PLOS ONE

RESEARCH ARTICLE

LAG-1: A dynamic, integrative model of learning, attention, and gaze

Jordan Barnes[1], **Mark R. Blair**[1]*, **R. Calen Walshe**[2], **Paul F. Tupper**[3]

1 Department of Psychology, Simon Fraser University, Burnaby, BC, Canada, **2** Center for Perceptual Systems, University of Texas, Austin, Texas, United States of America, **3** Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada

* mblair@sfu.ca

Check for updates

OPEN ACCESS

Citation: Barnes J, Blair MR, Walshe RC, Tupper PF (2022) LAG-1: A dynamic, integrative model of learning, attention, and gaze. PLoS ONE 17(3): e0259511. https://doi.org/10.1371/journal.pone.0259511

Editor: David Keisuke Sewell, The University of Queensland, AUSTRALIA

Received: March 27, 2021

Accepted: October 21, 2021

Published: March 17, 2022

Copyright: © 2022 Barnes et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supplementary Information files.

Funding: Author: J.B. Grant: (GXSO116) Funder: National Science and Engineering Research Council (NSERC) Funder role? No. Author: M.B. Grant: (327301) Funder: National Science and Engineering Research Council (NSERC) Funder role? No. Author: P.F.T. Grant: (RGPIN-2019-06911) Funder: National Science and Engineering Research Council (NSERC) Funder role? No.

Abstract

It is clear that learning and attention interact, but it is an ongoing challenge to integrate their psychological and neurophysiological descriptions. Here we introduce LAG-1, a dynamic neural field model of learning, attention and gaze, that we fit to human learning and eye-movement data from two category learning experiments. LAG-1 comprises three control systems: one for visuospatial attention, one for saccadic timing and control, and one for category learning. The model is able to extract a kind of information gain from pairwise differences in simple associations between visual features and categories. Providing this gain as a reentrant signal with bottom-up visual information, and in top-down spatial priority, appropriately influences the initiation of saccades. LAG-1 provides a moment-by-moment simulation of the interactions of learning and gaze, and thus simultaneously produces phenomena on many timescales, from the duration of saccades and gaze fixations, to the response times for trials, to the slow optimization of attention toward task relevant information across a whole experiment. With only three free parameters (learning rate, trial impatience, and fixation impatience) LAG-1 produces qualitatively correct fits for learning, behavioural timing and eye movement measures, and also for previously unmodelled empirical phenomena (e.g., fixation orders showing stimulus-specific attention, and decreasing fixation counts during feedback). Because LAG-1 is built to capture attention and gaze generally, we demonstrate how it can be applied to other phenomena of visual cognition such as the free viewing of visual stimuli, visual search, and covert attention.

Introduction

An important part of learning is learning what's important. Experts can be distinguished from novices based on the sources of information they use to make classifications. For example, expert bird watchers use subtle but predictive features of birdsong to distinguish between different bird species, identify birds using their specific names, and describe birds using behavioural traits, whereas novices do not [1, 2]. Analogous findings about the features used by experts and novices to make classifications have been found in biology [3, 4], physics [5], and computer programming [6], among many other areas. Knowledge about what's important

Competing interests: The authors have declared that no competing interests exist.

guides where people look. In reading, for example, many indicators of a new, or struggling, reader are obvious in their eye movements: excess fixations, fixations to previous words, or to determiners like "the" [7]. Similarly, novice drivers tend to look more at the road ahead and underutilize other important sources of information, relative to more experienced drivers [8]. Laboratory studies looking at eye movements during learning tend to show that the overall number of gaze fixations, the fixations to irrelevant information, and the durations of fixations decrease as skills are increased [9–11]. Learning influences the allocation of gaze and attention, but the reverse is also true. Providing novices with attentional instruction can speedily close their performance gap with experts; for example, in a classic study of chicken sexing—a notoriously difficult perceptual challenge—researchers were able to markedly improve novice performance by training them how to prioritize the most informative parts [12].

One paradigm that is excellent for eliciting the interactions of learning and attention is category learning [13–18]. Experiments with this paradigm are simple enough to be modelled, yet constitute a full sweep through the cognitive system, including: perception, attention, memory, and decision making, all packed into hundreds of incremental learning trials occurring over cognitively distinguished time scales [19, 20]. In a typical category learning study, the stimulus that is presented to the participant is a category exemplar composed of 3–4 visual features. Participants then choose one of the possible categories (usually between 2–4 categories). At first, participants are just guessing: they do not yet know which features are indicative of which categories. Each successive trial presents a new stimulus, solicits a response, and typically provides corrective feedback, enabling participants, over hundreds of trials, to learn how to categorize accurately. Usually, some of the features are more relevant than others for predicting the correct category, leading participants to eventually selectively attend to those features.

Category learning researchers have measured attention using both indirect methods—like with knowledge transfer tasks, wherein a research participant is required to categorize unseen exemplars after experience with a training set, and thus reveal what they deemed important (e.g. [21])—and direct methods, that require participants to explicitly reveal stimulus information using mouse movements or clicks for example [22–25]. Probably the most natural way of capturing the deployment of attention during category learning, however, is using eye tracking. Many general eye tracking findings have now been documented (see [26] for a summary of measures across 10 different experiments). Perhaps the simplest and most common finding is that the number of eye movements to stimulus features in a categorization trial starts high and drops over the course of an experiment. In most cases, it settles to just above the number of features necessary to consistently generate a correct classification of the stimulus [10, 27, 28]. The duration of individual fixations to features has also been shown to decrease as participants learn [29]. Participants spend more of their time looking at relevant features than irrelevant ones [10, 11], and they fixate important features earlier in a trial [10, 11, 25, 29]. Finally, the amount of time that participants spend viewing features during feedback drops as the participants learn how to more efficiently attend in the task [30]. In all of these findings, behavioural measures of attention are strongly related to behavioral measures of learning.

The full set of findings relating to learning and attention in category learning is challengingly diverse. Participants are choosing what features to look at and for how long, and in what order. They are deciding when they know enough to choose a category and which category is correct. They are then choosing if, when, and how long to look at the feedback and represented stimulus features, and when to move on to the next trial. Participant's sampling of information is continuous, and intertwined with the learning process itself. Further, behaviour adapts at a wide variety of timescales: saccades operate on the order of 100 ms, fixations about 200–400 ms, task length fixation ordering 1000–5000 ms, and learning-related changes reveal themselves over whole experiments lasting 180,000 ms. We know of no existing model which links

Table 1. Eye tracking, timing and category learning findings.

Finding	Empirical source	Model
Reaction time reduction	Homa and Fish (1975), McColeman et al. (2014) [26, 31].	Lamberts (1998), Logan (2002), Nosofsky and Palmeri (1997) [32–34].
Fixation count reduction	McColeman et al. (2014), Rehder and Hoffman (2005) [10, 26].	Barnes, McColeman, Blair, and Walshe (2014), Nelson and Cottrell (2007) [28, 35].
Fixation duration reduction	Blair, Watson, Walshe, and Maj (2009), McColeman et al. (2014) [11, 26].	N/A
Fixation ordering	Blair, Watson, Walshe, and Maj (2009), Chen, Meier, Blair, Watson, and Wood (2013), Rehder and Hoffman (2005) [10, 11, 29].	Rombouts, Bohte, Martinez-Trujillo, and Roelfsema (2015) [36].
Reduced feedback use	Bourne, Guy, Dodd, and Justesen (1965), Watson and Blair (2008) [30, 37].	N/A

N/A indicates findings that are modelled for the first time here.

https://doi.org/10.1371/journal.pone.0259511.t001

learning with attention and gaze that can make quantitative predictions about the full set of findings relating learning, and gaze in these tasks. This is in no small part because no one has tried. Instead, researchers have chosen to carve off smaller portions of the problem, trading generality for tractability. Some examples of findings and their corresponding models can be found in Table 1. To us, these phenomena seem to naturally arise from the interaction of a simple learning mechanism, a simple attentional priority map, and a simple saccade timing system.

LAG-1

In the present paper, we introduce LAG-1, a computational model of learning, attention and gaze designed to process common experimental manipulations in category learning tasks. The name emphasizes the three cognitive components we are attempting to combine. It also emphasizes the temporal nature of the integration of these processes: learning, attention and gaze, moment by moment. In this section, we describe the model in detail. We start with a high level description of its core theoretical commitments, and then explain how its structure relates to both the theoretical claim and to its behavioural predictions by showing how activation flows through the various components of the model. Finally, for those looking for exact implementational details and correspondences with neurophysiology, we provide some description of the neurophysiological and behavioural studies relevant to each component of the model and provide the equations for each component in the supplementary information.

At the heart of the model is a theoretical idea about how learning, attention, and gaze interact: simple associative learning affords an additional source of information about the importance of features beyond just the base-rate driven weights: later in learning, the contrast of these weights also tells you how diagnostic the features are. Measures of expected information gain are commonly employed to model visual attention (e.g., [38]), where this sort of quantity is integrated with bottom-up visual signals in an attentional priority map (e.g., [39]), that strongly influences saccade initiation [40]. According to LAG-1, these linkages between learning and attention, and between attention and saccadic initiation are the primary cause of the documented covariation between learning and gaze seen in the myriad measures of participant behaviour in the category learning paradigm. Thus, implementing these ideas in a

computational framework should allow us to predict, at least qualitatively, the changes to gaze that correspond with the changes in participant performance.

To test LAG-1, we simulate two category learning experiments that also recorded eye gaze [10, 25]. The empirical findings relevant to these simulations are listed in Table 1. Because the central phenomenon of interest to LAG-1 is the interrelationship between attention and learning, we report learning performance, via category responses and stimulus generalization. There are also a number of generic timing measures that can be fit, such as how long participants spend before making a categorization response, or how long they spend viewing feedback before continuing on to the next trial.

Implementing a model that is continuous in space and time and also neurally plausible—one of our chief design considerations—requires a suitable modelling framework. We implement LAG-1 using ideas from Dynamic Neural Field Theory (DNFT). This framework has an extensive track record of model-based explanations in developmental [41–44] and cognitive psychology [45–47]. Neural fields are well-suited to modelling the brain's numerous topographic maps, like those implicated in the making of eye movements [48, 49]. For example, the characteristic properties of dynamic neural fields have been used to explain the critical distances inherent to averaging saccades by modelling the functional relationships of the superior colliculus [46]. DNFT models have also provided important insights into the temporal dynamics of cognitive processes like object recognition, by naturally reflecting the time needed to resolve competition proportionate to the multidimensional metric confusability of object features (the "Lyupanov time" [50]), e.g., distinguishing a lime from a lemon takes longer than distinguishing a lime from banana) [51]. We note that, while the logic of modelling with DNFT is compelling, there are alternatives, such as simplifying the lateral influences of spatial configuration using discrete dynamical units that still preserve the temporal dynamics of interest, e.g., [52]. By building LAG-1 using DNFT however, the underlying structure of the model naturally accommodates any temporal and spatial aspects of cognitive processes in ways that match the computational properties of real neural systems.

We designed LAG-1 primarily to process visual category learning experiments, but very few changes were needed in order to have it process another important class of attention tasks: visual search. The benchmark measures for visual search include very different kinds of phenomena than in category learning, such as the distractor-heterogeneity effect, and the feature-target similarity effect, as well as several other findings which have been well documented elsewhere, e.g., [53]. It is noteworthy that temporal models of visual search are accounting for benchmark phenomena by allowing knowledge of the critical features of a target to amplify those particular feature dimensions as search items compete with one another for attention [54, 55]: attention weights have a long history in category learning models, and this suggests that progress in one domain may naturally complement the other.

As there are by now many dynamical cognitive process models of attention and learning, it is encouraging to see models try to build on previous efforts, especially when it comes to identifying neural processes that collectively act as a functional subsystem untethered from a specific model. For example, [56] builds on the influential category learning model "COmpetition between Verbal and Implicit Systems" (COVIS) [57], by replacing the original visual processing methods of the model with more precise models of cortical and subcortical structures at each level of the visual processing hierarchy. By exploiting the modularity of these visual processes the extended model increased the category learning phenomena that could in principle be accounted for, such as: the effect of feedback delays, or dimensional relevance shifts. LAG-1 similarly aims to be "plug-n-play", at the level of neurophysiologically plausible subsystems, where for example, the comparatively simple associative learning subsystem in LAG-1 could

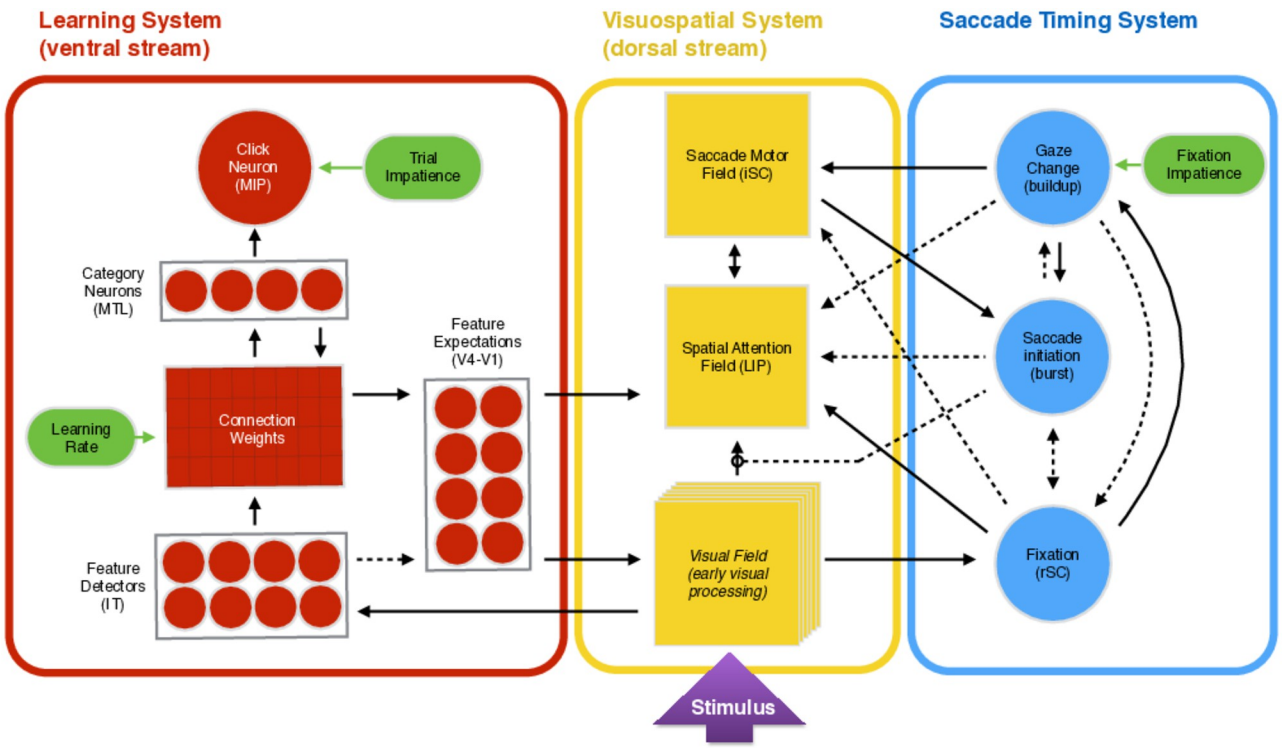


Fig 1. Model structure. The major components of LAG-1 and the excitatory (solid) and inhibitory (dashed) connections between them. There are three functional subsystems: Visuospatial Processing, Learning, and Saccade Timing. Labels in brackets refer to neuroanatomical regions that have similar functions. Green ovals represent the free parameters we varied to fit individual subjects. Circles represent single model neurons, rectangles denote fields, and circles within rectangles specify single neurons with some level of lateral interaction.

https://doi.org/10.1371/journal.pone.0259511.g001

be swapped out or extended by a more complex feedback-driven learning system capable of learning non-linear category structures.

LAG-1 is thus part of a wider group of dynamic models of cognition that are taking on the challenge of capturing the temporal nature of cognitive processes by directly modelling neural processes at the level of sensorimotor interactions between learning and attention.

The simplest illustration of the full LAG-1 model, and one that may be sufficient for readers not interested in the mathematical aspects of the work, is shown in Fig 1. We embedded the primary theoretical ideas in a model of how the three relevant subsystems (learning, attention, and gaze) might interact. LAG-1 actively samples the visual field, forms associations between categories and features values, and initiates actions influenced by both the sampling and the learning. The labels and connections shown in the figure are meant to broadly conform to what is known about the structure and function of the relevant aspects of the human brain.

Information flow through LAG-1

In the category learning experiments modelled, the task comprises several hundred learning trials, each of which has two phases: a response phase and a feedback phase. In the response phase, a fixation cross is first shown in order to center the participant's gaze, after which a stimulus is presented. The participant then views the features of the stimulus, and makes a response by clicking a button associated with one of the categories. The category response then initiates the feedback phase wherein both the same stimulus and the correct category are

presented. When the participant is ready to start the next trial, they click a button, and the fixation cross for the next trial appears. Both of the phases of the trial are self-paced: the stimulus and the correct category can be observed for as long as the subject (or the model) wishes.

The way that LAG-1 processes information during the experiment is depicted in Fig 1. At the start of each trial a stimulus is presented to the model as input to the Visual Field. Stimulus features have particular values, coupled to that particular location on the screen, and the model, because it represents everything spatially, is sensitive to the locations associated with particular combinations of feature values. As spatial competition for attention gets resolved by the lateral dynamics of local-excitation and global-inhibition, gaze will shift to a saccade threshold-crossing peripheral location of the Visual Field, and extract the value of the feature there. This information then passes into the Learning System, activating the appropriate Feature Detectors whose self-sustaining activation acts as a kind of visual working memory. Feature Detectors then propagate their activation through a weight matrix and into the Category Neurons. Recurrent connections from the Category Neurons (propagating backward through a copy of the same weight matrix) pass activation to Feature Expectation Neurons that are associated with those active categories. A critical feature of the system is that activation of these neurons is modulated by a gain mechanism that gives additional boosts of activation to features based on their expected information gain. The calculation of gain is explained in detail in the "Category Neurons" section of the Learning System overview. This top-down information is integrated with both bottom-up information propagating in from the Visual Field, and directly on to the Spatial Attention Field, such that active categories are driving attentional choices, and not just adapting what is salient [39]. Activity on the Spatial Attention Field passes into the Saccade Motor Field, which, because of an inhibitory signal at the current fixation location, can be thought of as a map of possible saccade targets. The pressure to initiate a saccade (and thus target the peak location in the Saccade Motor Field) is controlled by a trio of neurons, and builds up slowly from the beginning of a fixation. Strong activation at the location currently activated can delay saccades, and strong activation at potential saccade target locations can speed them up. Because correct categorization requires information about the value of more than a single feature in these experiments, and because only one feature can be fixated at a time, the model must make a series of eye movements, as humans must, in order to gather the information necessary to make a correct classification. As the model views the features of the stimulus, information is sustained by both the Feature Detectors and the Category Neurons, until activation of a motor response crosses a fixed threshold (the Decision Click Neuron); in the human experiments this would be like a mouse click or button press response.

Once a category response has been chosen, the feedback phase of the trial begins. The correct category enters the Visuospatial Processing system as a button that appears in the periphery of the Visual Field. If viewed, a boosting signal excites the Category Neuron corresponding to the correct category. LAG-1 may continue to look at features of the stimulus, boosting their activation among the Feature Detectors, and countering any memory decay for those values that may have occurred since having last looked at them. During the feedback phase the connections between active features and active categories are strengthened according to simple Hebbian learning. The longer LAG-1 studies the feedback, the more the synaptic weights connecting the active features and categories are strengthened. Once the Click Decision Neuron has again exceeded threshold, the next trial is initiated.

As changes to the associative weights between the Feature Detector Neurons and the Category Neurons accumulate over learning trials, the dynamics of the system begin to change. The strength of an association, reflecting base rates, and the category selectivity of Feature Detectors, reflecting information gain, work together to modify top-down attention. These top-down signals decrease the chance that irrelevant information will be fixated, as well as

decreasing fixation durations. Increased activation of the Category Neurons, and faster accessing of the useful information leads to faster reaction times. Increased activation of the Category Neurons also leads to less time being spent during feedback. LAG-1 thus reflects the idea that learning influences attention and attention influences gaze, as well as the inverse relation: what is viewed, and for how long, changes what is learned.

LAG-1 has only three fitted individual subject parameters—shown in green in Fig 1: Learning Rate, Fixation Impatience, and Trial Impatience. The Learning Rate modulates how rapidly the connections between Feature Detectors and categories change per unit time. The Fixation Impatience parameter modulates the growth rate of the pressure to initiate a saccade, which is known to be dynamically adjusted to improve reward rate (c.f., [58, 59]), and the Trial Impatience similarly controls the growth rate of the pressure to make a response (i.e., either by clicking with a category response, and thus initiating the feedback phase, or, during the feedback phase, by clicking to initiate the next trial). These parameters have the effect of preventing the model from perseverating on a difficult decision and ensuring a level of global stability. These impatience parameters could be thought of similarly to "urgency signals" found in the speeded choice literature (e.g., [60, 61]): the idea being to keep the system moving by applying a strong compulsion to move the eye or make a decision after a consistent amount of time has elapsed.

Detailed description of LAG-1 and neurophysiologial context

In this section we provide the formal description of the model, as well as identify research relevant to its structure and functioning. The model structure, illustrated in Fig 1, can be a useful map to the three subsystems, their component parts, and can even help one parse the various model equations (e.g., it can help to compare the inputs in the equation to the inputs shown in the figure). Readers that are less interested in the computational details may still wish to read the overviews provided for the three subsystems to get a feel for how they work. For readers who are interested in the technical details but may be unfamiliar with, or could benefit from a refresher on, the differential equations that form the foundation of the model, we have produced a detailed primer that can be found in the supplementary information: "S1 Appendix: A Primer to Dynamic Neural Field Theory".

