneider WX, Steil JJ. Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. Cognitive Computation. 2010;2(4):326–34323.
22. Liu H, Yu Y, Sun F, Gu J. Visual–tactile fusion for object recognition. IEEE Trans Autom Sci Eng. 2017;14(2):996–1008.
23. Poria S, Cambria E, Howard N, Huang GB, Hussain A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing. 2016;174:50–9.
24. Golomb JD, Chun MM, Mazer JA. The native coordinate system of spatial attention is retinotopic. J Neurosci. 2008;28(42):10654–62.
25. Dorris MC, Pare M, Munoz DP. Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. J Neurosci. 1997;17(21):8566–79.
26. Aboudib A, Gripon V, Coppin G. A biologically inspired framework for visual information processing and an application on modeling bottom-up visual attention. Cogn Comput. 2016;8(6):1007–26.
27. Klein RM. Inhibition of return. Trends Cogn Sci. 2000;4(4):138–47.
28. Findlay JM, Brown V, Gilchrist ID. Saccade target selection in visual search: the effect of information from the previous fixation. Vis Res. 2001;41(1):87–95.
29. McPeek RM, Skavenski AA, Nakayama K. Concurrent processing of saccades in visual search. Vis Res. 2000;40(18):2499–516.
30. Klein RM, MacInnes WJ. Inhibition of return is a foraging facilitator in visual search. Psychol Sci. 1999;10(4):346–52.
31. Smith TJ, Henderson JM. Looking back at Waldo: oculomotor inhibition of return does not prevent return fixations. J Vis. 2011;11(1):3–3.
32. Fecteau JH, Bell AH, Munoz DP. Neural correlates of the automatic and goal-driven biases in orienting spatial attention. J Neurophysiol. 2004;92(3):1728–37.
33. Fecteau JH, Munoz DP. Salience, relevance, and spiking neurons: a priority map governs target selection. Trends Cogn Sci. 2006;10:382–90.
34. Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. J Exp Psychol Hum Percept Perform. 1992;18(4):1030–44.
35. Yarbus AL. Eye Movements and Vision, New York: Plenum. (Originally published in Russian 1965); 1967
36. DeAngelus M, Pelz J. Top-down control of eye movements: Yarbus revisited. Vis Cogn. 2009;17(6–7):790–811.
37. Ballard D, Hayhoe M, Pelz J. Memory representations in natural tasks. J Cogn Neurosci. 1995;7(1):66–80. https://doi.org/10.1162/jocn.1995.7.1.66.
38. Land M, Hayhoe M. In what ways do eye movements contribute to everyday activities? Vis Res. 2001;41(25–26):3559–65. https://doi.org/10.1016/S0042-6989(01)00102-X.
39. Castelhano MS, Mack ML, Henderson JM. Viewing task influences eye movement control during active scene perception. J Vis. 2009;9(3):6–6.
40. Mills M, Hollingworth A, Van der Stigchel S, Hoffman L, Dodd MD. Examining the influence of task set on eye movements and fixations. J Vis. 2011;11(8):17–17.
41. Kardan O, Henderson JM, Yourganov G, Berman MG. Observers' cognitive states modulate how visual inputs relate to gaze control. J Exp Psychol Hum Percept Perform. 2016;42(9):1429–42.
42. Macdonald JSP, Mathan S, Yeung N. Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. Front Psychol. 2011;2:82.
43. Aston-Jones G, Cohen J. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci. 2005;28:403–50.
44. Kotsiantis S, Zaharakis ID, Pintelas PE. Supervised machine learning: a review of classification techniques. Artif Intell Rev. 2007;26(3):159–90.
45. Greene M, Liu T, Wolfe J. Reconsidering Yarbus: a failure to predict observers' task from eye movement patterns. Vision research, 62, 1–8.Henderson J., Shinkareva, S., Wang, J., Luke, S. & Olejarczyk, J. PLOS One: Predicting Cognitive State from Eye Movements; 2012. p. 2013. https://doi.org/10.1371/journal.pone.0064937.
46. Henderson JM, Shinkareva SV, Wang J, Luke SG, Olejarczyk J. Predicting cognitive state from eye movements. PLoS One. 2013;8(5):e64937.
47. Marat S, Rahman A, Pellerin D, Guyader N, Houzet D. Improving visual saliency by adding 'face feature map'and 'center bias'. Cogn Comput. 2013;5(1):63–75.
48. Kootstra G, de Boer B, Schomaker LR. Predicting eye fixations on complex visual stimuli using local symmetry. Cogn Comput. 2011;3(1):223–40.