In the supplementary information titled "S2 Appendix: Companion equations for the formal description of LAG-1and neurophysiologial context.", we provide the equations for each of the components described at a high-level here. Wherever possible we have attempted to give model parameters intuitive labels—for example, the Fixation Neuron is labeled u_x, the Gaze Change Neuron is u_g, and the Feature Detection Neurons are $u_{\mathrm{ft}_{\mathrm{det}}}(j, t)$. The output of each neuron and field is bounded by a sigmoid function. Values having been transformed by a sigmoid function f are indicated with a *, e.g., $u^* = f(u)$. More information about this is, provided in the supplementary information S1 Appendix. Finally, references to structural components of the model are capitalized, such as the Visual Field, as opposed to generically talking about the visual field of view.

Modelling complex non-linear systems like LAG-1 requires that the correct balance be found between the various forces affecting the activity of a neuron or field. In LAG-1 we use u_c to indicate a simple scaling parameter. There are many of these throughout the model equations as they are a convenient way to adjust the relative magnitudes between forces at play in the equation. That being said, once fields/neurons were stabilized, these scaling parameters were not changed. That is, u_c parameters were not adjusted at all during the data fitting process. We think of these parameters as a mathematical (and practical) necessity, but not critical for understanding the work as a theory; for this reason they are not discussed in detail at each

occurrence. Values for scaling parameters used in any model equations throughout this paper can be found in the supplementary information titled "S5 Appendix: Parameter tables and best fits."

Visuospatial Processing system overview. Experiment input to LAG-1 is represented as a stack of two dimensional layers having the same size as the Visual Field. The location of a red feature, for example, is represented as a set of 1s among an array of zeros at the appropriate spatial location in the layer that represent red input. LAG-1 extracts information from this input in two places, the Visual Field, which codes the spatial location of all elements in the stimulus regardless of color value, and the Feature Detection Neurons, which code the feature values for any feature that is currently fixated. On the Visual Field, information about the stimulus is coded in retinotopic coordinates, where the fovea centers the frame of reference. At this early stage of processing the featural dimensions of a stimulus are not intrinsically bound with location. Similar to the "what"/"where", ventral/dorsal, processing streams observed in the brain, LAG-1 must use "complementary processes" that have hierarchical and laminar components to integrate different dimensions of the input into a sensible package [62, 63].

Spatial information from the Visual Field projects into the Spatial Attention Field of LAG-1 where spatial competition for attentional priority resolves. Priority maps in the parietal cortex are known to predict movement behaviours given a wide range of considerations such as goals, reward history, novelty, as well as category decisions and information gain [39]. Parietal priority maps connect those ideas to actual choices to do things in the real world. Activity on the Spatial Attention Field can be thought of as both selective and predictive of eye movements. Where area LIP projects to the SC, and through a thalamic route receives input from the SC, the Spatial Attention Field and Saccade Motor Field in LAG-1 also have bidirectional relationships. In both the model and the brain, these structures are essential to relating overt and covert attentional dynamics [64].

Converting representations between different frames of reference can be accomplished with the aid of a movement shift operator or coupling fields that represent information in differing frames of reference. An earlier DNFT model of saccadic remapping during multiple-object-tracking (MOT), offers a precise approach to the mechanics of reference frame transformations between neural fields [65]. In the present version of LAG-1 we simplify things by directly coding a transformation operator between the spatiotopic frames of reference used by the Spatial Attenion and Saccade Motor Fields, and the retinotopic coordinates of the Visual Field.

Visual Field. The Visual Field in LAG-1 is a retinotopic map that represents the locations of features relative to the fovea. The center of the Visual Field defined in Equation 11 is always the fovea, while the activations of the Visual Field are translated in space to reflect their change in distance from the center as the eye moves. Experiment stimulus inputs provided to the Visual Field are transformed by Equation 12 The processes comprised by the Visual Field and the input stimulus transformation are an abstraction of the kind of early visual processes that begin at the retina and culminate at the psychologically interpretable visual categories represented by complex neurons in V4 and IT [66].

Spatial Attention Field. Turning knowledge into behaviour at the right place, and at the right time, is a large part of the functions of the priority maps in the parietal cortex [39, 67]. The posterior parietal region of the parietal cortex coordinates a number of motor actions like reaching, grabbing and saccade targeting [68]. Neurons in lateral intraparietal area (LIP) have been shown to vary systematically with the location [69], category [70], reward [71], posterior likelihood [38, 72] and task relevance [73, 74] of individual eye movements. This makes LIP a very important source of attention related projections to more proximate motor control of the eye like the superior colliculus [75, 76].

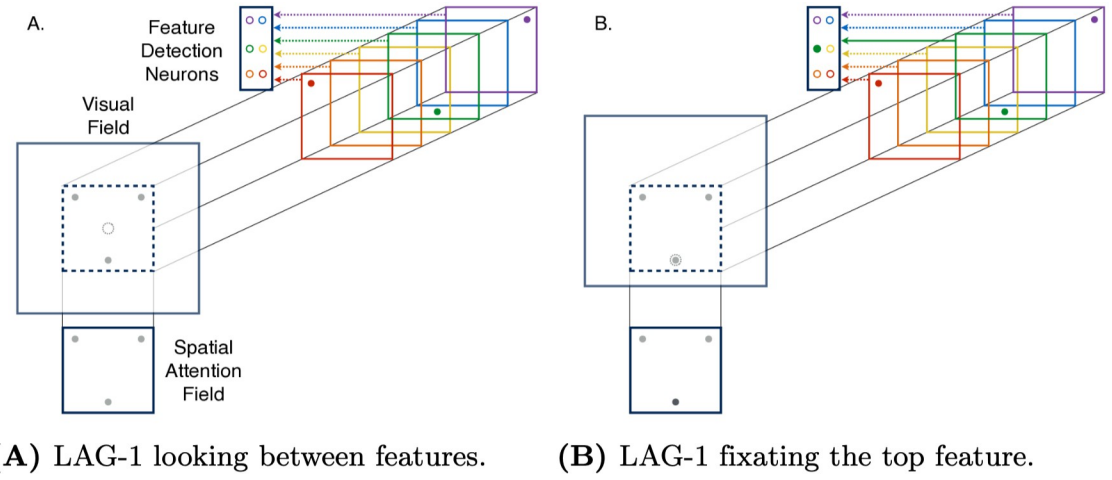


Fig 2. Information processing schema. Schematic of the relationships between experiment input, the Feature Detection Neurons, Visual Field, and the Spatial Attention Field. A) The fovea is indicated by the dashed grey circle in the center of the Visual Field. B) After an eye movement to the bottom middle feature, the green sensitive Feature Detector is activated and the Spatial Attention Field is boosted at its associated location. The Visual Field is coding spatial information in retinotopic coordinates while the Spatial Attention Field is coding information in spatiotopic coordinates.

https://doi.org/10.1371/journal.pone.0259511.g002

The Spatial Attention Field in LAG-1 represents changes in attentional priority to locations in a spatiotopic reference frame. The different frames of reference of the Spatial Attention Field and the Visual Field are depicted in Fig 2. Notice that there is no change in the location of features on the Spatial Attention Field between Fig 2A and 2B, despite the shifting of the bottom feature onto the fovea of the Visual Field.

Changes in activation on the Spatial Attention Field are described by similar equations as those used for the Visual Field, in Equation 13.

Saccade Motor Field. Candidate saccade target locations compete on the Saccade Motor Field [48, 49]. Unlike the Spatial Attention Field which binds a retinue of competing attentional priorities with spatial locations, the Saccade Motor Field resolves competition for attention at locations *other than* the current locus of fixation. The dynamics of this field resolves the competition between locations to be the target of the next eye movement according to Equation 17.

Saccade Timing System overview. Under normal viewing conditions, humans make about three saccades per second [77]. The period of relative spatial stability between saccades is the fixation duration. The parameterizations used in the present simulations were chosen for their rough correspondence with the normal range of saccade and fixation durations. In the model, as in the brain, the timing of an eye movement is affected by previous experience. In people, this includes factors like expected processing difficulty, but category learning experiments are designed to minimize such effects, so changes in fixation duration exhibited by LAG-1 are primarily the result of learning the relevance of different features. We have also observed fixation duration differences in LAG-1 when using it for visual search where disorganized inputs yield spatial interactions that speed or slow its saccade onset latency [40, 46, 58, 67, 78].

As the name suggests, the Saccade Timing System is the primary arbiter of decisions to release fixation and foveate a new location. The trio of neurons controlling this system are: the Gaze Change Neuron, the Fixation Neuron, and the Saccade Initiation Neuron. These model neurons have a functional correspondence with brain stem neurons referred to as: "build-up

neurons", that increase their firing rate over the course of saccade preparation; "fixation neurons", that remain tonically active during a fixation; and "burst neurons", that spike just prior to a saccade [79–81]. Again, a saccade is initiated by LAG-1 when the Saccade Initiation Neuron crosses a threshold. When the eye is at rest, the Fixation Neuron inhibits this Saccade Initiation Neuron, and the Gaze Change Neuron increases the pressure to initiate a saccade as the duration of a fixation goes grows.

The Saccade Timing System is influenced by three exogenous inputs. First, the Fixation Impatience parameter influences how rapidly the activation of the Gaze Change Neuron grows, and reflects stable individual differences in fixation durations. That such a parameter could vary systematically between individuals is motivated by timing differences in gaze behaviours between individuals. Individuals prone to making longer fixations will do so across numerous kinds of tasks and this trait covaries with other gaze measures, such as the typical length of their saccades [82, 83]. Second, the Fixation Neuron receives inputs from the Visual Field when a feature is being fixated. Finally learning induced increases in activation for relevant features on the Spatial Attention Field and Saccade Motor Fields push the Saccade Initiation Neuron over threshold faster.

Gaze Change Neuron. The Gaze Change Neuron plays a critical role in compelling LAG-1 to explore the visual environment. The behaviour of this neuron, defined by Equation 19, has notable similarities with build-up neurons found in caudal regions of the intermediate layers of the superior colliculus [49]. The Fixation Impatience parameter, λ_{P2}, increases the influence of these requests with every time step when $\lambda_{P2} > 1$. Putting the bulk of a saccade initiation decision on to a preprogrammed timer makes these decisions more resilient to noise and moment-to-moment attentional capture. In the real world people can dynamically adjust such timers over the course of a few trials based on factors like the expected processing difficulty or the inter-trial interval [58, 59].

Fixation Neuron. The Fixation Neuron, defined in Equation 20, suppresses the impetus to move the eye. It accomplishes this by exciting the foveal location of the Spatial Attention Field, thereby damping extrafoveal locations of the field by way of a stronger global inhibition projection to regions other than the boosted fovea. The only input to this neuron is the sum of featural information within the foveally masked location of the Visual Field. Anatomically, fixation neurons are observed to project tonic inhibition to saccadic burst neurons in the time between saccades, such that inhibiting them allows saccades to be programmed and executed faster [84, 85].

Saccade Initiation Neuron. A saccade is initiated by LAG-1 when the Saccade Initiation Neuron reaches threshold, according to Equation 21. This neuron has functional similarities with saccadic burst neurons, which show phasic activity in the moments just prior to a saccade [79].

Learning system overview

LAG-1 learns during the feedback phase of every trial by Hebbian modification of the weights connecting active Feature Detection Neurons with Category Neurons activated by fixating a correct category presented on the screen. If an incorrect guess was made, anti-Hebbian learning simultaneously decouples the active Feature Detection Neurons and Category Neurons. For example, if a blue, yellow and green set of Feature Detection Neurons is co-active with category A during the feedback phase, applying the learning rule will increase the particular weight that connects the presynaptic Feature Detection Neuron and the correct Category Neuron, proportionate to the current strength of the weight. If the correct category was not A in this example, but B, then the blue, yellow, and green Feature Detection Neurons activated by

fixating these features, would reduce their connection strength to A, and increase it to B during the feedback phase.

Attending to feedback signals is an essential part of category learning [30, 37]. The category learning experiments we simulate are self-paced: participants click a response button associated with a particular category which brings up the feedback display, and when they are ready they click again to advance to the next trial. Participants can, if they choose, rapidly click to advance to the next trial without spending time looking at the feedback or stimulus. Human participants who have mastered the categories often choose to do this in order to finish faster —LAG-1 too, has this option. The decision to click and advance the experiment to the next phase is represented by a Click Decision Neuron having two thresholds. Crossing the first threshold moves the trial from the response phase to the feedback phase, crossing the second threshold ends the trial.

At the start of the experiment, when LAG-1 knows nothing about the category and feature relationships, impatience and experiment feedback signals provide the bulk of the input needed for the Click Decision Neuron to cross its thresholds. As knowledge about the category structure develops, this knowledge will result in higher activation of particular Category Neurons, which, when combined with experiment feedback signals, make for faster reaction times in the feedback phase. When the category associations are very strong, feedback signals may be unnecessary for crossing the second threshold, and thus, LAG-1, like the human participants, can stop looking at feedback altogether when it has mastered the categories.

In addition to registering the physical attributes of the features being looked at, Feature Detection Neurons act like a visual working memory in that they can sustain their activation even after looking elsewhere. Each of these neurons is connected to the others, as well as to a set of Category Neurons that do not themselves have a sensory connection to the experiment stimulus except through the the Feature Detection Neurons.

Activity propagates from the Category Neurons into the Click Decision Neuron, as well as feeding back through a mirror of the weights that connect the Feature Detection Neurons with the Category Neurons, to activate the Feature Expectation Neurons. These neurons mirror the Feature Detection Neurons in number but project back toward the spatial locations of the Visual and Spatial Attention Fields where the feature values are known to be associated. This is how LAG-1 allows feature-based attention to influence the selection of eye movements [86, 87].

At the top of the visual attention hierarchy in the brain are category sensitive neurons in areas like inferotemporal (IT) and medial temporal lobe (MTL) which get activated by ongoing sensory combinations of visual input and which have reentrant projections to successive layers of the visual cortex, and categorical projections to LIP respectively. Neurons in these regions effectively sensitize visual neurons according to category expectations [70, 88–92]. Neurons in these regions are reported to have a number of attribute-invariant preferences for all kinds of complex categories like cars and celebrities [93, 94]. When activated, neurons with these categorical preferences alter sensory expectancies across modalities and bias the spatial priority of action competitions resolved in the parietal cortex [73, 95].

Because our primary theoretical interest was integrating learning, attention and gaze, and not replicating a fully developed category learning model we chose to use a relatively simple learning system (cf. [96]). The only weights that are adjusted within each experiment are those between the Feature Detection Neurons and the Category Neurons. This means that LAG-1 is, for now, only capable of linear association, limiting its general ability to test rules and learn categories that are not linearly separable [21, 96, 97]. Implementing a one–level–more–detailed category learning system for LAG-1 would require a richer version of visual feature-based categorization inspired by the functions of IT/MTL/hippocampus [93], reentrant relationships

between temporal, visual, and parietal cortices [63, 72, 98], prefrontal cortex for choosing among differing rules and strategies [57, 99], and reward modulated Hebbian plasticity for more explicit temporal difference learning [100–102].

Feature Detection Neurons. The Feature Detection Neurons represent the non-spatial attributes of a stimulus. Feature values are depicted in Fig 2 as color patches in different locations of the screen. In the experiments we simulate, the stimulus features are binary valued. That is, each feature can have only one of two possible attributes for a given trial. LAG-1 represents this as two features connected to the same region of space in the Spatial Attention Field and Visual Field, and with inhibitory connections between the two Feature Detection Neurons tuned to each location.

The Feature Detection Neurons, defined by Equation 22, are agnostic to the property of the stimulus they are representing. The color patches we use are a convenience and could be thought of as representing any feature dimension of the stimulus. The color categories represented by the Feature Detection Neurons in our examples could be understood as explicit representations in a psychological color space like that found in inferior temporal (IT) cortex, as easily as any other physical property of the stimulus [66, 103].

Category Neurons. Categories in LAG-1 are represented by discrete neurons, as defined in Equation 24, with one neuron for each category in the task. Importantly, the Category Neurons do not have to be directly maintained by the excitatory projections of the Feature Detection Neurons: they can be completely self-sustaining once activated enough. The Category Neurons also compete with one another via the same simple global inhibition as the Feature Detection Neurons.

Activation propagating from Feature Detection Neurons to the Category Neurons is attenuated by gain, defined by Equation 25. The total difference between the synaptic weights connecting the Feature Detector Neurons associated with a particular location, for each Category Neuron can be used as an indicator of the information value of the feature dimension at that location for that category. For example, if Feature 1 and Feature 1' both have a connection to Category Neuron A of strength 0.5 (thus a difference of zero), then Category A will be equally active, regardless of the value of Feature 1. In contrast, if Feature 1 has a connection of.02 and Feature 1' has a connection of.98 with Category Neuron A (and thus the difference between them is large) then this location is treated as very diagnostic of this category.

The purpose of gain in LAG-1 is to temper the effects of associative learning based on stimulus base-rates, with a derived measure of information value. It is known that certain LIP neurons represent a kind of gain as in the form of expected posterior log likelihood ratios of reward and information that would be returned by saccades to each location of the field [38, 72, 101].

Feature to category association. The weight matrix connecting the Feature Detection Neurons and the Category Neurons, changes as the model proceeds through the experiment. The values $W(i, j, t)$ store the strength of the association between Category Neuron i, and Feature Detection Neuron j, at time t. On each time step of trial feedback, these weights will undergo at least two of the three types of associative learning as defined in Equation 26: 1) increasing the weights between active Feature Detection Neurons and the above-threshold Category Neuron, 2) decreasing all the weights proportionate to the average increase stimulus-specific increase (homeostatic or normalizing "decay"), and 3) decreasing the weights connecting the chosen Category Neuron and the active Feature Detector Neurons.

Feature Expectation Neurons. At the start of an experiment, LAG-1 has no information about which features are relevant to the categorization task. As the experiment progresses, it learns how the presence of a particular feature value at a particular location predicts particular categories. The way these expectations are translated into changes in behaviour is through the

selective activation of Feature Expectation Neurons by the Category Neurons, as well as inhibition projecting from the Feature Detection Neurons. Category Neurons are connected to the Feature Expectation Neurons, defined in Equation 30, using a copy of the synaptic weight values as those that connect the Feature Detection Neurons with the Category Neurons. The recurrent projections from the Feature Expectation Neurons into the Visual and Spatial Attention Fields change the salience/priority of competing saccade target locations.

Click Decision Neuron. The activity of the Click Decision Neuron gradually increases over the course of the trial as LAG-1 looks around and activates categories, and as its decision impatience grows. There are two critical threshold events signaled by this neuron. Crossing the first threshold corresponds to the subject pressing a button to select a category, and transitions the trial into the feedback phase, wherein information representing the correct answer appears on the Visual and Spatial Attention Fields. When the Click Decision Neuron, defined in Equation 31, crosses its second threshold, the trial ends. The amount of time it takes to cross the gap from the first threshold to the second threshold is variable, such that high enough impatience and/or category knowledge can push these neuron across threshold quickly enough to count as a "double-click" that skips feedback entirely.

Simulations

The present work is based on two key ideas. The primary *theoretical* idea that LAG-1 instantiates is that learning, attention, and gaze, connect in a way that allows category-feature associations and expected information gain to influence spatial attention which in turn influences the selection of saccadic targets. This theoretical idea explains why the different subsystems of LAG-1 are connected the way they are. The primary *empirical* idea behind the model, is that the qualitative learning-related changes in behaviour, both gaze and choice, that have been documented in the experiments that we simulate are largely the result of these linkages between systems. Thus we have two key proposals; one is the functional connection between systems, and one is the unified causal source of the variety of choice and gaze behaviours shown in these tasks. If LAG-1, built on our key theoretical connections, produces the learning-related behaviour changes (e.g., that accuracy goes up throughout learning, that more informative features are prioritised earlier within a trial, that irrelevant features are increasingly ignored, that there is a reduction in fixation counts across learning, and so on) in both category learning choices and in gaze behaviours found in human subjects, then we would take that as evidence for both the theoretical and empirical ideas. If the model reproduces only some of the findings, we will take that as evidence that either the theoretical or empirical ideas, or both, require modification. If there are many findings that the model cannot qualitatively reproduce, then we will take that as evidence that the idea is incorrect, or incomplete.

Breadth and precision of the modelling

It is important to recognize that our primary interest is in the breadth of LAG-1, rather than its precision. That is, we are less concerned about the precise quantitative magnitiudes than we are in seeing if LAG-1's behaviours (both choice and oculomotor) change with learning the way people's behaviours do. Our core argument is that relatively simple connections between learning and attention and gaze underlie a wide variety of findings. In an effort to assess this idea in the simulations, we make several modelling decisions which undoubtedly makes getting good quantitative fits more difficult, but which are truer to our project goals.

First we keep the model flexibility low so that we can unambiguously attribute successful simulations to our core claims. The mathematics that allows the model to fixate in two spatial dimensions and run continuously in time are obviously complex, requiring many equations

and many fixed parameter values. However, parameters used in the fitting process—those that can actually change the model's behaviour to produce better quantitative fits—are very few. Our three free parameters (Learning Rate; Trial Impatience; Fixation Impatience) are considerably fewer than comparable models, e.g., [32].

Second, consistent with our claim that the myriad of behavioural phenomena shown in these task are related, we fit a very broad range of findings. Yet, each additional measure we include constrains the model further, and will impact the quantitative goodness of our fits. We note the breadth of LAG-1's predictions in this task are unparalleled to our knowledge. Quantitatively comparing models that do not fit the same phenomena is not common practice, and there are clear problems to doing so. For example, we note that for data that a model does not address, the likelihood of that data given the model is zero. Under this construal, the broad model *always* wins, even if its fits are terrible. Obviously there is something unsatisfactory here, in that a narrower model may still have value, even if it cannot address every finding that a broad model can. Yet, it seems equally unsatisfactory to quantitatively compare overlapping models only on the datasets to which they both apply: essentially placing zero value on fitting more data.

Finally, we reserve a few measures, leaving them out of the fitting process entirely, so that we can verify the accuracy of LAG-1's natural predictions. In LAG-1, the connection between attention and learning processes is relatively constrained: both the magnitude of the connection strength between features and categories, and the information value of those connections, are quantities that derive purely from the knowledge contained in one set of associative weights. We reasoned that if this account is correct, then, as long as we captured important neurofunctional relationships, other complex phenomena might emerge from the basic constraints in the model without additional coaxing on our part. Demonstrating that LAG-1 can make appropriate predictions about unfitted measures is an important way of further supporting our core argument.

Experimental data being fit

Testing LAG-1 requires an experiment in which both learning and gaze are simultaneously measured, and in which changes in learning lead to changes in gaze. As we know of no other process models that target the relationships between learning and attention and gaze found in category learning in this particular way, instead of direct model comparisons, as is often done when looking at very specific effects or representational difference between models, the approach here is to try and match the human data for all the findings and effects listed in Table 1.

In this section, we fit LAG-1 to data produced in two different studies: [10] and [25]. We describe them in brief, here, with a fuller description of the methods available below. We chose the [25] data as it uses a category structure that has been run many times in our lab with different experimenters, stimuli, instructions, base-rate manipulations, information access costs, and response types [11, 26, 27, 29, 30]. The structure elicits the theoretically important behaviour called stimulus-specific attention, wherein attention is allocated differently based on the properties of the stimulus itself. This finding is important because many existing models have one set of attention weights that change only during response feedback, and so cannot change the allocation based on the features of present stimulus. Eliciting stimulus-specific patterns of eye movements will be an important test of LAG-1.

The second data set is from the [10] study using the "5/4" categories of [17]. This eye tracking study was one of the earliest category learning experiments to use eye movements as a confirmational measure of attention weight predictions. The results suggested that prototype

models underestimate attention to the least relevant stimulus dimension in the 5/4 categories (e.g. [32, 104]). Though the 5/4 categories have a long, if contentious (e.g. [105]), history in the study of categorization, using it here allows us to ensure that LAG-1 can model both the appropriate learning and generalization data as well as eye tracking measures. Simulating the [10] experiment also requires that LAG-1 process variable-length experiments with different numbers of features, different arrangements of the features, and different numbers of categories than the [25] study. The reaction time data for the response and feedback sections of the data were not available, and so these measures were not fit in Simulation 2, even though LAG-1 automatically produces reaction times when fit to the remaining data.

In Simulation 1, four behavioural measures were fit: accuracy, fixation count, the probability of fixating irrelevant information on a trial, and the average fixation durations, over the experiment. The data on each of these measures was summarized by an initial intercept and a slope of change in the measure over the experiment, yielding a target vector with eight elements for each human subject. In line with our breadth-focused modelling goals we leave several aspects of the human behaviour to emerge organically from the model. Reaction times, within-trial fixation probabilities, and time spent on features during feedback were left out of the fit function, allowing the fits to these measures to emerge from the model organically.

In Simulation 2, we used the individual subject transfer probabilities (for all 16 stimuli), their learning points (the trial number on which the participant began two consecutive error-free blocks), and a measure of the changes in the allocation of attention to stimulus features of high and low diagnosticity. We were unable to precisely reconstruct all of the individual subject data reported in [10] for the attentional allocation component of the fit measure, so we instead used the reported averages in three measures of attentional allocation. The first two numbers were the average starting and ending fixation counts to the individual features. The third component was the difference between the least informative (Feature 2) and the most informative features (Feature 1) on this measure (~ 0.3). The target vector for Simulation 2 thus had 19 elements. Again we allow the model to predict several aspects of the behavioural data with Simulation 2: in particular the within-trial fixation probabilities and the overall fixation proportions.

The experiments record all of the participants' various responses, when they occurred, and the location of their gaze at all points in time. The simulations of the model are similarly data-rich. Likewise, the attentional and gaze processes in LAG-1 are also spatial and the output of LAG-1 is thus just like the raw data from humans. In fact, we use the same computer code for both humans and simulations to convert the raw data to the aggregated and binned measures presented below. This applies to gaze also: just as with humans, raw gaze location described in x,y coordinates is recorded at a time scale set to approximate a 120hz eye tracker, and this is aggregated into fixations using the same modified gaze dispersion algorithm [106] that was used for the human participants in Simulation 1. Once the final best-fitting simulations are produced (see the supplementary information "S3 Appendix: Fitting procedure" for a full description of the fitting procedures, and supplementary information "S5 Appendix: Parameter tables and best fits" for the best-fitting parameter values) the result is a simulation yoked to each human subject: matching both the experimental situation, and, as well as possible, the impatience and learning rate characteristics of that subject.

Free parameter relationships with model behaviour

Before we cover the results of the simulations, it might be helpful to the reader to get a sense of how the model behaves under various values of the three free parameters. Model output at varying levels of the Learning Rate, Fixation Impatience and Trial Impatience is shown in Fig

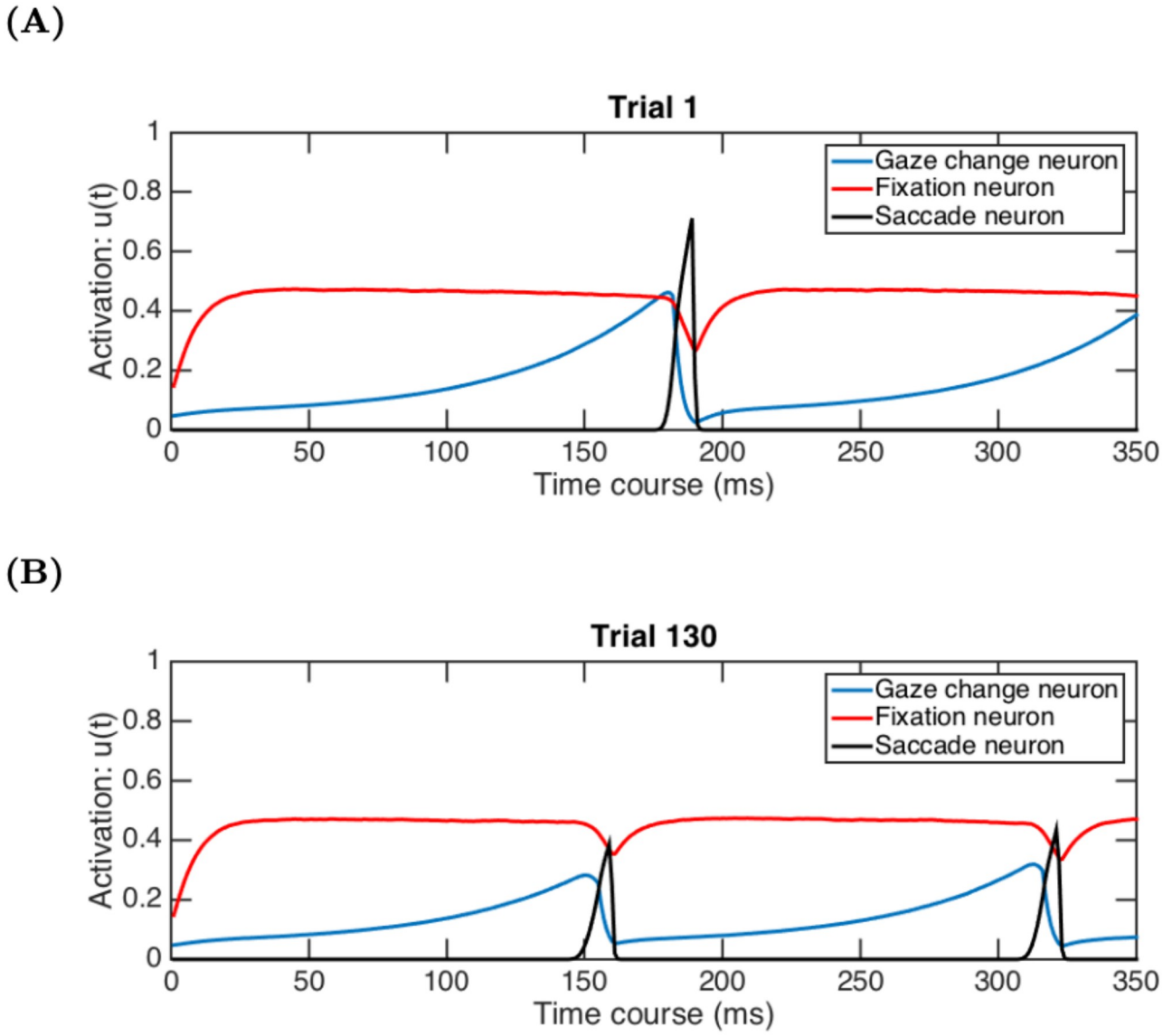


Fig 3. Saccadic timing neurons, Two examples of the contemporaneous activity of the three neurons of the Saccade Timing System, at two different points in learning. A) Before the model has made any strong feature-category associations, different stimulus features are equally interesting to it. Because there is little preference for particular features at this point, competition to determine the next saccade target takes longer to resolve than later in the experiment, extending the fixation duration in time. In these cases, the Gaze Change Neuron (blue) eventually surpasses the Fixation Neuron (red), just prior to the phasic spike in activity of the Saccade Initiation Neuron (black) which initiates a saccade. B) After the model has established stronger associations between features and categories, the knowledge that particular feature locations contain useful information, speeds targeting by boosting priority at these locations in the Spatial Attention Field. This is reflected in the modified timing relationships of the three neurons, where the Gaze Change Neuron does not reach the same level as the Fixation Neuron before the Saccade Initiation Neuron spikes. As a result, fixation durations are 30–40 ms faster than at the start of the experiment.

https://doi.org/10.1371/journal.pone.0259511.g003

3. Each parameter combination used to generate the plots is chosen from the storage table of LAG-1 for Simulation 1 output according to the closest models to low, medium and high levels of one of the parameters when the other two parameters are held to the overall medians observed in the individuals. Fig 3A shows how Learning Rate modulates accuracy by

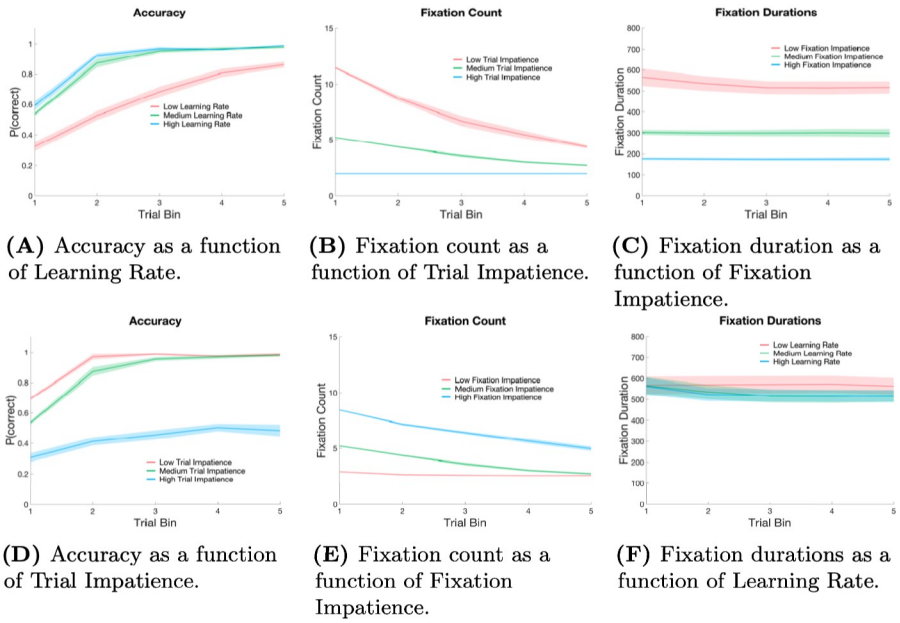
simultaneously plotting LAG-1's learning curves given low, medium, and high Learning Rates, taken from the best fits of the 42 subjects of Simulation 1. The free parameters map onto behaviour in relatively straightforward ways. As the Learning Rate increases, the connections between the Feature Detection Neurons and Category Neurons change more rapidly, leading to improved accuracy. In Fig 3B, as the Trial Impatience increases, the activation of the Click Decision Neuron grows more rapidly, resulting in faster decisions, and therefore fewer fixations. Finally, as the Fixation Impatience parameter increases, as shown in Fig 3C, activation of the gaze change neuron grows more quickly, resulting in shorter fixations.

However, complex interactions amongst the parameters are also important, and meaures of accuracy, fixation duration, and fixation count, can be influenced by changes to the other parameters as well. In Fig 3D, the Trial Impatience parameter strongly influences the learning curve by scaling the amount of time spent on feedback. Fixation counts in Fig 3E are influenced by Fixation Impatience because they change the number fixations that can occur within a particular interval of time. Fixation durations may moderately shrink over the course of experiment as learning increases the recurrent input to the Spatial Attention Field, eliciting the onset of saccades more rapidly. But to underscore how complex the predictions owing to these interactions can be, in the final plot, it is only when Fixation Impatience is high, that this decline is noticeable in Fig 3F.

Simulation 1

The eye tracking and category learning data for the first set of simulations comes from the publicly available data set originally reported in [25]. Participants in this experiment were instructed to classify fictional microorganisms, defined as having three critical organelle features (see Fig 4). Each feature could take on two possible values, yielding eight stimuli in total, which were associated with four different categories (A1, A2, B1, B2). Features are arrayed in one of three locations. As is typical for category learning studies, the assignment of image pairs to feature dimensions (feature one, two, and three), and features to locations was roughly counterbalanced across participants. For LAG-1, we simplify the stimuli from complex cell organelles down to three colored features with two possible values, each with its own layer contributing input to the Visual Field (see Fig 2). After making a decision, the participants were given feedback presenting the stimulus again, their own response, and the correct response. Only participants who had at least 70% of their gaze collected in total were included for fitting, and only trials having more than 75% of gaze collected were included in the analyses. Participants had to meet a learning criterion of 24 trials in a row correct in order to be included, leaving 42 individuals in the data set. Of the 480 trials of data available from each subject, we look at only the first 360 in order to reduce simulation time; human behaviour changed very little after that point. More comprehensive details of this experiment can be found in the original paper and on the data repository website.

Stimulus-specific attention—the useful finding that emerges from this category structure—occurs after the participants learn to classify the categories correctly. As can be seen in Table 2, Feature one is always useful, in that it can tell the difference between categories A and B. Features two and three are contingently useful: that is, if Feature one takes a 0 value, then Feature two is important for telling the difference between A1 and A2, while Feature three is no help at all. Likewise, if Feature one has a value of 1, then Feature three can say whether the stimulus is a B1 or B2 but Feature three is irrelevant. Participants can, and do, learn this: the data of within-trial fixation probabilities after learning this category structure show that Feature one is the most likely to be fixated early in the trial and the second most relevant feature for that stimulus, either Feature two or three, is most likely to be fixated later in the trial. Models using a



(A) Accuracy as a function of Learning Rate.

(B) Fixation count as a function of Trial Impatience.

(C) Fixation duration as a function of Fixation Impatience.

(D) Accuracy as a function of Trial Impatience.

(E) Fixation count as a function of Fixation Impatience.

(F) Fixation durations as a function of Learning Rate.

Fig 4. Isolated parameter influences on model. In each of the six figures a comparison of the model's behaviour is presented at three different levels of the parameters. Each set of parameters is chosen from the closest models available when restricted to the range set by the lowest and highest observed in the best fits of individuals in Experiment 1. In each case, one free parameter is compared while the other two parameters are held to the overall medians observed within that same range. A) Model accuracy over three levels, low, medium and high, of Learning Rate ($[2.5 \times 10^{-6}, 1.1 \times 10^{-5}, 1.65 \times 10^{-5}]$), when Fixation Impatience is 1.7 and Trial Impatience is 1.8. B) Model fixation counts over three levels, low, medium and high, of Trial Impatience ([1.6, 1.8, 2.05]), when Learning Rate is 1.1×10^{-5} and Fixation Impatience is 1.5. C) Model fixation durations over three levels, low, medium and high, of Fixation Impatience ([1.5, 1.7, 1.9]), when Learning Rate is 1.1×10^{-5} and Trial Impatience is 1.8. D) Model accuracy over three levels, low, medium and high, of Trial Impatience ([1.6, 1.8, 2.05]), when Fixation Impatience is 1.7 and Learning Rate is 7×10^{-6}. E) Model feature fixation probability over three levels, low, medium and high, of Trial Impatience ([1.6, 1.8, 2.05]), Fixation Impatience is 1.7 and Learning Rate is 1.1×10^{-5}. F) Model feature fixation durations over three levels, low, medium and high, of Learning Rate ($[2.5 \times 10^{-6}, 1.1 \times 10^{-5}, 1.65 \times 10^{-5}]$), Fixation Impatience is 1.7 and Trial Impatience is 1.8. Shading represents standard deviations of the model for that level of parameters.

https://doi.org/10.1371/journal.pone.0259511.g004

single set of weights cannot shift attention depending on the currently viewed stimulus within a trial, and so cannot account for this finding [11, 16, 107, 108], but some models can account for this at an aggregate feature level [109–111].

Simulation 1 results—direct fits

We first look at model performance on measures that we fit directly—that is, some aspect of the measure was a component of the fit function. These basic measures reflect the learning of the categories, and the corresponding changes in the allocation of attention.

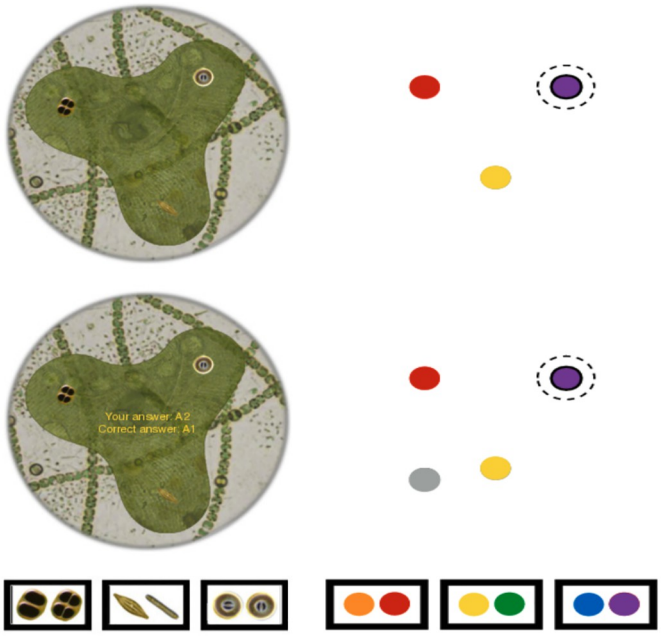
Learning curves. The first measure we look at is classification accuracy. Chance performance in this task is 25%, setting a baseline from which a learner will increase over the course

Table 2. Category structure used in [25].

Feature 1	Feature 2	Feature 3	Category
0	0	0/1	A1
0	1	0/1	A2
1	0/1	0	B1
1	0/1	1	B2

The two values that each feature can take on are represented by 0s and 1s. If a feature is irrelevant for correct classification of a category, it is represented in the table by virtue of being either 0 or 1, denoted as 0/1.

https://doi.org/10.1371/journal.pone.0259511.t002



(A) Experiment stimuli. **(B)** Model stimuli.

Fig 5. Simulation 1: Experiment and model stimuli examples. A) An example stimulus from the response phase (top) and feedback phase (bottom) from [25]. Each feature type subtends approximately 1.7˚ of visual angle, separated by 10.6˚. B) The features as they look to LAG-1 during the response phase (top), and feedback phase (bottom) showing the location of the feedback button. Features are represented as simple color patches (top right). The feedback button provides the correct answer when foveated by the model. Also depicted by a solid black outline top the top right feature is 1 standard deviation of LAG-1's fovea, given a fixation to the top right hand feature, where the dashed line reflects 2 standard deviations from highest acuity. Stimuli are counterbalanced between subjects, such that features appear at different locations and have differing relevance.

https://doi.org/10.1371/journal.pone.0259511.g005

of the experiment. Fig 5A shows the averaged learning curve for both the model and the human participants' learning performance.

Fixation durations. Fixation durations reflect an important aspect of attentional optimization: look too fast, and crucial information can go unnoticed (e.g. [112]), look too slow and you waste time, potentially slowing learning down at the trial level. The mean fixation durations for the human subjects, and the accompanying model fits seen in Fig 5B, are just above 300ms. These durations typically decrease over the course of the experiment. Similarly, LAG-1 produces fixation durations that decrease across the experiment.

Fixation counts. The total number of fixations on a trial is one of the more obvious indicators of task efficiency. As seen in Fig 5C, the human data show a general trend to start 6 fixations per trial and finish between 3 and 4 fixations per trial. LAG-1 starts with roughly 7 fixations per trial, and by the end is indistinguishable from the human data.

Fixating irrelevant information. Another method of describing the optimization or efficiency of attention is the probability of fixating an irrelevant feature over the course of an experiment [10, 26]. The data in Fig 5D compares model and humans on this measure, the measure of which is calculated by assigning a 1 or 0 to each trial based on whether or not the irrelevant feature was fixated. Under our current best fitting parameters, LAG-1 has a bias to look at the irrelevant feature more often at the start of the experiment than humans do. In fact, at the start of the experiment, LAG-1 almost invariably looks at all features, and so is more likely to fixate all features, not just the irrelevant one. The finding that people do not attend all

features at the beginning of learning is not unique to this study: many previous studies have reported similar findings, though typically not as extreme, showing that some features are not being fixated even when the fixation count is high [10, 25, 26]. There could be a number of explanations for this. One possibility is that participants are engaging in rule testing strategies such that they fixate only one or two features. Previous research has observed a tendency for participants to start by using simple, unidimensional rules in similar tasks [113, 114]. Regardless of the real cause, it is clear that participants in the experiments here just do not look at all the features the way that LAG-1 does, and some adjustment to the model will eventually be needed to better fit this human data.