49. Dodd MD, Van der Stigchel S, Hollingworth A. Novelty is not always the best policy: inhibition of return and facilitation of return as a function of visual task. Psychol Sci. 2009;20:333–9.
50. Bahle B, Mills M, & Dodd MD. Human classifier: Observers can deduce task solely from eye movements. Attention, Perception, & Psychophysics. 2017; 1–11.
51. Borji A, Lennartz A, Pomplun M. What do eyes reveal about the mind?: algorithmic inference of search targets from fixations. Neurocomputing. 2015;149:788–99.
52. Hess EH, Polt JM. Pupil size as related to interest value of visual stimuli. Science. 1960;132:349–50.
53. Beatty J, Kahneman D. Pupillary changes in two memory tasks. Psychon Sci. 1966;5:371–2.
54. Kahneman D. Attention and effort. Engelwood Cliffs, NJ: Prentice Hall; 1973.
55. Laeng B, Ørbo M, Holmlund T, Miozzo M. Pupillary stroop effects. Cogn Process. 2011;12:13–21.
56. Gabay S, Pertzov Y, Henik A. Orienting of attention, pupil size, and the norepinephrine system. Atten Percept Psychophys. 2011;73(1): 123–9. https://doi.org/10.3758/s13414-010-0015-4.
57. Posner MI, Fan J. Attention as an organ system. In: Pomerantz JR, editor. Topics in integrative neuroscience: from cells to cognition. 1st ed. Cambridge: Cambridge University Press; 2008. p. 31–61.
58. Rajkowski J, Kubiak P, Aston-Jones G. Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. Soc Neurosci Abstr. 1993;19:974.
59. Rajkowski J, Majczynski H, Clayton E, Aston-Jones G. Activation of monkey locus coeruleus neurons varies with difficulty and performance in a target detection task. J Neurophysiol. 2004;92:361–71.
60. Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern. 1973;3(3):32–57. https://doi.org/10.1080/01969727308546046.
61. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett. 2010;31(8):651–66.
62. Vincent BT. Bayesian accounts of covert selective attention: a tutorial review. Atten Percept Psychophys. 2015;77(4):1013–32.
63. Druzdzel MJ. SMILE: structural modeling, inference, and learning engine and GeNIe: a development environment for graphical decision-theoretic models. In: Aaai/Iaai; 1999, July. p. 902–3.
64. Moon TK. The expectation-maximization algorithm. IEEE Signal Process Mag. 1996;13(6):47–60.
65. Kardan O, Berman MG, Yourganov G, Schmidt J, Henderson JM. Classifying mental states from eye movements during scene viewing. J Exp Psychol Hum Percept Perform. 2015;41(6):1502–14.
66. Fishel J, Loeb G. Bayesian exploration for intelligent identification of textures. Front Neurorobotics. 2012;18 https://doi.org/10.3389/fnbot.2012.00004.
67. Murphy KP. Dynamic bayesian networks: representation, inference and learning. University of California, Berkeley: Doctoral dissertation; 2002.
68. Bylinskii, Z., Judd, T., Borji, A., Itti, L., Durand, F., Oliva, A., & Torralba, A. (2015). Mit saliency benchmark.
69. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001;24:167–202.
70. Akaike H. A new look at the statistical model identification. Automatic Control, IEEE Transactions on. 1974;19(6):716–23.
71. Posner M, Dehaene S. Attentional networks. Trends Neurosci. 1994;17:75–9.
72. Banich M, Milham M, Atchley R, Cohen N, Webb A, Wszalek T, et al. Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. Cogn Brain Res. 2000;10(1–2):1–9, ISSN 0926-6410. https://doi.org/10.1016/S0926-6410(00)00015-X.
73. Tanner J, Itti L. Goal relevance as a quantitative model of human task relevance. Psychol Rev. 2017;124(2):168–78.
74. Hanes DP, Wurtz RH. Interaction of the frontal eye field and superior colliculus for saccade generation. J Neurophysiol. 2001;85(2): 804–15.
75. Bruce CJ, Goldberg ME. Primate frontal eye fields. I. Single neurons discharging before saccades. J Neurophysiol. 1985;53(3):603–35.
76. Trappenberg T, Dorris M, Munoz D, Klein R. A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. J Cogn Neurosci. 2001;13(2):256–71.
77. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 2002;3:201–15.
78. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci. 2010;10(2):252–69.
79. Joshi S, Li Y, Kalwani RM, Gold JI. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron. 2016;89(1):221–34.
80. Barack DL, & Platt ML. Engaging and Exploring: Cortical Circuits for Adaptive Foraging Decisions. In *Impulsivity* (pp. 163–199). Springer International Publishing. 2017