Simulation 1 results—indirect/predictive fits

In this section we look at three additional measures which were not in the fit function at all. These measures vary from being moderately constrained by the basic attentional behaviours (reaction time is related to fixation durations and fixation counts, for example), to being mostly unconstrained in its fit to fixation durations and counts (e.g., within-trial fixation probabilities, including the important stimulus-specific attention finding).

Reaction times. Reaction time measures can give some insight about attention in category learning (e.g. [31, 115]), and have been modelled before using a variety of methods (see [32–34]). The reaction time data in Fig 5E looks like what one would expect: the responses of both the human participants and LAG-1 speed up with experience. In LAG-1, learning increases the input from the Category Neurons that drive the Click Decision Neuron. The model reproduces the general trend of decreasing reaction times across the experiment, and the model, like the humans, begins at about 4 seconds. The model is slightly too fast by the end, and generally has less variability than the human participants. In humans there are of course additional processes that will affect reaction time that are not captured in LAG-1, such as mind wandering and exogenous distractions, limiting the variability that can be fit on this measure.

Within-trial fixation probabilities. Within-trial feature fixation probabilities have previously been reported as another way of quantifying attentional priority [10, 11]. This measure, used in Fig 6, reveals the general order in which the features are fixated. Note that this measure describes within-trial behaviour, whereas all of the data used in the fitting procedure are at the level of the changes across the experiment; there is nothing in the fitting procedure that rewards approximating this within-trial data. The category structure is designed to elicit different optimal fixation orders for the A categories (see Fig 6A for the human data and Fig 6B for LAG-1), differently than the B categories (Fig 6C for the human data and Fig 6D for LAG-1). The trends observed in the participants when looking at just the features of interest (where at any given point in the trial the total probability sums to unity—fixations outside AOIs are excluded from these data for both humans and simulations), show that for the A categories, there is a Feature one precedence, followed by Feature two, and the irrelevant Feature three dropping in probability over the trial. For the B categories Features two and three are flipped as a result of the change in feature relevance. These same trends are present in LAG-1. As before, the model fixates irrelevant information a bit too often, and thus the differences are less dramatic.

Time spent on features during feedback. Fig 5F reports the average time spent looking at features of the stimulus during the feedback phase of each trial over the course of the experiment. LAG-1 captures the qualitative trend after this initial large decline. Studies of feedback during category learning are rare [30, 37] and we do not yet know enough about the processes involved, or how they change as knowledge about the task grows, to make strong empirically

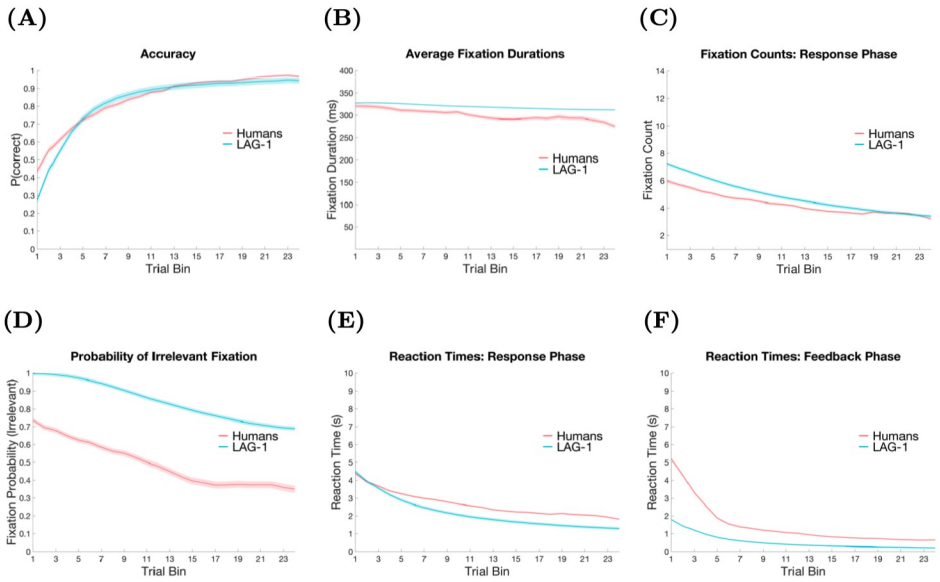


Fig 6. Simulation 1 model fits. Each measure is depicted over 360 trials averaged into 24 bins of 15 trials. Color shading represents standard error of the mean accumulated from individuals and population. A) Trial accuracy means and variability are fit well. B) Fixation duration variability is fit well but the final late reduction means are diverge. C) Fixation count means, variability and change are qualitatively fit well. D) Probability of fixating the irrelevant shows roughly the same change and variability over the experiment, however the scale is off by about 25% at all points. E) Reaction time means, variability and change are qualitatively fit well. F) Total time looking at features during feedback has roughly the right change and variability but the scale is off by 50% at all points.

https://doi.org/10.1371/journal.pone.0259511.g006

derived recommendations, but it would appear that LAG-1 offers a straightforward way to account for some of the variance of this measure.

Individual fits. The fits shown thus far have been averages across many individual participants and simulations. To give the reader a better sense of how LAG-1 addresses differences across individual participants, we show fits to three individual subjects chosen for their differences in behaviour. Amongst these three participants, there is fast and slow learning, short and long fixaton durations, and many and few fixations per trial. Fig 7 shows data for three human participants and for the several model simulations run for each participant. The individual fits for the remaining participants can be found in the supplementary information beginning at Fig 21, in "S6 Appendix: Individual fit visualizations.". Overall LAG-1 seems responsive to individual differences, and produces large and small numbers for counts, durations, and accuracy. As far as individual fits, there are no obvious peculiarities, though, for some participants, LAG-1 fails to fit the magnitude of a measure, or its slope. As in the aggregated data, it is clear that LAG-1 is not capturing the variability of human data, though it seems better for accuracy than for the other measures. We kept noise relatively constrained for these processes, but additional noise, for Trial Impatience in particular, seems warranted. We note, though, that some of this variability is extreme and seems more likely to be strategic, or to be related to phenomena like mind-wandering [116] than to be a matter of simple noise.

Simulation 2

The second simulation is of response data and eye movement data from [10]. Participants in this experiment were instructed to classify fictional schematic drawings of insects, defined as

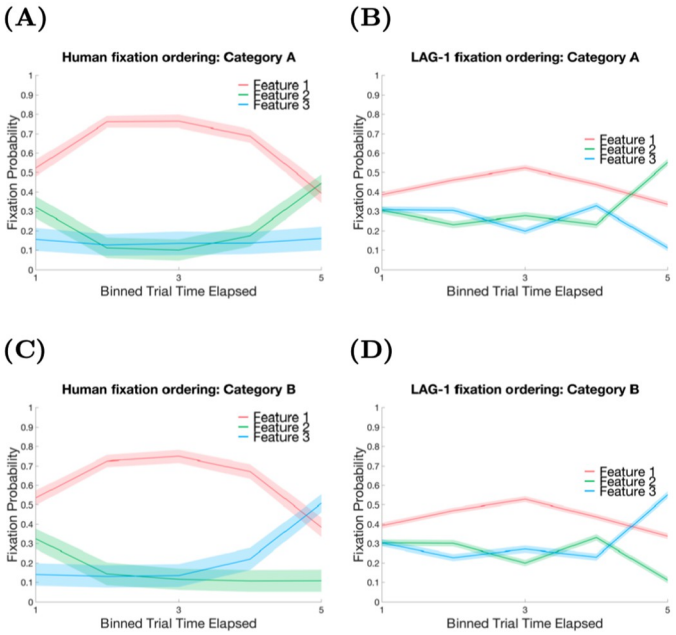


Fig 7. Within-trial feature fixation probabilities. Within-trial feature fixation probabilities averaged across all post criterion point trials and scaled to 100 data points for each trial and averaged into five bins. A) Human data for A categories and with LAG-1's comparable performance in B. C) Human data for the B categories with LAG-1 again beside it in D. The three important features of the data are: the primacy of Feature one through most of the trial; the primacy of the alternative feature at the end of the trial; and the change in the identity of the second feature across categories (For Category A stimuli it is Feature two, and for Category B stimuli it is Feature three). LAG-1 qualitatively captures all three findings.

https://doi.org/10.1371/journal.pone.0259511.g007

having four diagnostic features (see Fig 8). Two of these features were highly diagnostic (and equivalently so) of the category, one was of medium diagnosticity, and one of low diagnosticity. Each feature could take one of two possible values, yielding sixteen stimuli in total. Nine stimuli were shown during training and were associated with two categories, leaving seven for classification during a feedback-free transfer phase (see Table 3). After choosing a category, the participants would get feedback, and see the stimulus again. After 189 trials or 18 correct trials in a row, whichever came sooner, a transfer phase would begin that twice tested all of the stimuli, including the seven previously unseen stimuli. There were 64 participants that met the learning criterion with minimal loss of gaze data. Further details of this experiment can be found in the original paper.

The best fitting models in Simulation 2 were found by comparing individuals with LAG-1 on 2 subject-level measures and 2 population-level measures. Where Simulation 1 had 8 components to the error function, Simulation 2 uses the 16 transfer responses of each individual, as well as their individual learning speeds, and the average fixation counts at the start and end of the experiment, for a total of 19 components as the objective function (weights listed in Table 8 S5 Appendix). Because we did not have individual data for all measures, we do not show fits to individual subjects as we did in Simulation 1.

Simulation 2 results—direct fits

Again, we first look at model performance on measures that we fit directly—that is, some aspect of the measure was a component of the fit function.

Transfer probabilities. The first measure we look at is transfer behaviour. Fig 9 shows the transfer responses for both human participants and our LAG-1 simululations. These responses

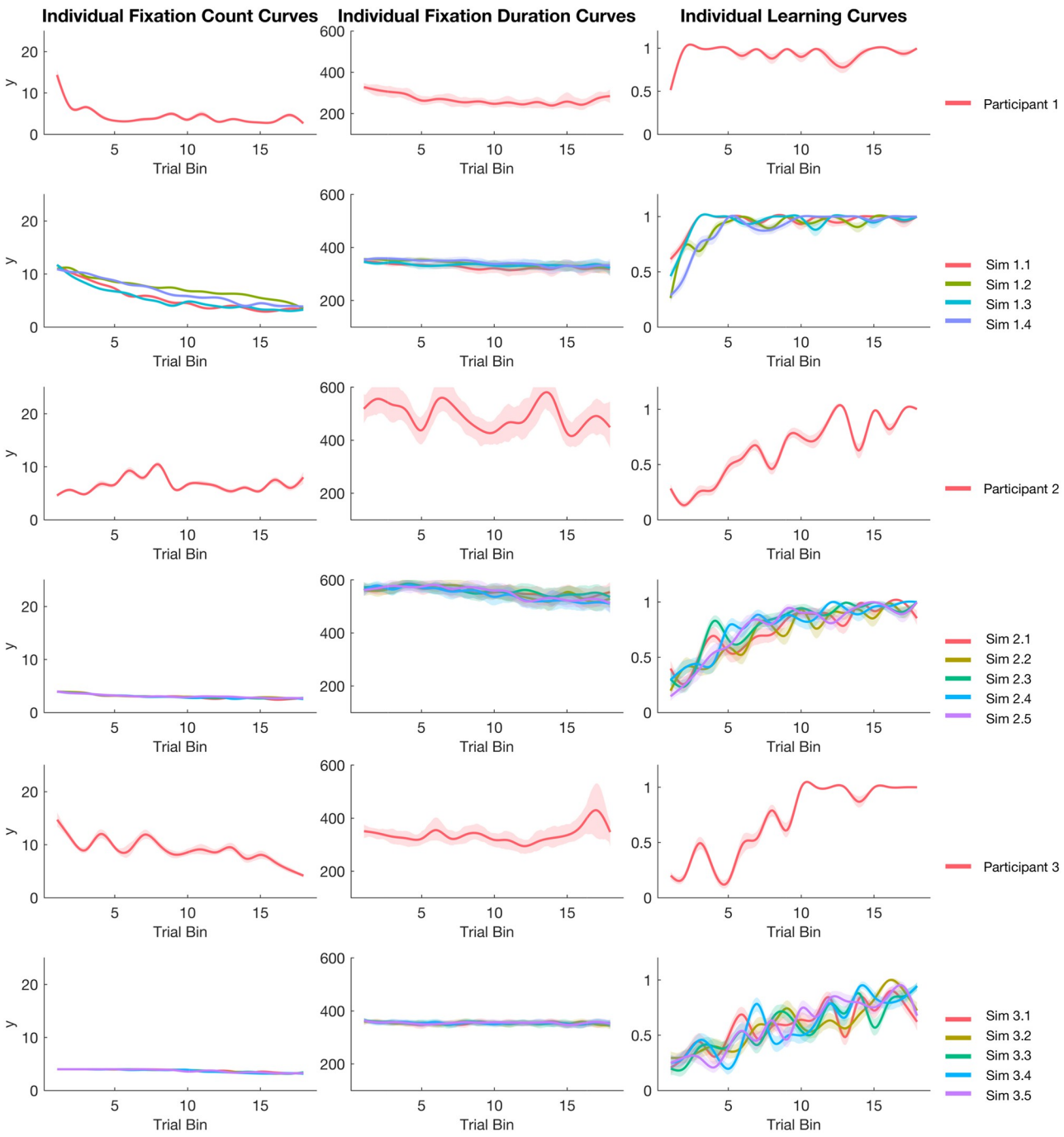


Fig 8. Fits to individuals. Data from three individual participants for fixation count, fixation duration, and accuracy. Below each participant are LAG-1's fits to that participant's data. Though the fits are not perfect in all cases, the model is showing both more and fewer fixations (counts), shorter and longer fixations (in milliseconds), and slower and faster learning (proportion correct) in accordance with individual differences.

https://doi.org/10.1371/journal.pone.0259511.g008

Table 3. Category structure used in [10].

Feature 1	Feature 2	Feature 3	Feature 4	Category
1	0	1	1	A1
0	0	1	1	A2
0	1	1	1	A3
1	1	1	0	A4
1	1	0	1	A5
1	0	1	0	B1
1	0	0	1	B2
0	1	0	0	B3
0	0	0	0	B4
0	1	1	0	T1
0	0	1	0	T2
1	1	1	1	T3
0	0	0	1	T4
1	1	0	1	T5
0	1	0	1	T6
1	0	0	0	T7

Feature values of 0 and 1 represent the two values of each feature can take on. Comparing the feature values to the categories will reveal that Feature one is of low diagnosticity, Feature two is of medium diagnosticity, and features three and four are of high diagnosticity, relatively speaking.

https://doi.org/10.1371/journal.pone.0259511.t003

are to both trained A and B category stimuli, as well as the seven transfer stimuli not presented during training.

Fixation probabilities and counts. Trial-level fixation measures reported in [10], are presented in Fig 10 along with the comparable output from the best fitting LAG-1 models. It is clear that by the end of the experiment human participants are fixating the least diagnostic

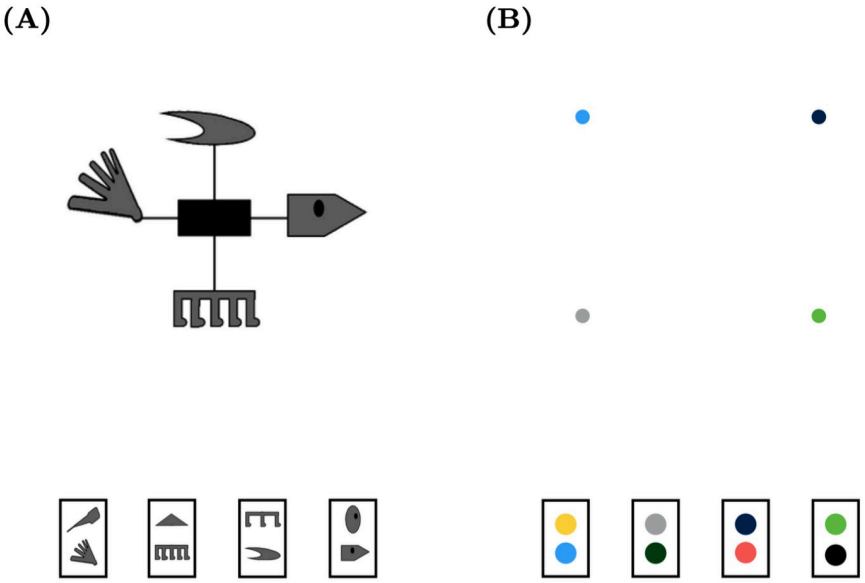


Fig 9. Simulation 2: Experiment and model stimuli examples. A) Example stimulus used by [10]. The four features of interest were reported to subtend approximately 4˚ of visual angle each, with the stimulus height and width being 12˚. B The four color features, spaced 24 spatial units horizontally and vertically and 34 spatial units diagonally, presented to the model.

https://doi.org/10.1371/journal.pone.0259511.g009

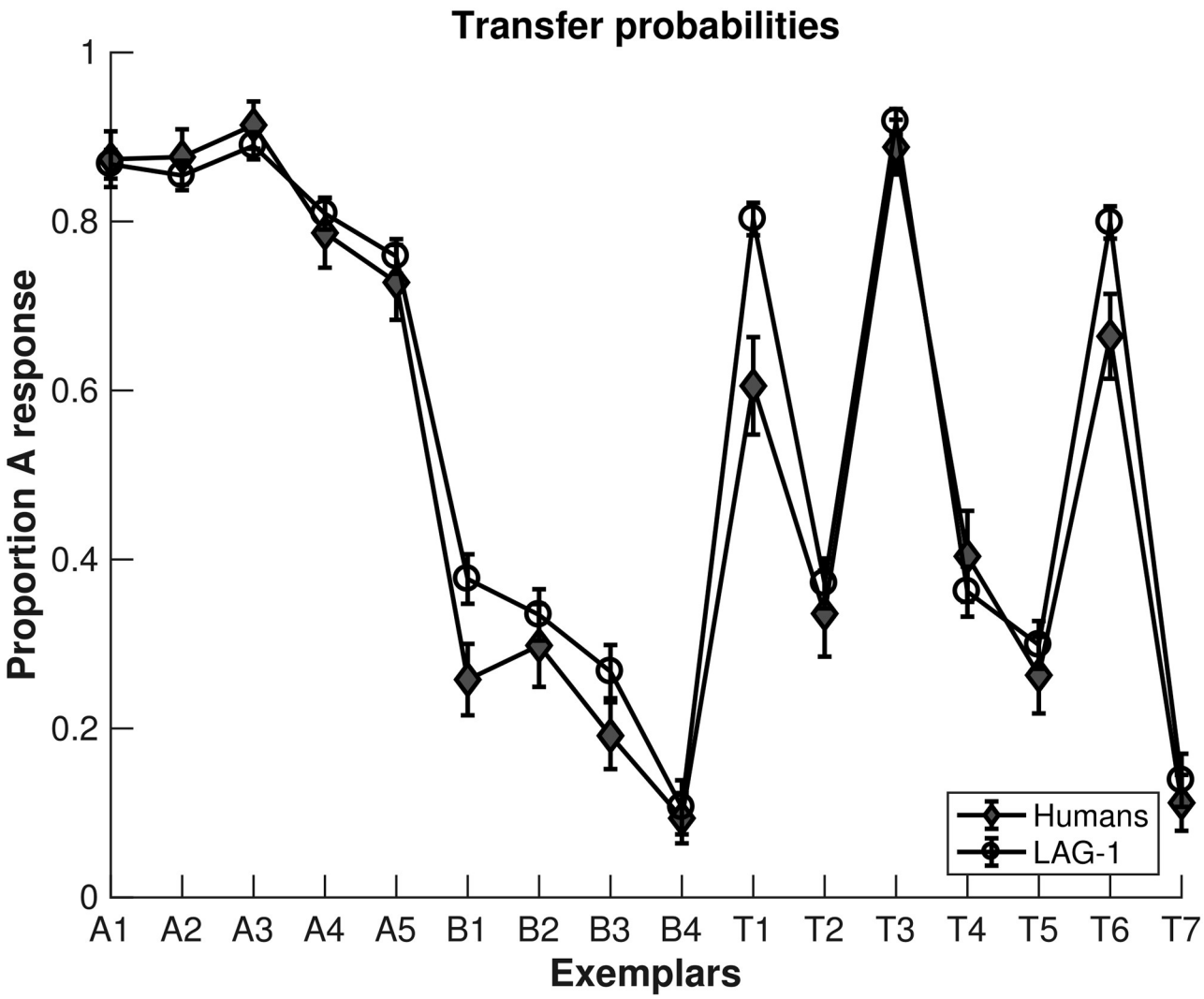


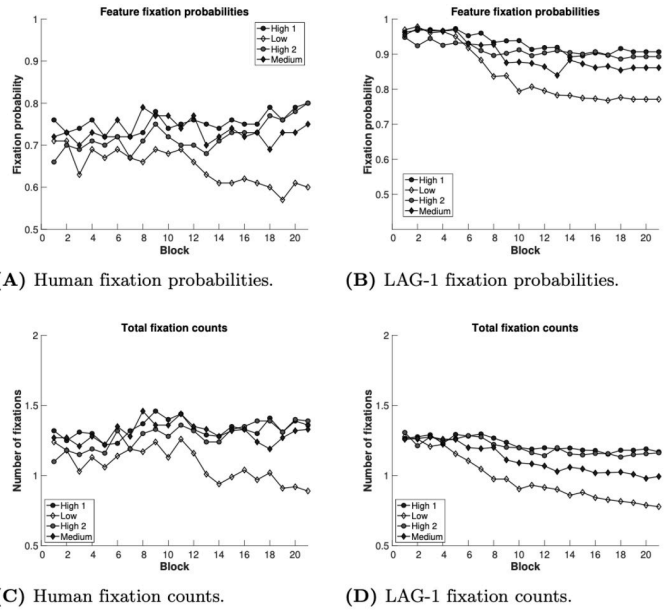
Fig 10. Simulation 2: Model fits to transfer responses. Average categorizations on the 16 transfer stimuli, over all models/subjects. Error bars represent standard error of the mean.

https://doi.org/10.1371/journal.pone.0259511.g010

dimension, Feature 2, less often and for less overall time than the other features. The fixation probabilities to each feature are generally higher in LAG-1, just as they were in Simulation 1 (as indicated by the probability of fixating the irrelevant feature measure over the course of the experiment). However, LAG-1 does show analogous reductions in fixation probability and fixation counts, to the least-diagnostic feature.

Simulation 2 results—indirect/predictive fits

Fixation proportions. As the purpose of [10] was to use an overt empirical measure of attentional optimization to test model predictions, one of the measures they reported was the proportion of time spent on each feature in the transfer phase of the experiment (see Fig 11). Most category learning models are fit with a capacity-constrained level of attention, meaning that on any trial the weights to each dimension are normalized in some way; this means that



(A) Human fixation probabilities. **(B)** LAG-1 fixation probabilities.

(C) Human fixation counts. **(D)** LAG-1 fixation counts.

Fig 11. Simulation 2 model fits. A) The probability of fixating a particular feature reported for each of feature observed in human subjects. B) The probability of fixating a particular feature reported for each of feature observed in the best fits of LAG-1. C) The fixation counts to each feature averaged over all trials observed in human subjects. D) The fixation counts to each feature averaged over all trials observed in the best fits of LAG-1.

https://doi.org/10.1371/journal.pone.0259511.g011

the total time on each feature has to be represented as a proportion. In our simulations, LAG-1, like the human participants, reproduces the rank ordering of the features by diagnosticity, with high diagnosticity features receiving the most observations, then medium, then low.

Within-trial fixation probabilities. The last measure we report here is similar to the within-trial fixation probability measure that we reported in Simulation 1. In Fig 12 the first four seconds of LAG-1's average within-trial fixation probability to each feature is reported alongside the human data presented in [10]. LAG-1 reproduces the approximate ordering of peak fixations to features of differing diagnosticity: the higher diagnosticity features have peak fixation probabilities earlier than less diagnostic features. The lowest diagnosticity feature is fixated much less often in the first few seconds of the trial for both humans and the model. As in Simulation 1, we do not use within-trial information as part of the fitting function; that and, the largely arbitrary relationship between fixation orders and the overall category structure, make these data difficult to fit precisely.

Discussion

Our goal in developing LAG-1 is to provide an account of the integration of learning, attention and gaze. In this model, as in the brain, knowledge about where to look emerges from a hierarchy of reentrant visual and categorical processes [88, 90, 117]. LAG-1, as a theory, proposes that the expected information gain can be derived from simple learned associations and used to guide attention. This information is integrated with bottom-up visual signals in a priority map, from which the entire space of saccade target locations is considered simultaneously. The competitive dynamics of this field change both the saccade location choices and saccade onset latencies, thus illustrating an actual mechanism for changes in gaze patterns and changes in fixation durations over learning. Using the model's three free parameters (Fixation Impatience, Trial Impatience and Learning Rate) we fit LAG-1 to human data [10, 25] from two

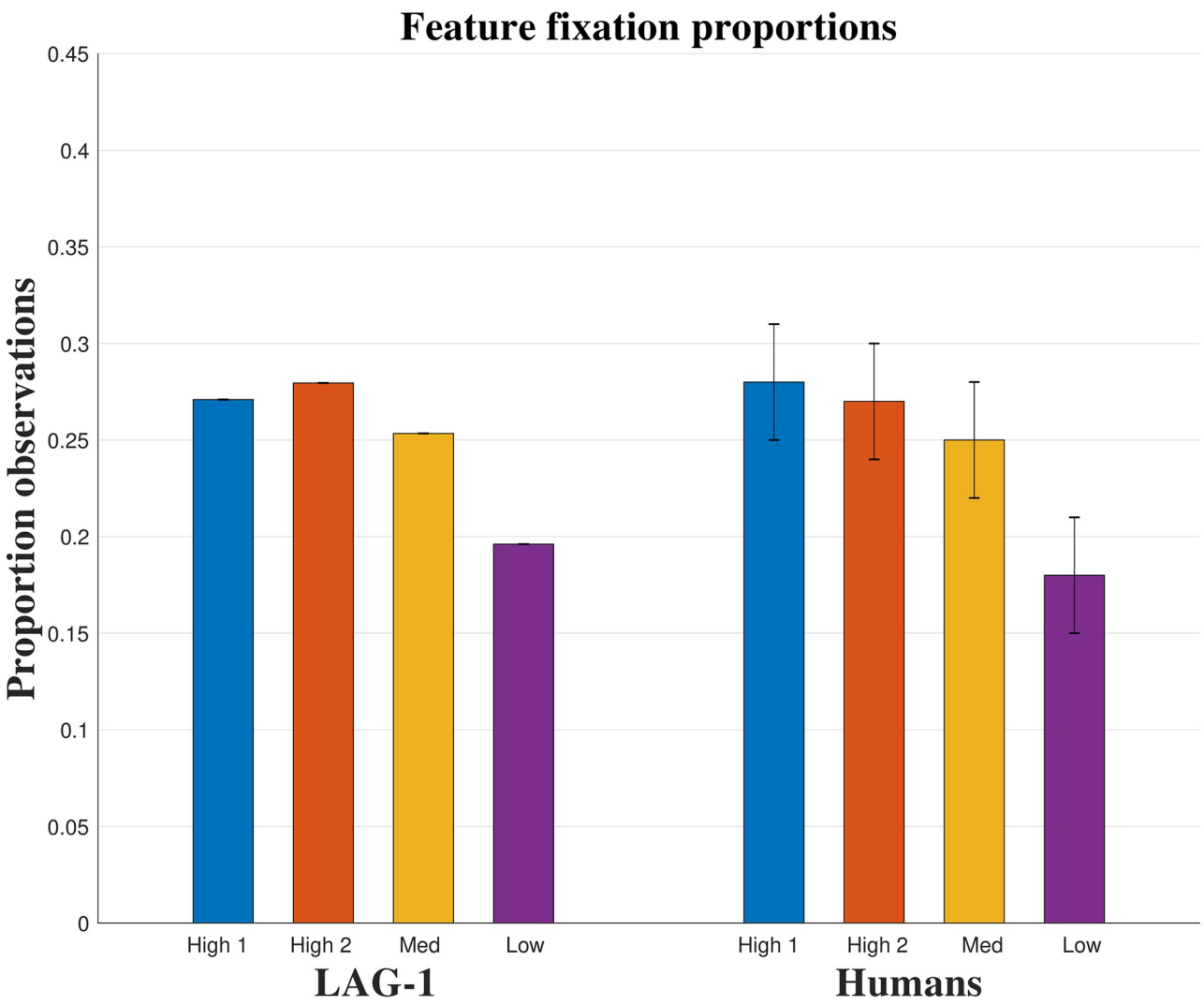


Fig 12. Simulation 2: Fixation proportion fits. The fixation proportions aggregated across just the transfer trials of the experiment. LAG-1 data is on the left, while the data from [10] are re-represented on the right.

https://doi.org/10.1371/journal.pone.0259511.g012

well established and replicable category learning tasks [11, 17]. We fit classic learning measures (i.e., learning curves, transfer probabilities, reaction time), as well as indicators of attention (i.e., feature fixation proportions, fixation count, fixation duration, fixation orders, feedback processing time), many of which have never before been modelled. The aim of our simulations was to test the idea that the diverse learning-related phenomena found in participant behaviour emerged from interactions such as those that LAG-1 embodies. If the model's behaviour changes across learning in the same qualitative ways that human participants' behaviour does, across all the measures, it would provide support for our hypothesis.

Simulation 1 used data from [25] based on the category structure first used in [11] to demonstrate stimulus-responsive attention. This structure is important because the relevance of the features depends on the category to which the stimulus belongs, meaning that using a single attentional strategy for all stimuli is inefficient. Participants have been shown, via eye

tracking, to increasingly ignore the stimulus features that are, for that stimulus, irrelevant. LAG-1 ably captures learning curves and reaction times across the experiment. Further, it simulates all the qualitative eye tracking findings previously reviewed in Table 1. Fixation durations, fixation counts, and probability of fixating irrelevant features all decrease as the model learns. The model naturally predicts the correct relative within-trial fixation probabilities, and, importantly, shows different fixation probabilities for different categories, thus displaying stimulus-responsive attention [11]. Finally, the model correctly predicts a reduction in the time spent viewing stimulus features during the feedback phase [30].

In Simulation 2, we worked with data published by [10] using the 5/4 categories of [17]. This structure is important because of its broad use in the field, and one of the key contributions of Rehder and Hoffman's paper is confirmation, via eye tracking, that, as predicted by the generalized context model even the least discriminative dimension still receives a moderate attention weight [10, 18]. In this simulation, LAG-1 qualitatively matched the transfer responses that have been replicated many times. Further, the model qualitatively captured the eye movement data: more fixations for features that were more diagnostic, and predicts similar within-trial feature fixation probabilities. Importantly, the model also predicts [118]' key finding, that the least diagnostic information nonetheless receives a moderate amount of attention.

Support for our claim that human learning and gaze behaviours in category learning tasks arise from a system similar to LAG-1 derives from our ability to fit the participants data with the model. Our claim is strongest if we: a) use fewer rather than greater numbers of free parameters, because then our fits flow from the structure of the model and not mathematical flexibility provided by additional parameters; b) fit more findings rather than fewer findings, because it strengthens our argument that all gaze related behaviours at multiple timescales flow from these core principles, and c) fit some aspects of the data indirectly, without using them as part of the fit function, because it speaks to the inevitability of these effects given a system of the kind we proposed, and guarantees that the findings are not due to the fitting process, but are instead endemic to the model. We interpret our results as strong support for our claims because learning influences gaze in the way the model predicts in all cases. But, while these three modelling decisions are appropriate to make the strongest possible claim, they also undoubtedly hinder our ability to get the magnitudes exactly right. We believe our choice to focus on modelling breadth rather that precision was the appropriate way to evaluate our claims. Overall, we view the present findings as a remarkably successful first step toward an integrative model.

Opportunities for improvement

While LAG-1 does a fine job fitting the important qualitative findings, which are summarized in Table 4, there is significant room for improvement in getting the magnitudes correct. While many of the findings might be improved by tweaking one or more of the numerous fixed parameters that influence model timings, there are two major areas in particular where LAG-1 diverges from the human data in ways that seem more fundamental. First, the model does not show enough feature neglect (i.e., inattention to critical features), and second, the variability on many measures, like fixation duration and probability of fixating an irrelevant feature, is too small.

Feature neglect is most obvious in Fig 10: LAG-1 fixates too many features. At the start of the experiment, humans ignore a stimulus feature about 25% of the time. Intuitively, this seems strange. After all, how can one learn which features are predictive of the category if they are not viewed? In addition to causing difficulty for LAG-1, this finding also contradicts the assumptions made by models such as ALCOVE, which spreads attention evenly across all

Table 4. Summary of simulation results.

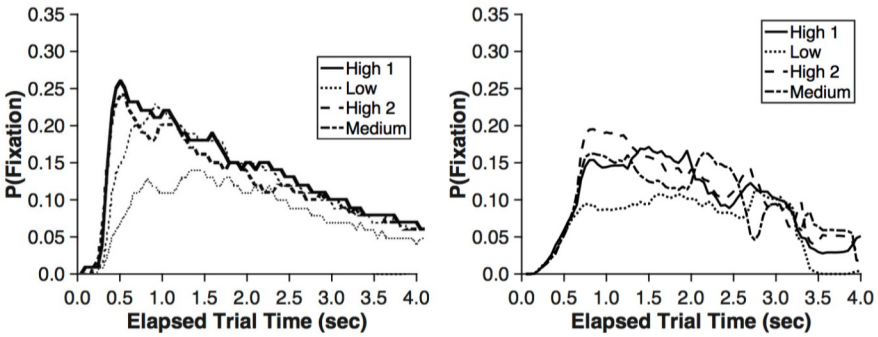
Measure	Simulation 1	Simulation 2
Learning curves	Appropriate scale and reductions (overall).	N/A
Learning transfer	N/A	Appropriate generalization patterns.
Fixation counts	Appropriate scale and reductions (overall).	Appropriate scale and reductions (by feature).
Fixation durations	Appropriate average intercept/change, within-individual variability needs improvement.	N/A
Fixation orders	Appropriate average intercept/change, within-individual variability needs improvement.	N/A
Feature fixation proportions	N/A	Appropriate average intercept/change, within-individual variability needs improvement.
Reaction times	Appropriate average intercept/change, within-individual variability needs improvement.	Appropriate average intercept/change, within-individual variability needs improvement.
Feedback processing time	Appropriate average intercept/change, within-individual variability needs improvement.	N/A

N/A specifies findings not modelled for that simulation.

https://doi.org/10.1371/journal.pone.0259511.t004

features at the beginning of an experiment [16]. It is not only in the [10] data that we find this pattern however, we also see this in the [25] data which had a different number of stimulus features. Indeed, [26] reports the probability of ignoring irrelevant information from 10 experiments, and the range of feature neglect seen in the first block of the experiment was between 10% and 40%. One possibility is that this is a function of missing fixations due to eye tracker error; but two facts mitigate against this idea. First, we ran simulations using LAG-1's gaze data wherein we dropped gaze points (because we used LAG-1's data, there were no naturally missing values) in the same proportion, and with the same sequence length, as in humans, to see if the missing data would cause reduced fixations, and thus feature neglect. It did not. In fact, the overall number of fixations in the experiment increased. Second, we reanalysed the human data from the [25] study, using only trials in which we had 100% of the gaze data and found there was still significant feature neglect. [10] did note that this may be indicative of participants using rule-based strategies initially, but they ultimately conclude that there's likely a mixture of strategies at work. Without further study, it is difficult to say for sure, but we note that there is at least some feature neglect even in the case of mouse driven interfaces such as used in the second experiment of [25]. Regardless, LAG-1 does not show significant feature neglect, and this fact influences the model's fit to several other measures. For instance, the within-trial fixation probabilities do not quite match the human data (Fig 6), primarily because it looks too often at irrelevant features. Overall, LAG-1 does well with the learning-related changes in fixation probabilities, but does not capture all of their overall levels.

The other area where LAG-1 could provide better fits is in predicting the variability across, and within, individuals. This is most noticeable in our attempt to account for fixation durations (Fig 5B), but can also be seen to a lesser extent in measures like the probability of fixating an irrelevant feature (Fig 5D). Given the fact that almost every equation in the model has a noise parameter, it is sensible to ask if we might achieve the appropriate variability by simply increasing the noise on some of these. Possibly. Changing the variability, by increasing noise in one of the components of the model, can have a large impact—one that reverberates throughout all the dynamics of the system. In Fig 13 one can see that the neurons of the



(A) Human within-trial fixation probabilities.

(B) LAG-1 within-trial fixation probabilities.

Fig 13. Simulation 2: Within-trial fixation probabilities fits. A) Human within-trial fixation probabilities averaged over the first four seconds all transfer trials reported in Rehder and Hoffman (2005). B) Comparable within trial fixation probabilities for LAG-1. High (relevance) 1 refers to Feature 1 in Table 3, Low is Feature 2, High 2 is Feature 3, and Medium is Feature 4.

https://doi.org/10.1371/journal.pone.0259511.g013

Saccade Timing System have very little noise. As an example, imagine that noise was added to the neurons of this system. This could increase the variability in fixation durations, but it would also influence ongoing Hebbian learning between features and categories; a fixation that lasts 800ms would end up having 4 times greater category association than would a 200ms fixation, and that, in turn can increase the variability in the speed of learning. Time sensitive activity is known to modify fundamental motor control over the eye [59]: a fact that causes additional challenges for modelling efforts. We felt it was more important, as a first step, to reproduce the primary attentional learning findings shown in Table 1, than to capture the variablity more generally, though we strongly agree with the idea that variabilty is important to fit [119].

Application of LAG-1 to other category learning situations

There are many variations of basic category learning experiments, and we have fit only a small subset of them. The studies we fit had binary valued stimuli, spatial separation of the stimulus features, self-paced responses, self-paced feedback viewing, and the re-presentation of the stimulus during feedback. But, what if they did not? Can LAG-1 handle task variants?

LAG-1 can easily handle any temporal manipulations—for example, adding a specific stimulus presentation duration rather than the self-paced presentation—with no modifications. Manipulations like changing the feedback presentation duration, presenting each stimulus feature for a different amount of time, presenting stimulus features in different orders can also be easily be implemented. Note that many of these experimental manipulations seem likely to influence the performance of the model, and are thus predictions; in LAG-1, as a theory, time matters for many things. For example, if one were to run LAG-1 through two conditions, one with a short feedback phase, and one with a long feedback phase, LAG-1 would predict that the long feedback phase would lead to stronger associations between active features and

categories because of the time dependence of Hebbian learning. We note that this prediction about a temporal manipulation falls outside the scope of what extant category learning models can address because they are generally not modelling cognitive processes in time, and we also note that this prediction has some empirical support [30, 37].

LAG-1 can also deal with spatial manipulations with little to no modifications to the core model. LAG-1 can accommodate different spatial configurations, and make predictions about how this will influence measures like fixation order, reaction time, and even learning. Again, the configuration of the stimulus seems very likely to influence patterns of attention and gaze, and in turn, performance and learning. Imagine an experiment in which, in one condition the features are spread out in a triangle shape, and another condition in which two features are near each other on the left, and the other feature is on the right. This spatial configuration would likely lead to faster reaction times, because they require less time to inspect the features. The activation of two features next to each other (if they are sufficiently close) could join to create a larger peak on the visual field, and thus be more likely to attract attention. What if we compared a situation where the two features on the left were relevant for the correct classification of the stimulus and the feature on the right was irrelevant to one in which the relevant features were separated? It seems plausible that having the relevant features next to each other (and fixated first) might lead to a bit faster learning in that condition, which is to say that the interaction of learning and attention is influenced by the spatial factors that LAG-1 incorporates.

Other kinds of changes to the experiment can be implemented with varying difficulty. Removing the visual feedback indicator (to simulate a sound indicating accuracy, for instance) would be very simple to implement. We would omit the feedback button as input to the Visual Field. The correct category node can still be activated by feedback and the associations between features and categories can change as usual. This requires a change to the experiment (changing what is displayed to the model) but no change to the core model. More substantial changes would be required for something like comparing information sampling using eye movements to information sampling using hand movements. Manual information sampling (e.g., using a computer mouse to click on features to reveal them) is less variable, and more efficient, though slower, and shows similar changes throughout learning [25]. To the extent that we think of information sampling and the allocation of attention using the eyes to be broader phenomena (e.g. "active sampling", [120]), additional mechanisms that implement manual selection of information—such as clicking of menus in a computer program—would be needed. Accepting stimuli with continuous dimensions rather than binary valued features is something LAG-1 is already coded to do, but were not used in the eye tracking experiments we simulate and added unnecessary complexity for the current purposes, so we omitted it here. Having more complex visual processing sufficient to categorize natural images would require more extensive modification. Work by [51] is a good example of how a DNFT-based model might approach this challenge.

Application of LAG-1 to a broader range of experimental situations

Our goal in creating LAG-1 was to build and test a model that can capture the interactions between learning, attention, and gaze. Category learning is a useful task in which to test this integration: participants learn categories and attend to features of varying relevance; further, both response behaviours and gaze behaviours change in concert throughout the experiment. Only a model that specifies how learning is related to the allocation of gaze, and vice versa, can illuminate this co-evolution. A wide range of experimental paradigms within the broader realm of visual cognition depend on these systems, however, and in this section we will provide

a rough sketch of how LAG-1 might be expanded in a variety of ways to address to such situations.

Bottom-up salience. In category learning tasks there is a clear goal: participants are choosing to fixate particular locations to get information that helps them choose the correct category. The emphasis of the task is on learning to direct top-down attention toward task relevant features, the locations of which are well known. The free-viewing of scenes, in contrast, is a function of bottom-up salience. Though LAG-1 does not include many aspects of bottom-up attention–in category learning experiments stimulus differences in bottom-up salience are controlled for by the choice of features and counterbalancing—nevertheless, adding the influence of bottom-up salience into the model is straightforward.

To demonstrate this we added an additional input to the Visual Field (implemented as an additional layer) that acts as a preprocessed salience map. To handle the wider range of field inputs, we also changed the function that chooses saccadic targets from a greedy, maximum activation based selection, to a probabilistic choice, like was used for the category decision. This allowed the model to fixate a broader range of locations. These small changes to the model allowed us to use an image as a stimulus. To do this, the image must be converted to a salience map that, when input to the Visual Field, can drive fixations to the most salient (in the bottom-up sense) areas of the image. Though we left the category learning system in place, there is no input to the Feature Detection Neurons, and so no top-down attention is generated. Fig 14 shows the original image (Roy Lichtenstein's 1973 painting: "Things on the Wall"), the salience map it produced, a heat map of a typical human free-viewing the painting, and a heat map of LAG-1 fixations viewing the image. Note that this implementation of the model predicts bottom-up salience influences the Visual Field earlier than top-down forces because the recurrent loop through the whole system takes extra time. Indeed, short latency saccades are associated with bottom-up salience and longer latency saccades are more influenced by behavioural goals [121].

For expediency, in this demonstration we created the salience map with a common Matlab package [122]. For example, we could implement the components of the salience map (i.e., color, orientation, intensity) as layers of the Visual Field, and calculate the salience from that activation. Abrupt onset and motion are also known to act as a source of salience (e.g., [123, 124]). Both of these factors might be implemented using large changes in activation at a particular location of the Visual Field to boost attention (see [45] for a DFNT-based approach to change detection). It is intriguing to consider that a system with robust, top-down attention may actually *require* a strong abrupt onset salience mechanism to function effectively because significant top-down attentional boosting of known, or expected features and locations, might otherwise make fixating new, unanticipated stimuli too improbable and slow.

Uncertainty in feature location. In category learning the main focus is on the learning the connection between features and categories. In other situations the location of the information sought is unknown. One of the most iconic visual cognition tasks is visual search [125]. In visual search the categories used in the task (target and distractor) are simple and explained from the beginning; no learning is necessary, or in most experiments helpful or desired. Another difference is that while feature locations are static and known to learners in category learning tasks, feature locations are unknown in visual search. While the specifics of what is known and what needs to be learned differ between visual search and category learning, there is a general level at which the tasks are similar: one is presented an image, then one looks around for information in the image, and then makes a response. LAG-1 performs search in a guided fashion (e.g., [126]). It has a selective processing stream—the Learning system which only processes information about features that are fixated—but, it also has a non-selective one: the Visuospatial System, which processes information about basic visual

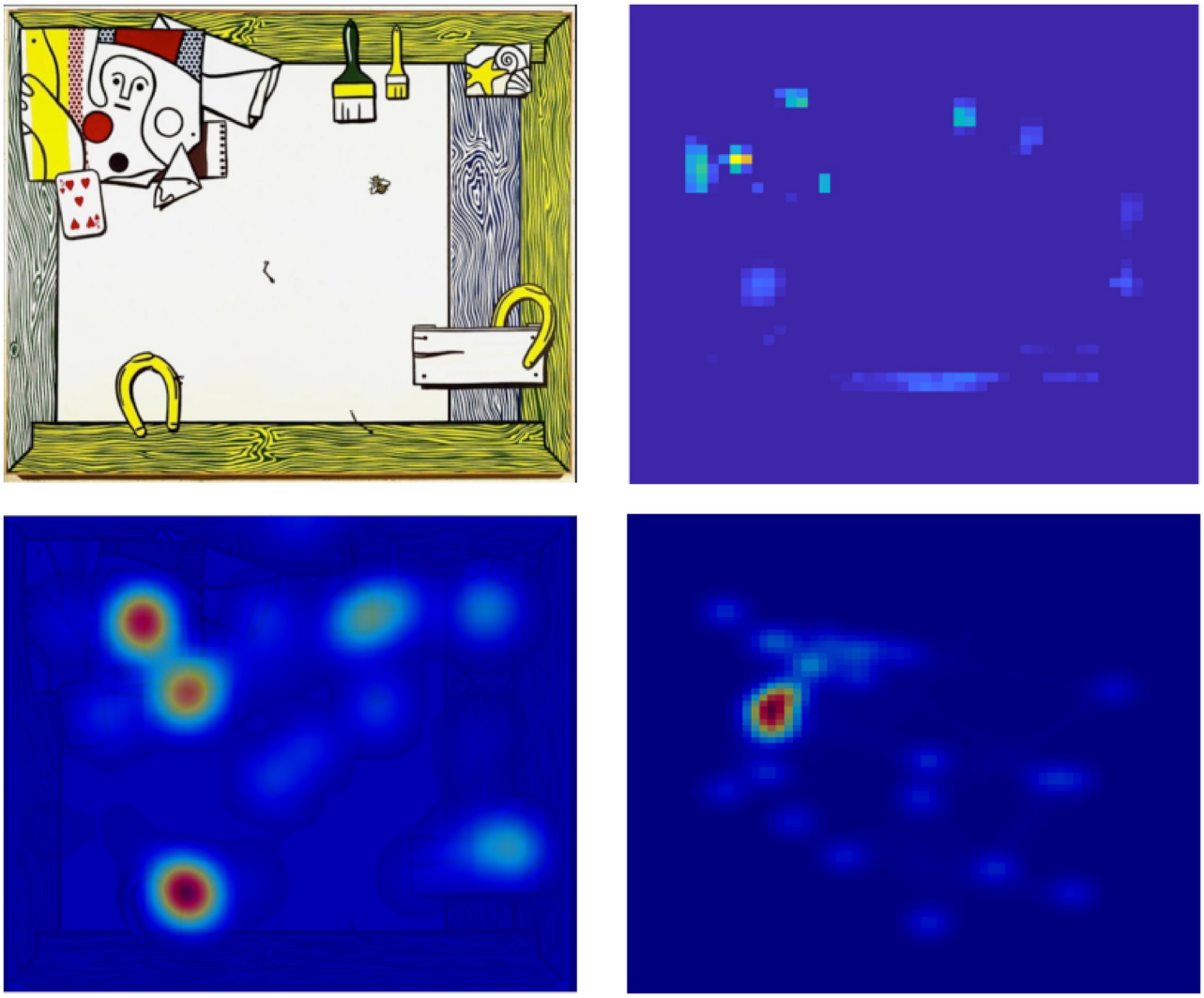


Fig 14. Salience map representation of model during free-viewing. Top left) "Things on the Wall" by Roy Lichtenstein. Top right) a salience map generated by three boundary detection iterations at different spatial scales, integrated across the dimensions of: color, orientation, and intensity. Bottom left) Fixation heat map from a typical human subject. Bottom right) Fixation heatmap from one trial of LAG-1.

https://doi.org/10.1371/journal.pone.0259511.g014

properties (e.g. color) and their locations. Scene statistics and other bottom-up influences on salience are not currently part of the model, but could be incorporated in the same way as a saliency map is used in the free-viewing section above.

The most straightforward way of doing visual search in LAG-1 is to take "target" and "distractor" to be categorical designations (in place of category A and B). Features that indicate target and distractor categories are set in advance (i.e., hard coding the difference into the information gain matrix in this demonstration). In the model there are two recurrent pathways from the category level back into the Visuospatial System. One goes to the locations of particular features on the Spatial Attention Field. In category learning, this linkage between features and their location in space makes sense because participants know the feature locations—which do not move—from the instructions. In a visual search task, however, using this to guide search seems unrealistic because the participant has no knowledge of how particular features connect to particular locations on the Spatial Attention Field. To address this LAG-1 has

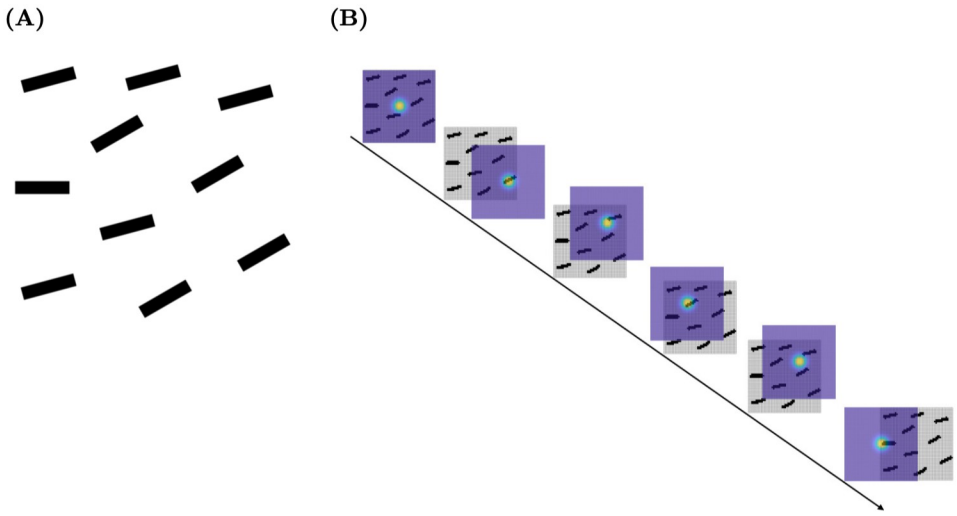


Fig 15. LAG-1 applied to an example visual search problem. A) Visual Field input array recreated from [53]. B) An example trial of LAG-1's visual search scan path. Each frame is centered at the successive fixation location until the target is foveated.

https://doi.org/10.1371/journal.pone.0259511.g015

a second connection: this one links back to the various layers of the Visual Field and can thus produce low-level feature-based attention [86]. Search in LAG-1 is thus guided using this second recurrent pathway. Knowledge about the target one is looking for recurrently boosts activation in the corresponding layers in the Visual Field via gain on the appropriate Feature Expectation Neurons. This second pathway raises the activation of the locations of those features and thus the chance that they will be selected as a visual target from the resultant boost on the Saccade Motor Field, in other words, guided search.

Consider a task in which participants are searching for a vertical line amongst lines of 15˚. In simulating such a task, as we have done with LAG-1 and shown in Fig 15, the feature-specific layers of LAG-1's Visual Field would code orientations of differing degrees (instead of the six different colors we used in the category learning simulations, the six layers used in Simulation 1 can code six different orientations. Top-down signals boost the Feature Expectation Neurons (or field, in this case) associated with the target orientation, leading to increased activation on the corresponding slice of the 3D Visual Field. This boost causes an increased likelihood of an eye movement to that location, relative to unboosted locations, and thus more efficient search. In this implementation of visual search, LAG-1 will look around from item to item, and when an item is fixated the Feature Detection Neurons are activated, which in turn activates the associated category nodes (target or distractor). The model will only stop when the model fixates the target pushing the Category Neuron for target to its threshold, or until the Trial Impatience has driven a click decision (e.g. on target absent trials). LAG-1 does not search in parallel (though again, its choice of target locations is biased toward the boosted location): it must confirm that a particular object is the target by fixating the target. This simple version of visual search, again, implemented with only minor changes from the category learning implementation, undoubtedly will produce the common findings of serial search slopes and RT differences between target present and target absent displays. Search will also be less efficient in the case of search for a conjunction of features: distractor locations that share properties with the target would also receive a boost, decreasing the relative advantage of target over distractor locations, and thus decreasing the target's advantage. Overall, what LAG-1

would predict is that search is more or less efficient depending on the degree of overlap between target and distractor properties [127].

Foundational findings in visual search include accuracy and response time differences as the items of the search array differ by category, such as letters and colours, how they flank one another, and their variability. Several of these seem amenable to simulation by a visual search version of LAG-1. Knowledge effects, such as the influence of target and distractor similarity, fall out of the category learning system [128] as described above, as presumably would category effects [129]. Flanking, wherein search is slowed if distractor orientations are on both sides of the target orientation, also seems to have a relatively straightforward explanation that derives from the spatial dynamics of neural fields. In a more generalized version of our visual search extension, input from a small line length of 15˚ could more strongly activate neurons tuned to that orientation and weakly activate ones of nearby frequencies leading to a center surround bump on the field leading to a flanking effect. In LAG-1, this finding would make sense in that activation of a target could slightly activate orientations that are associated with distractors due to overlap (again, we are imagining a 3d visual field—with x and y as spatial dimensions and z is orientation—and a Feature Detection Field that codes orientation continuously). If distractors flank both sides, there is more overlap, and therefore more competition between target and distractor nodes. This competition would straightforwardly weaken the relative advantage of the target within the recurrent channel. This account makes the assumption that the orientations are close enough that there is some overlap in the activation bumps on the field. The model thus makes the prediction that the flanker effect would be strong in cases where the differences between targets and distractors are small, but may be weak, or non-existent if distractors are very different than targets, and thus have no overlap on the field (see Fig 19 in S1 Appendix).

Some findings do not seem to have a natural explanation given the current structure of LAG-1. The finding that search is slowed if there is more variation in distractors, even if that variation is farther from the target, (a 30˚ target among 15˚ distractors is found faster than a 30˚ target among both 15˚ and 0˚ distractors) has no obvious explanation for example [53, 130]. We can also see no easy method of accounting for search asymmetries such as the result that a 0˚ target among 15˚ distractors is easier to find than the reverse. More substantial modifications to the model seem necessary to account for such phenomena. By design, in most versions of visual search, learning mechanisms do not aid performance. However, in contextual cueing a display in a visual search task is repeated multiple times, and participants in these experiments show improved search times for these repeated displays [131]. LAG-1 has no method of producing this effect in its present form; learning in the model associates features with categories, but the location of the features is presumed to be known (consistent with the category learning tasks to which it is applied in the current work). [132] simulated contextual cueing using a simple two layer network in which distractor locations are associated with the target location upon each presentation of the display. This model restricts learning to those distractors that are nearby the target. Adding a similar mechanism to LAG-1 would be relatively straightforward. An additional field—a Visual Memory Field—could be linked via connections to the Visual Field, such that when the target is found (or more broadly, when reinforcement is presented), connections are strengthened between the Visual Field, and the Visual Memory Field around the area of fixation. Such a modification would more or less embed the [132] model into the larger LAG-1 framework.

Adding broader feature extraction to address covert attending. In the current work, one focus of ours has been on modeling overt attention using eye tracking data. To account for these human behavioural phenomena we have proposed, in simple terms, the processes that lead the eyes to orient toward features for category learning, and in doing so, we have also

specified how features and feature locations receive processing advantages which give them more influence in the complex, interactive processes enabled by the system as a whole. These processing advantages, because they are not directly noticeable, would count as one form of covert attention.

Though the distinction between, and indeed the relationship between, overt and covert attention is captured in the model, there is room for extensions that might account for a larger set of data, but which would not (likely) hamper the performance of the model in category learning tasks. In the current work, we have assumed that eye movements result from a shift of attention, and the uptake of the new information, centered on the Visual Field, is passed on with full clarity to the Learning System. This assumption does not hold for the broader world of visual cognition, and likely not even for some aspects of simple category learning tasks. For example, visual search tasks can be created that allow people to perform them while maintaining fixation to a central point, that is, without eye movements [133]. Further, even in some category learning studies conducted in our lab, we have noticed that certain participants, after hundreds of training trials, continue to perform perfectly, but without fixating their gaze all the features necessary for correct classification. This suggests two things: that information in peripheral vision may be used in object recognition, and that learning seems to influence, or is at least related to, people's ability to do this. The current implementation of LAG-1 cannot account for these findings.

One way of augmenting LAG-1 that might better accord with these findings is to include a separate eye movement field coding a distinct force component [81]. Normally, eye movement signals center the fovea over the target, providing high acuity information to the category learning system. But what happens if an eye movement is suppressed? Top-down signals could prevent the gaze shift by inhibition of the field coding force, yet still allow the Saccade Motor Field to indicate a target, and for the Spatial Attention Field to shift to another location on the Visual Field. This would produce covert attentional shifts, as the eye would be centered on one location, while attention—in the form of the Spatial Attention and Saccade Motor Fields—would be focused somewhere else. In cases where the Spatial Attention Field and the fovea were not over the same location, the information released to the category learning system would depend on the visual acuity at the focus of attention. This extension would describe overt and covert attentional systems in distinct terms, and allow the model to be applied to tasks for which covert and overt systems are strongly dissociated. In the case of covert attention during category learning, experience increases the ability of features to activate categories, eventually allowing even weak peripheral feature values to be strong enough to lead to a classification, and thus, no need for an additional fixation.

One limitation of the current version of LAG-1 is that it requires some calibration to handle different experimental situations. To some extent this is true for most models—in that changing the number of categories, or the number of features requires a new structure—but because of the dynamics in LAG-1, additional calibration is required. In Simulation 2, for example, the input comprised four stimulus dimensions rather than the three used in Simulation 1. This difference changes the competitive dynamics on the Spatial Attention Field, in that the height of a typical bump on the field is lower when there are more features, due to their being more global inhibition on the field. Smaller peaks change the time to reach saccadic thresholds, and influence the highly interactive model components in many other ways. Also, there were only two categories in Simulation 2 whereas there were four categories in Simulation 1. This change modifies the total energy of the Feature Detection, Category, and Feature Expectation layers, and, in turn, leads to different behaviours. It would be better if LAG-1 did not require this kind of tuning as clearly human learners do not. The issue is complex, however, and while it may be easy to take a negative view of LAG-1 on this matter, it is useful to remember that

there are many examples of sensitivity adjustments in humans that take noticeable amounts of time to resolve. The visual system, for instance, cannot process extremely bright light, such as the noonday sun, and adaptation to low light conditions takes time: full dark-adaptation can take 30 minutes, and depends on complex adjustments, not just in the visual cortex, but even in the chemistry of the retina [134]. Also, the adult human visual system has already undergone a long process of self-organization in tuning appropriate expectations based on the demands of the physical environment [135]. We have no such developmental process in LAG-1. We have however, tried to take some tentative steps in thinking about suitable homeostatic processes (e.g., [136]). Developing more general mechanisms for self-calibration will be an important target for future research.

In summary, LAG-1 has broad applicability to the diverse tasks and phenomena within visual cognition. We have taken some first steps here, having implemented a few changes that allow for application of LAG-1 to free-viewing and simple visual search tasks, but more work needs to be done to broaden the model's account to include contextual cueing and covert attention. That such effects can be considered modelling targets gives credence to the idea that integrating attentional, fixational, and associative processes in a single model is an important goal.

Contributions and conclusion

While we have discussed at length the details of the model, the methods used in the simulations and the fits to the data, it is important to emphasize that, at the broadest level, the theme of the present work is *integration*. It proposes connections between different neurofunctional systems, and between different empirical phenomena operating at different timescales. Alan Newell identified four timescales of cognition, emphasizing that biological, cognitive, rational, and social phenomena, are not just different classes of phenomena, but that they occur at different speeds [19]. One important aim, then, is to understand how biological phenomena that are rapid, give rise to slower cognitive phenomena (such as a classification decision), and in turn give rise to still slower phenomena (such as category learning). LAG-1 gives just such an account; one that spans most of Newell's temporal bands of cognition. At the finest scale the model resolves changes in neural activation with short timesteps ($\sim 10^{-2}$ seconds). These neural changes build over time to drive saccadic eye movements ($\sim 10^{-1}$ seconds), which are in turn ordered according to the expected information gain derived from Hebbian learning at the trial level ($\sim 10^{0}$ seconds), shaping the perception of abstract categories learned on the order of $\sim 10^{1}$ seconds. Finally, as associations strengthen and decline over trials, the model describes a continuous relationship between average learning curves and average attentional learning ($\sim 10^{3}$ seconds).

LAG-1 is not an attempt to be a better model of category learning, or of oculomotor control, or of attention. Instead, it is a step toward building bridges *between* established work in these areas by providing a theory of how they might interact. As a result of the integrative approach to the construction of the model, LAG-1 can be applied to more than just category learning performance (e.g. accuracy, transfer probabilities). LAG-1 can simultaneously account for measures that have only occasionally been modelled (e.g. reaction time, probability of fixating features), and further, to measures that have never been directly modelled (e.g. time spent viewing feedback). We fit learning phenomena, and we fit attention phenomena, but most importantly, we account for their interactions (e.g. learning-related changes in fixation

durations). We know of no related model that has shown good qualitative fits to as many different empirical findings as LAG-1.

Even just the existence of a dynamic theory of attention, learning and gaze provides theoretical motivation for novel experiments with temporal and spatial manipulations that could reveal a whole host of psychologically interesting phenomena related to learning and attention. Not only can it motivate new kinds of thinking and research about attention and learning in action, but it can also be constrained by, and improved based on, findings from a wide variety of investigations in the cognitive sciences.

Supporting information

S1 Appendix. Primer to dynamic neural field theory [16, 40, 51, 66, 69, 117, 137–149].
(PDF)

S2 Appendix. Companion equations for the formal description of LAG-1 and neurophysiologial context [16, 99, 108, 150, 151].
(PDF)

S3 Appendix. Fitting procedure [10].
(PDF)

S4 Appendix. Supplementary equation parameters.
(PDF)

S5 Appendix. Parameter tables and best fits.
(PDF)

S6 Appendix. Individual fit visualizations.
(PDF)

Acknowledgments

We would like to thank Bob Rehder and Kim Meier for sharing eye tracking data, as well as all past and present members of the SFU Cognitive Science lab, in particular Caitlyn McColeman and Kat Dolguikh, for their helpful comments.

Author Contributions

Conceptualization: Mark R. Blair, R. Calen Walshe.

Data curation: Jordan Barnes.

Formal analysis: Jordan Barnes, Paul F. Tupper.

Investigation: Jordan Barnes, Paul F. Tupper.

Methodology: Mark R. Blair, R. Calen Walshe, Paul F. Tupper.

Project administration: Jordan Barnes.

Supervision: Mark R. Blair, Paul F. Tupper.

Validation: Mark R. Blair, R. Calen Walshe, Paul F. Tupper.

Visualization: Jordan Barnes.

Writing – original draft: Jordan Barnes, Mark R. Blair, R. Calen Walshe, Paul F. Tupper.

Writing – review & editing: Jordan Barnes, Mark R. Blair, R. Calen Walshe, Paul F. Tupper.

References

1. Johnson KE, Mervis CB. Microgenetic analysis of first steps in children's acquisition of expertise on shorebirds. Developmental Psychology. 1994; 30(3):418–435. https://doi.org/10.1037/0012-1649.30.3.418
2. Johnson K, Mervis CB. Effects of varying levels of expertise on the basic level of categorization. Journal of experimental psychology General. 1997; 126(3):248–277. https://doi.org/10.1037/0096-3445.126.3.248 PMID: 9281832
3. Boster JS, Johnson JC. Form or Function: A Comparison of Expert and Novice Judgments of Similarity among Fish. American Anthropologist. 1989; 91:866–889. https://doi.org/10.1525/aa.1989.91.4.02a00040
4. Pearsall NR, Skipper JEJ, Mintzes JJ. Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. Science Education. 1997; 81(2):193–215. https://doi.org/10.1002/(SICI)1098-237X(199704)81:2%3C193::AID-SCE5%3E3.0.CO;2-A
5. Chi MTH, Feltovich PJ, Glaser R. Categorization and Representation of Physics Problems by Experts and Novices. Cognitive Science. 1981; 5:121–152. https://doi.org/10.1207/s15516709cog0502_2
6. Davies SP. Knowledge restructuring and the acquisition of programming expertise. International Journal of Human-Computer Studies. 1994; 40(4):703–726. https://doi.org/10.1006/ijhc.1994.1032
7. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychological bulletin. 1998; 124(3):372–422. https://doi.org/10.1037/0033-2909.124.3.372 PMID: 9849112
8. Underwood G, Chapman P, Brocklehurst N, Underwood J, Crundall D. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. Ergonomics. 2003; 46(6):629–646. https://doi.org/10.1080/0014013031000090116 PMID: 12745692
9. Gegenfurtner A, Lehtinen E, Säljö R. Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. Educational Psychology Review. 2011; 23(4):523–552. https://doi.org/10.1007/s10648-011-9174-7
10. Rehder B, Hoffman AB. Thirty-something categorization results explained: selective attention, eye-tracking, and models of category learning. Journal of experimental psychology Learning, memory, and cognition. 2005; 31(5):811–29. https://doi.org/10.1037/0278-7393.31.5.811 PMID: 16248736
11. Blair MR, Watson MR, Walshe RC, Maj F. Extremely selective attention: eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. Journal of experimental psychology Learning, memory, and cognition. 2009; 35(5):1196–206. https://doi.org/10.1037/a0016272 PMID: 19686015
12. Biederman I, Shiffrar MM. Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1987; 13(4):640–645. https://doi.org/10.1037/0278-7393.13.4.640
13. Shepard RN, Hovland CI, Jenkins HM. Learning and memorization of classifications. Psychological Monographs: General and Applied. 1961; 75(13):1–42. https://doi.org/10.1037/h0093825
14. Anderson JR. The Adaptive Nature of Human Categorization. Psychological Review. 1991; 98(3):409–429. https://doi.org/10.1037/0033-295X.98.3.409
15. Homa D, Cultice JC. Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1984; 10(1):83–94. https://doi.org/10.1037/0278-7393.10.1.83
16. Kruschke JK. ALCOVE: an exemplar-based connectionist model of category learning. Psychological review. 1992; 99(1):22–44. https://doi.org/10.1037/0033-295X.99.1.22 PMID: 1546117
17. Medin DL, Schaffer MM. Context theory of classification learning. Psychological Review. 1978; 85(3):207–238. https://doi.org/10.1037/0033-295X.85.3.207
18. Nosofsky RM. Attention, similarity, and the identification-categorization relationship. Journal of experimental psychology General. 1986; 115(1):39–61. https://doi.org/10.1037/0096-3445.115.1.39 PMID: 2937873
19. Newell A. Précis of Unified theories of cognition. Behavioral and Brain Sciences. 1992; 15:425–492. https://doi.org/10.1017/S0140525X00069478 PMID: 24924001
20. Ballard DH, Hayhoe MM, Pook PK, Rao RPN. Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences. 1997; 20(04):723–42; discussion 743–67. https://doi.org/10.1017/S0140525X97001611 PMID: 10097009
21. Blair MR, Homa D. Expanding the search for a linear separability constraint on category learning. Memory & cognition. 2001; 29(8):1153–1164. https://doi.org/10.3758/BF03206385 PMID: 11913752
22. Matsuka T, Corter JE. Observed attention allocation processes in category learning. Quarterly journal of experimental psychology (2006). 2008; 61(7):1067–1097. https://doi.org/10.1080/17470210701438194

23. Wood MJ, Fry M, Blair MR. The Price is Right: A High Information Access Cost Facilitates Category Learning. Engineering. 2009;(2004):236–241.

24. Wood MJ, Blair MR. Informed inferences of unknown feature values in categorization. Memory & cognition. 2011; 39(4):666–674. https://doi.org/10.3758/s13421-010-0044-1 PMID: 21264594

25. Meier KM, Blair MR. Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. Cognition. 2013; 126(2):319–25. https://doi.org/10.1016/j.cognition.2012.09.014 PMID: 23099124

26. McColeman CM, Barnes JI, Chen L, Meier KM, Walshe RC, Blair MR. Learning-Induced Changes in Attentional Allocation during Categorization: A Sizable Catalog of Attention Change as Measured by Eye Movements. PLoS ONE. 2014; 9(1):e83302. https://doi.org/10.1371/journal.pone.0083302 PMID: 24497915

27. Blair MR, Watson MR, Meier KM. Errors, efficiency, and the interplay between attention and category learning. Cognition. 2009; 112(2):330–336. https://doi.org/10.1016/j.cognition.2009.04.008 PMID: 19481733

28. Barnes JI, McColeman CM, Blair MR, Walshe RC. RLAttn: An actor-critic model of eye movements during category learning. In: Knauff M, Pauen M, Sebanz N, Wachsmuth I, editors. Proceedings of the 36th Annual Conference of the Cognitive Science Society. vol. 1. Austin, TX: Cognitive Science Society; 2014. p. 1892–1897. Available from: https://mindmodeling.org/cogsci2014/papers/332/.

29. Chen L, Meier KM, Blair MR, Watson MR, Wood MJ. Temporal characteristics of overt attentional behavior during category learning. Attention, Perception, and Psychophysics. 2013; 75(2):244–256. https://doi.org/10.3758/s13414-012-0395-8 PMID: 23151960

30. Watson MR, Blair MR. Attentional Allocation During Feedback: Eyetracking Adventures on the Other Side of the Response. In: Knauff M, Pauen M, Sebanz N, Wachsmuth I, editors. Proceedings of the 30th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society; 2008. p. 345–350.

31. Homa D, Fish R. Recognition reaction time in long-term memory as a function of repetition, lag, and identification of positive and negative search sets. Journal of Experimental Psychology: Human Learning & Memory. 1975; 104(1):71–80.

32. Lamberts K. The time course of categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1998; 24(3):695–711. https://doi.org/10.1037/0278-7393.24.3.695

33. Logan G. An instance theory of attention and memory. Psychological Review. 2002; 109(2):376–400. https://doi.org/10.1037/0033-295X.109.2.376 PMID: 11990323

34. Nosofsky RM, Palmeri TJ. An exemplar-based random walk model of speeded classification. Psychological review. 1997; 104(2):266–300. https://doi.org/10.1037/0033-295X.104.2.266 PMID: 9127583

35. Nelson JD, Cottrell GW. A probabilistic model of eye movements in concept formation. Neurocomputing. 2007; 70(13-15):2256–2272. https://doi.org/10.1016/j.neucom.2006.02.026 PMID: 22787288

36. Rombouts JO, Bohte SM, Martinez-Trujillo J, Roelfsema PR. A Learning Rule That Explains How Rewards Teach Attention. Visual Cognition. 2015; 23(1-2):179–205. https://doi.org/10.1080/13506285.2015.1010462

37. Bourne LE, Guy DE, Dodd DH, Justesen DR. Concept Identification: the Effects of Varying Length and Informational Components of the Intertrial Interval. Journal of experimental psychology. 1965; 69 (6):624–629. https://doi.org/10.1037/h0022018 PMID: 14304315

38. Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J. Parietal neurons encode expected gains in instrumental information. Proceedings of the National Academy of Sciences. 2017; p. 201613844. https://doi.org/10.1073/pnas.1613844114 PMID: 28373569

39. Bisley JW, Goldberg ME. Attention, intention, and priority in the parietal lobe. Annual review of neuroscience. 2010; 33:1–21. https://doi.org/10.1146/annurev-neuro-060909-152823 PMID: 20192813

40. Schneegans S, Spencer JP, Schöner G, Hwang S, Hollingworth A. Dynamic interactions between visual working memory and saccade planning. Journal of Vision. 2014; 10(7):537–537. https://doi.org/10.1167/10.7.537

41. Thelen E, Schöner G, Scheier C, Smith LB. The dynamics of embodiment: A field theory of infant perseverative reaching. Behavioral and Brain Sciences. 2001; 24(1):1–34. https://doi.org/10.1017/S0140525X01003910 PMID: 11515285

42. Kruschke JK, Kappenman ES, Hetrick WP. Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31(5):830–845. PMID: 16248737

43. Schutte AR, Spencer JP, Schöner G. Testing the Dynamic Field Theory: Working Memory for Locations Becomes More Spatially Precise Over Development. Child Development. 2003; 74(5):1393–1417. https://doi.org/10.1111/1467-8624.00614 PMID: 14552405

44. Spencer JP, Schöner G. Bridging the representational gap in the dynamic systems approach to development. Developmental Science. 2003; 6(4):392–412. https://doi.org/10.1111/1467-7687.00295

45. Johnson JS, Spencer JP, Luck SJ, Schöner G. A Dynamic Neural Field Model of Visual Working Memory and Change Detection. Psychological Science. 2009; 20(5):568–577. https://doi.org/10.1111/j.1467-9280.2009.02329.x PMID: 19368698

46. Wilimzig C, Schneider S, Schöner G. The time course of saccadic decision making: Dynamic field theory. Neural Networks. 2006; 19(8):1059–1074. https://doi.org/10.1016/j.neunet.2006.03.003 PMID: 16942860

47. Bastian A, Schöner G, Riehle A. Preshaping and continuous evolution of motor cortical representations during movement preparation. European Journal of Neuroscience. 2003; 18(7):2047–2058. https://doi.org/10.1046/j.1460-9568.2003.02906.x PMID: 14622238

48. Robinson DA. Eye movements evoked by collicular stimulation in the alert monkey. Vision research. 1972; 12:1795–1808. https://doi.org/10.1016/0042-6989(72)90070-3 PMID: 4627952

49. Munoz DP, Fecteau JH. Vying for dominance: dynamic interactions control visual fixation and saccadic initiation in the superior colliculus. Progress in brain research. 2002; 140:3–19. https://doi.org/10.1016/S0079-6123(02)40039-8 PMID: 12508579

50. Gepperth A, Lefort M. Learning to Be Attractive: Probabilistic Computation with Dynamic Attractor Networks. In: 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). Cergy-Pontoise, France: IEEE; 2016. p. 270–277.

51. Faubel C, Schöner G. Learning to recognize objects on the fly: a neurally based dynamic field approach. Neural networks: The official journal of the International Neural Network Society. 2008; 21(4):562–76. https://doi.org/10.1016/j.neunet.2008.03.007 PMID: 18501555

52. Sewell DK, Stallman A. Modeling the Effect of Speed Emphasis in Probabilistic Category Learning. Computational Brain & Behavior. 2020; 3(2):129–152. https://doi.org/10.1007/s42113-019-00067-6

53. Wolfe JM, Horowitz TS, Van Wert MJ, Kenner NM, Place SS, Kibbi N. Low Target Prevalence Is a Stubborn Source of Errors in Visual Search Tasks. Journal of Experimental Psychology: General. 2007; 136(4):623–638. https://doi.org/10.1037/0096-3445.136.4.623 PMID: 17999575

54. Smith PL, Sewell DK. A Competitive Interaction Theory of Attentional Selection and Decision Making in Brief, Multielement Displays. Psychological Review. 2013; 120(3):589–627. https://doi.org/10.1037/a0033140 PMID: 23915085

55. Grieben R, Tekülve J, Zibner SKU, Schneegans S, Schöner G. Sequences of Discrete Attentional Shifts Emerge from a Neural Dynamic Architecture for Conjunctive Visual Search That Operates in Continuous Time. Proceedings of the 40th Annual Conference of the Cognitive Science Society. 2018; p. 1–6.

56. Cantwell G, Riesenhuber M, Roeder JL, Ashby FG. Perceptual Category Learning and Visual Processing: An Exercise in Computational Cognitive Neuroscience. Neural Networks. 2017; 89:31–38. https://doi.org/10.1016/j.neunet.2017.02.010 PMID: 28324757

57. Ashby FG, Alfonso-reese LA, Turken U, Waldron EM. A Neuropsychological Theory of Multiple Systems in Category Learning. Psychological Review. 1998; p. 442–481. https://doi.org/10.1037/0033-295X.105.3.442 PMID: 9697427

58. Trukenbrod Ha, Engbert R. ICAT: a computational model for the adaptive control of fixation durations. Psychonomic bulletin & review. 2014; 21(4):907–34. https://doi.org/10.3758/s13423-013-0575-0 PMID: 24470305

59. Haith AM, Reppert TR, Shadmehr R. Evidence for Hyperbolic Temporal Discounting of Reward in Control of Movements. Journal of Neuroscience. 2012; 32(34):11727–11736. https://doi.org/10.1523/JNEUROSCI.0424-12.2012 PMID: 22915115

60. Hawkins GE, Wagenmakers EJ, Ratcliff R, Brown SD. Discriminating Evidence Accumulation from Urgency Signals in Speeded Decision Making. Journal of Neurophysiology. 2015; 114(1):40–47. https://doi.org/10.1152/jn.00088.2015 PMID: 25904706

61. van Maanen L, Fontanesi L, Hawkins GE, Forstmann BU. Striatal Activation Reflects Urgency in Perceptual Decision Making. NeuroImage. 2016; 139:294–303. https://doi.org/10.1016/j.neuroimage.2016.06.045 PMID: 27355435

62. Goodale M, Milner A. Separate visual pathways for perception and action. Trends in neurosciences. 1992; 15(1):20–5. https://doi.org/10.1016/0166-2236(92)90344-8 PMID: 1374953

63. Grossberg S. Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. Neural Networks. 2013; 37:1–47. https://doi.org/10.1016/j.neunet.2012.09.017 PMID: 23149242

64. Krauzlis RJ, Lovejoy LP, Zénon A. Superior colliculus and visual spatial attention. Annual review of neuroscience. 2013; 36:165–82. https://doi.org/10.1146/annurev-neuro-062012-170249 PMID: 23682659

65. Spencer JP, Austin A, Schutte AR. Cognitive Development Contributions of dynamic systems theory to cognitive development. Cognitive Development. 2012; 27:401–418. https://doi.org/10.1016/j.cogdev.2012.07.006 PMID: 26052181

66. Lehky SR, Sejnowski TJ. Seeing white: Qualia in the context of decoding population codes. Neural computation. 1999; 11(6):1261–1280. https://doi.org/10.1162/089976699300016232 PMID: 10423495

67. Kiani R, Hanks TD, Shadlen MN. Bounded Integration in Parietal Cortex Underlies Decisions Even When Viewing Duration Is Dictated by the Environment. Journal of Neuroscience. 2008; 28(12):3017–3029. https://doi.org/10.1523/JNEUROSCI.4761-07.2008 PMID: 18354005

68. Cohen YE, Andersen RA. A common reference frame for movement plans in the posterior parietal cortex. Nature reviews Neuroscience. 2002; 3(7):553–62. https://doi.org/10.1038/nrn873 PMID: 12094211

69. Freedman DJ, Assad Ja. Distinct Encoding of Spatial and Nonspatial Visual Information in Parietal Cortex. Journal of Neuroscience. 2009; 29(17):5671–5680. https://doi.org/10.1523/JNEUROSCI.2878-08.2009 PMID: 19403833

70. Ferrera VP, Grinband J. Walk the line: parietal neurons respect category boundaries. Nature neuroscience. 2006; 9(10):1207–8. https://doi.org/10.1038/nn1006-1207 PMID: 17001336

71. Sugrue LP, Corrado GS, Newsome WT. Choosing the greater of two goods: neural currencies for valuation and decision making. Nature reviews Neuroscience. 2005; 6(5):363–375. https://doi.org/10.1038/nrn1666 PMID: 15832198

72. Yang T, Shadlen MN. Probabilistic reasoning by neurons. Nature. 2007; 447(7148):1075–1080. https://doi.org/10.1038/nature05852 PMID: 17546027

73. Toth LJ, Assad Ja. Dynamic coding of behaviourally relevant stimuli in parietal cortex. Nature. 2002; 415(6868):165–8. https://doi.org/10.1038/415165a PMID: 11805833

74. Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. Journal of neurophysiology. 2001; 86(4):1916–36. https://doi.org/10.1152/jn.2001.86.4.1916 PMID: 11600651

75. Bisley JW, Goldberg ME. Neuronal activity in the lateral intraparietal area and spatial attention. Science. 2003; 299(5603):81–86. https://doi.org/10.1126/science.1077395 PMID: 12511644

76. Gottlieb J. Attention, learning, and the value of information. Neuron. 2012; 76(2):281–95. https://doi.org/10.1016/j.neuron.2012.09.034 PMID: 23083732

77. Tatler BW, Hayhoe MM, Land MF, Ballard DH. Eye guidance in natural vision: reinterpreting salience. Journal of vision. 2011; 11(5):1–23. https://doi.org/10.1167/11.5.5 PMID: 21622729

78. Salthouse Ta, Ellis CL. Determinants of eye-fixation duration. The American journal of psychology. 1980; 93(2):207–34. https://doi.org/10.2307/1422228 PMID: 7406068

79. Munoz DP, Wurtz RH. Saccade-related activity in monkey superior colliculus. I. Characteristics of burst and buildup cells. Journal of neurophysiology. 1995; 73(6):2313–2333. https://doi.org/10.1152/jn.1995.73.6.2313 PMID: 7666141

80. Gandhi NJ, Keller EL. Spatial distribution and discharge characteristics of superior colliculus neurons antidromically activated from the omnipause region in monkey. Journal of neurophysiology. 1997; 78(4):2221–5. https://doi.org/10.1152/jn.1997.78.4.2221 PMID: 9325389

81. Sparks DL. The brainstem control of saccadic eye movements. Nat Rev Neurosci. 2002; 3(12):952–64. https://doi.org/10.1038/nrn986 PMID: 12461552

82. Castelhano MS, Henderson JM. Stable individual differences across images in human saccadic eye movements. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale. 2008; 62(1):1–14. https://doi.org/10.1037/1196-1961.62.1.1 PMID: 18473624

83. Poynter W, Barber M, Inman J, Wiggins C. Individuals exhibit idiosyncratic eye-movement behavior profiles across tasks. Vision Research. 2013; 89:32–38. https://doi.org/10.1016/j.visres.2013.07.002 PMID: 23867568

84. Munoz DP, Wurtz RH. Fixation cells in monkey superior colliculus. II. Reversible activation and deactivation. Journal of neurophysiology. 1993; 70(2):576–89. https://doi.org/10.1152/jn.1993.70.2.576 PMID: 8410158

85. Liversedge S, Findlay JM. Saccadic eye movements and cognition. Trends in cognitive sciences. 2000; 4(1):6–14. https://doi.org/10.1016/S1364-6613(99)01418-7 PMID: 10637617

86. Martinez-trujillo J, Treue S. Feature-Based Attention Increases the Selectivity of Population Responses in Primate Visual Cortex. Current biology. 2004; 14:744–751. https://doi.org/10.1016/j.cub.2004.04.028 PMID: 15120065

87. Ling S, Liu T, Carrasco M. How spatial and feature-based attention affect the gain and tuning of population responses. Vision Research. 2009; 49(10):1194–1204. https://doi.org/10.1016/j.visres.2008.05.025 PMID: 18590754

88. Di Lollo V. The feature-binding problem is an ill-posed problem. Trends in Cognitive Sciences. 2012; 16(6):317–321. https://doi.org/10.1016/j.tics.2012.04.007 PMID: 22595013

89. van Zoest W, Hunt aR, Kingstone A. Representations in Visual Cognition: It's About Time. Current Directions in Psychological Science. 2010; 19(2):116–120. https://doi.org/10.1177/0963721410363895

90. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience. 1999; 2(1):79–87. https://doi.org/10.1038/4580 PMID: 10195184

91. Nomura EM, Reber PJ. A review of medial temporal lobe and caudate contributions to visual category learning. Neuroscience and Biobehavioral Reviews. 2008; 32(2):279–291. https://doi.org/10.1016/j.neubiorev.2007.07.006 PMID: 17868867

92. Seger CA, Miller EK. Category Learning in the Brain. Annual Review of Neuroscience. 2010; 33(1):203–219. https://doi.org/10.1146/annurev.neuro.051508.135546 PMID: 20572771

93. Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci. 2000; 3(9):946–953. https://doi.org/10.1038/78868 PMID: 10966627

94. Litman L, Awipi T, Davachi L. Category-specificity in the human medial temporal lobe cortex. Hippocampus. 2009; 19(3):308–19. https://doi.org/10.1002/hipo.20515 PMID: 18988234

95. Freedman DJ, Assad Ja. Experience-dependent representation of visual categories in parietal cortex. Nature. 2006; 443(7107):85–8. https://doi.org/10.1038/nature05078 PMID: 16936716

96. Kruschke JK. Base rates in category learning. Journal of experimental psychology Learning, memory, and cognition. 1996; 22(1):3–26. https://doi.org/10.1037/0278-7393.22.1.3 PMID: 8648289

97. Gluck MA. Stimulus Generalization and Representation in Adaptive Network Models of Category Learning. Psychological Science. 1991; 2(January):50–55. https://doi.org/10.1111/j.1467-9280.1991.tb00096.x

98. Sporns O, Gally Ja, Reeke GN, Edelman GM. Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. Proceedings of the National Academy of Sciences of the United States of America. 1989; 86(18):7265–7269. https://doi.org/10.1073/pnas.86.18.7265 PMID: 2780571

99. Daw ND, O'Doherty JP, Dayan P, Dolan RJ, Seymour B. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441(7095):876–9. https://doi.org/10.1038/nature04766 PMID: 16778890

100. Seger Ca. The Basal Ganglia in Human Learning. The Neuroscientist. 2006; 12(4):285–290. https://doi.org/10.1177/1073858405285632 PMID: 16840704

101. Soltani A, Koch C. Visual saliency computations: mechanisms, constraints, and the effect of feedback. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30(38):12831–12843. https://doi.org/10.1523/JNEUROSCI.1517-10.2010

102. Luciw M, Sandamirskaya Y, Kazerounian S, Schmidhuber J, Schöner G. Reinforcement and shaping in learning action sequences with neural dynamics. In: IEEE ICDL-EPIROB 2014—4th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics; 2014. p. 48–55.

103. Komatsu H, Komatsu H, Ideura Y, Ideura Y, Kaji S, Kaji S, et al. Color selectivity of neurons in the inferior temporal cortex of the awake macaque monkey. The Journal of neuroscience: the official journal of the Society for Neuroscience. 1992; 12(2):408–24. https://doi.org/10.1523/JNEUROSCI.12-02-00408.1992 PMID: 1740688

104. Minda JP, Smith JD. Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. Journal of experimental psychology Learning, memory, and cognition. 2002; 28(2):275–292. https://doi.org/10.1037/0278-7393.28.2.275 PMID: 11911384

105. Blair MR, Homa D. As easy to memorize as they are to classify: the 5-4 categories and the category advantage. Memory & Cognition. 2003; 31(8):1293–1301. https://doi.org/10.3758/BF03195812 PMID: 15058690

106. Salvucci DD, Goldberg JH. Identifying fixations and saccades in eye-tracking protocols. Proceedings of the symposium on Eye tracking research & applications—ETRA'00. 2000; p. 71–78.

107. George DN, Kruschke JK. Contextual modulation of attention in human category learning. Learning & behavior. 2012; 40(4):530–41. https://doi.org/10.3758/s13420-012-0072-8 PMID: 22528785

108. Kruschke JK, Johansen MK. A model of probabilistic category learning. Journal of experimental psychology Learning, memory, and cognition. 1999; 25(5):1083–119. https://doi.org/10.1037/0278-7393.25.5.1083 PMID: 10505339

109. Yang LX, Lewandowsky S. Knowledge Partitioning in Categorization: Constraints on Exemplar Models. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004; 30(5):1045–1064. PMID: 15355135

110. Lamberts K. Information-accumulation theory of speeded categorization. Psychological review. 2000; 107(2):227–260. https://doi.org/10.1037/0033-295X.107.2.227 PMID: 10789196

111. Kruschke JK. Locally Bayesian learning with applications to retrospective revaluation and highlighting. Psychological review. 2006; 113(4):677–699. https://doi.org/10.1037/0033-295X.113.4.677 PMID: 17014300

112. Hooge I, Erkelens CJ. Adjustment of fixation duration in visual search. Vision Research. 1998; 38 (9):1295–1302. https://doi.org/10.1016/S0042-6989(97)00287-3 PMID: 9666997

113. Johansen MK, Palmeri TJ. Are there representational shifts during category learning? Cognitive Psychology. 2002; 45(4):482–553. https://doi.org/10.1016/S0010-0285(02)00505-4 PMID: 12480477

114. Love BC, Medin DL, Gureckis TM. SUSTAIN: a network model of category learning. Psychological review. 2004; 111(2):309–32. https://doi.org/10.1037/0033-295X.111.2.309 PMID: 15065912

115. Rosch E, Mervis CB. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology. 1975; 7(4):573–605. https://doi.org/10.1016/0010-0285(75)90024-9

116. Smallwood J, Schooler JW. The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. Annual Review of Psychology. 2015; 66(1):487–518. https://doi.org/10.1146/annurev-psych-010814-015331 PMID: 25293689

117. Rao RP, Sejnowski TJ. Spike-timing-dependent Hebbian plasticity as temporal difference learning. Neural computation. 2001; 13(10):2221–37. https://doi.org/10.1162/089976601750541787 PMID: 11570997

118. Rehder B, Hoffman AB. Eyetracking and selective attention in category learning. Cognitive psychology. 2005; 51(1):1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001 PMID: 16039934

119. Roberts S, Pashler H. How persuasive is a good fit? A comment on theory testing. Psychological Review. 2000; 107(2):358–367. https://doi.org/10.1037/0033-295X.107.2.358 PMID: 10789200

120. Gottlieb J. Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. Cortex. 2017; p. 1–11. https://doi.org/10.1016/j.cortex.2017.08.019 PMID: 28919222

121. Stritzke M, Trommershäuser J, Gegenfurtner KR. Effects of salience and reward information during saccadic decisions under risk. Journal of the Optical Society of America A. 2009; 26(11):B1. https://doi.org/10.1364/JOSAA.26.0000B1 PMID: 19884911

122. Walther DB, Koch C. SaliencyToolbox; 2016. Available from: http://www.saliencytoolbox.net/index.html.

123. Schreij D, Owens C, Theeuwes J. Abrupt onsets capture attention independent of top-down control settings. Perception & Psychophysics. 2008; 70(2):208–218. https://doi.org/10.3758/PP.70.2.208 PMID: 18372744

124. Jonides J, Yantis S. Uniqueness of abrupt visual onset in capturing attention. Perception & Psychophysics. 1988; 43(4):346–354. https://doi.org/10.3758/BF03208805 PMID: 3362663

125. Carrasco M. Visual attention: The past 25 years. Vision Research. 2011; 51(13):1484–1525. https://doi.org/10.1016/j.visres.2011.04.012 PMID: 21549742

126. Wolfe JM. Guided Search 2.0 A revised model of visual search. Psychnomic Bulletin & Review. 1994; 1(2):202–238. https://doi.org/10.3758/BF03200774 PMID: 24203471

127. Wolfe JM. What can 1 million trial tell us about visual search? Pyschological Science. 1998; 9(1):33–39.

128. Duncan J, Humphreys GW. Visual Search and Stimulus Similarity. Psychological Review. 1989; 96 (3):433–458. https://doi.org/10.1037/0033-295X.96.3.433 PMID: 2756067

129. Wolfe JM, Friedman-Hill SR. Visual search for oriented lines: The role of angular relations between targets and distractors. Spatial Vision. 1992; 6(3):199–207. https://doi.org/10.1163/156856892X00082 PMID: 1419930

130. Rosenholtz R. Visual Search for Orientation among Heterogeneous Distractors: Experimental Results and Implications for Signal-Detection Theory Models of Search. Journal of Experimental Psychology: Human Perception and Performance. 2001; 27(4):985–999. PMID: 11518158

131. Chun MM, Jiang Y. Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. Cogn Psychol. 1998; 36(1):28–71. https://doi.org/10.1006/cogp.1998.0681 PMID: 9679076

132. Brady TF, Chun MM. Spatial Constraints on Learning in Visual Search: Modeling Contextual Cuing. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33(4):798–815. PMID: 17683229

133. Monosov IE, Thompson KG. Frontal Eye Field Activity Enhances Object Identification During Covert Visual Search. Journal of Neurophysiology. 2009; p. 3656–3672. https://doi.org/10.1152/jn.00750.2009 PMID: 19828723

134. Barlow BHB, Fitzhugh R, Kuffler SW. Dark Adaptation, Absolute Threshold and Purkinje Shift in Single Units of the Cat's Retina. J Physiol. 1957; 137(137):327–337. https://doi.org/10.1113/jphysiol.1957.sp005816 PMID: 13463770

135. Perone S, Spencer JP. Autonomy in action: linking the act of looking to memory formation in infancy via dynamic neural fields. Cognitive science. 2013; 37(1):1–60. https://doi.org/10.1111/cogs.12010 PMID: 23136815

136. Jenkins GW, Barnes JI, Tupper P, Blair MR. A modeling link between cognitive and biological homeostasis. In: Proceedings of the 39th Annual Conference of the Cognitive Science Society; 2017. p. 588–593.

137. Kakade S, Dayan P. Dopamine: generalization and bonuses. Neural Networks. 2002; 15(4-6):549–559. https://doi.org/10.1016/S0893-6080(02)00048-5 PMID: 12371511

138. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986; 323(6088):533–536. https://doi.org/10.1038/323533a0

139. Eliasmith C, Stewart TC, Choo X, Bekolay T, DeWolf T, Tang Y, et al. A large-scale model of the functioning brain. Science. 2012; 338(6111):1202–5. https://doi.org/10.1126/science.1225266 PMID: 23197532

140. Georgopoulos AP, Schwartz A, Kettner R. Neuronal population coding of movement direction. Science. 1986; 233(4771):1416–1419. https://doi.org/10.1126/science.3749885 PMID: 3749885

141. Georgopoulos AP, Lurito J, Petrides M, Schwartz A, Massey J. Mental rotation of the neuronal population vector. Science. 1989; 243(4888):234–236. https://doi.org/10.1126/science.2911737 PMID: 2911737

142. Georgopoulos AP, Taira M, Lukashin A. Cognitive neurophysiology of the motor cortex. Science. 1993; 260(5104):47–52.

143. Jancke D, Erlhagen W, Dinse HR, Akhavan AC, Giese M, Steinhage A, et al. Parametric population representation of retinal location: neuronal interaction dynamics in cat primary visual cortex. The Journal of neuroscience: the official journal of the Society for Neuroscience. 1999; 19(20):9016–28. https://doi.org/10.1523/JNEUROSCI.19-20-09016.1999 PMID: 10516319

144. Lee C, Rohrer W, Sparks DL. Population coding of saccadic eye movements by neurons in the superior colliculus. Nature. 1988; 332(6162):19–21.

145. McPeek RM, Maljkovic V, Nakayama K. Saccades require focal attention and are facilitated by a short-term memory system. Vision Research. 1999; 39:1555–1566. https://doi.org/10.1016/j.visres.2008.05.025 PMID: 10343821

146. Theeuwes J, Belopolsky A, Olivers CNL. Interactions between working memory, attention and eye movements. Acta Psychologica. 2009; 132(2):106–114. https://doi.org/10.1016/j.actpsy.2009.01.005 PMID: 19233340

147. Hebb DO. The Organization of Behavior. 5th ed. New York, New York, USA: John Wiley & Sons Inc; 1949.

148. Ferster D. Is Neural Noise Just a Nuisance? Science. 1996; 273(5283):1811–12.

149. Müller NG, Mollenhauer M, Rösler A, Kleinschmidt A. The attentional field has a Mexican hat distribution. Vision research. 2005; 45(9):1129–37. https://doi.org/10.1016/j.visres.2004.11.003 PMID: 15707921

150. Hoffman JE, Subramaniam B. The role of visual attention in saccadic eye movements. Perception & psychophysics. 1995; 57(6):787–95.

151. Craig S, Lewandowsky S, Little DR. Error discounting in probabilistic category learning. Journal of experimental psychology Learning, memory, and cognition. 2011; 37(3):673–87. https://doi.org/10.1037/a0022473 PMID: 21355666

S1 Appendix. Primer to dynamic neural field theory.

This section explains the differential equations that describe the time evolution of the dynamical variables of the model. It is our intention here to make the modelling more accessible to readers unfamiliar with dynamical systems.

Neurally inspired computational cognitive models have been used to support research at a variety of different ontological levels including: transmitter systems [137], single neuron computation [117], error-driven learning (e.g. [16, 138]) as well as whole-brain models [139]. By writing differential equations for the activations of the units in these models, time-dependent behaviours can be modelled. This is done by stating how the activations of the units in the system change through time in the presence of some input (which may also change though time).

In LAG-1 we use two main types of processing units: single neurons and fields. While the concept of a neuron in a cognitive model may be familiar, the use of fields is less common. One way to think of a field is as an organized collection of neurons that work together to represent some variables or relationships of interest. LAG-1 uses neural fields primarily to represent spatial dimensions, but fields can also be used to represent other metric dimensions such as color, size or aspect ratio [51]. Individual neurons have preferential responses to particular values along those dimensions. The emergent activity in a population can be thought of as a localized bump of activation in a space that monotonically orders the individual neurons by their preferences. In other words, the location of the bump on the field is what is representing the information. Because the representations are distributed across many neurons, this method of interpretation is referred to as population coding [140, 141, 142]. Neural population codes have been reconstructed throughout the brain. Such population codes contribute to our fundamental understanding of parietal [69], temporal [66], striate [143] and mid-brain [144] regions to name a few—all of which are relevant to the functioning of LAG-1.

Single neuron equations

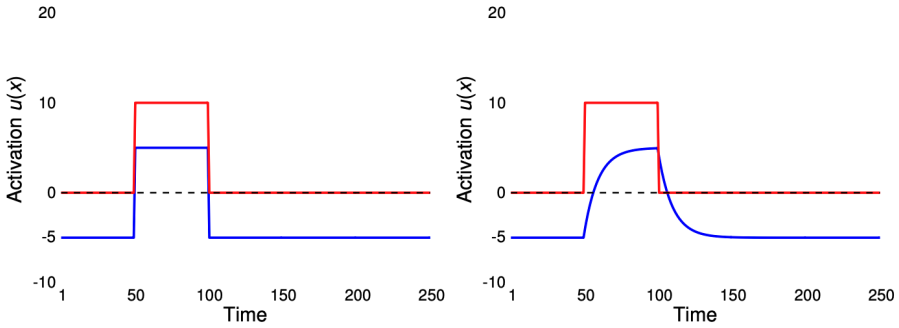
We now look at the mathematics that specifies single neuron activation in LAG-1. An individual neuron's change in activity across an interval of time is described by the equation:

$$\dot{u}(t) = -u(t) + h + S(t) \tag{1}$$

In Equation 1, $\dot{u}(t)$ (also written as du/dt) defines the rate of change in activation of a neuron u, at a point in time, t. Over a short interval of time, δt, the amount of change in u is approximately $\dot{u}(t)\delta t$. The parameter h is a constant, indicating a resting level, or baseline firing rate of the neuron u. Exogenous input to u at time t is represented by the specific input term $S(t)$. When there is no input to the model, $S(t) = 0$, the rate of change of u will also be zero so long as $u(t) = h$, i.e. a neuron at its resting level will stay at its resting level if there is no external input.

If the activation of u is not at its resting level, and no other input is provided, then the value of $u(t)$ will converge exponentially toward h as a function of time. If $S(t)$ changes from 0, it will displace the activation from u from its resting level h (either up or down). In Figure S1, the behaviour of such just such a neuron-like variable is depicted, given some non-zero input for a period of time, i.e. $S(t) = 10$ for $50 \leq t \leq 100$, and $S(t) = 0$ otherwise. As can be seen, the neuron's activation stays at its resting level, -5, then jumps up to the sum of the input and the resting level ($(10) + (-5) = 5$). It maintains this higher activation until the input is turned off, at which point the neuron's activation immediately falls back down to its resting level.

Implicit in Equation 2 is a time scale for its changes in activation. The time scale in Figure S1A is on the order of one unit of time, meaning that the transition in activation



(A) Timescale obscured. **(B)** Timescale expanded.

Figure 16. Characteristic timescales of activation. Activity of the neuron u plotted in blue and input to this neuron plotted in red. A) Activation of the neuron versus time when $\tau = 1$. B) Activation of the neuron versus time when $\tau = 10$.

is not clearly visible in this plot. We need to introduce a parameter to control this time scale in order to be able to match it to human behaviour. The time scaling parameter to do this is τ, included on the left-hand side of Equation 1, as shown in Equation 2.

$$\tau \dot{u}(t) = -u(t) + h + S(t) \tag{2}$$

The τ parameter dictates how long it takes for a change in input to impact on the variable. In this definition, larger values of τ mean that the system evolves more slowly. This can be seen by dividing the equation by τ and observing that the larger τ is, the less the rate of change of u is. Figure S1B shows the difference in activation over time, as scaled by τ.

Numerical simulation of the equations

The dynamics produced by Equation 2 are simple enough that we can write down an exact solution for $u(t)$. In more sophisticated models however, this is usually not possible. Instead, we must obtain a sequence of approximate values of the variables over a sample of time points. To do this, a step length, δt, must be chosen that specifies the constant distance between the computed points in time. Starting with an initial activation value for $u(0)$, the approximate values at each point in time are obtained as $u(\delta t)$, $u(2\delta t)$, $u(3\delta t)$, and so forth. Choosing a smaller δt provides a more accurate solution, but is more computationally expensive, as the number of steps that must be taken to simulate over a time interval of length T is $T/\delta t$. A larger value for δt means that the solutions can be obtained faster, but risks being inaccurate. Given an initial value for $u(t)$, an approximation for its value at $u(t + \delta t)$ is given by:

$$u(t) + \frac{\delta t}{\tau}(-u(t) + h + S(t)) \tag{3}$$

Notice that the new value of u given by Expression 3 is equal to the value of u at time t plus δt times the derivative of u. A good rule of thumb in selecting δt is that the timesteps need to be smaller than the shortest time scale in the problem at hand. For LAG-1 we discretized time to just below what was needed for the model to make accurate saccadic eye movements: one unit of model time is 10 ms of real time when $\delta t = 1$. If a human participant tends to fixate stimulus features for 300ms then the model should be similarly fixating for 30 time steps.

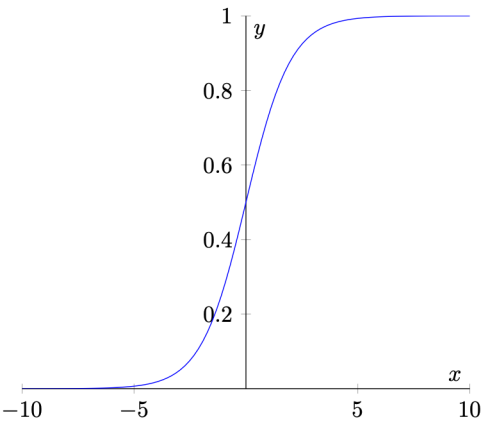


Figure 17. The sigmoid activation function. The sigmoid function defined in Equation 5. This function asymptotes to 0 and 1 as $x \to -\infty$ or $x \to +\infty$ respectively: $(f_{\beta=1,u_0=0})$.

Endogenous input and self sustaining activation

Recurrent connections can be added to systems like that defined in Equation 2. This can allow for a kind of memory.

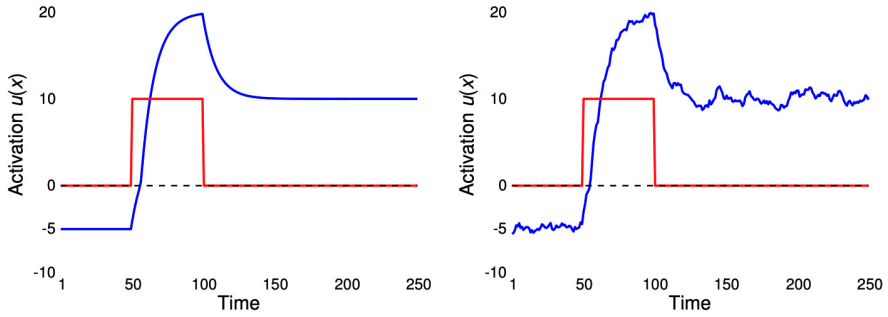
$$\tau \dot{u}(t) = -u(t) + h + S(t) + cf(u(t)) \tag{4}$$

In Equation 4, a variable $f(u(t))$, representing input to the neuron based on its own level of activation, is added to Equation 2, scaled by a constant parameter c. This input consists of the activation of the unit passed through a non-linear sigmoid function f, which takes any real valued number as input and returns a value between 0 and 1 as output. This is a normal operation, that keeps activation bounded while also having some useful mathematical properties for complex learning. Figure 17 plots the behaviour of an example sigmoid function defined in Equation 5 below. The sigmoid function is used in many places in LAG-1 with different values of the parameters β, u_0.

$$f_{\beta,u_0}(u) = \frac{1}{1 + e^{-\beta(u-u_0)}} \tag{5}$$

The introduction of the recurrent term for u allows the possibility of two distinct equilibrium states. When $u(t)$ is low, the input from the $f(u(t))$ term will also be small and may not significantly influence the overall level of u. When $u(t)$ is high enough however, the input from $f(u(t))$ will abruptly increase because of the non-linear sigmoid transformation, resulting in yet larger values of $u(t)$, and furthering the tendency to maintain $u(t)$ at a higher level of activation. The low activation state and the high activation states are referred to as stable fixed points and can be considered as distinct states: "off" and "on". The system may be pushed from off to on (and vice versa) from external input or sometimes just with noise. Given a recurrent input like this, the system can effectively store a single bit of information. Adjusting the resting level h changes the relative ease with which the system may transition between these fixed points.

In Figure 18, the system defined in Equation 4 is simulated, such that the activation of the neuron starts from its resting level, and is subsequently excited by some exogenous input. While this is happening, the influence of the recurrent excitatory input is also increasing. When the exogenous input is removed, the output of $f(u(x,t))$ has reached a point where it alone can sustain the system's activation preventing it



(A) Neuron with self-excitation. **(B)** Neuron with self-excitation and noise.

Figure 18. Dynamic variables with noise. A) Activation $u(t)$ (blue) and exogenous input $S(t)$ (red) as a function of time, for the system described in Eq. 2. B) The same system but with a small noise term added.

from falling back down to its resting level. In LAG-1, the Feature Detection Neurons, among others, have this property, allowing them to retain information about stimulus features as they are viewed over the course of a trial, even when the model has stopped looking at it [40, 145, 146]. Descriptions of neural information processing with these kinds of dynamics can be traced at least back to Hebb's fundamental hypotheses about the cell assembly and its relation to eye movements [147].

Noise Part 1

Variability in neural signaling is an inherent aspect of real-world neural networks [148]. To model noise, we can build on the systems defined earlier by incorporating a noise term, $\zeta(t)$.

$$\tau\dot{u}(t) = -u(t) + h + S(t) + cf(u(t)) + \zeta(t) \tag{6}$$

In Equation 6, $\zeta(t)$ represents Gaussian white noise: that is, noise having equal amplitude over all frequencies (scaled by $\tau^{1/2}$ for reasons that will be explained in the next section). Figure 18 shows the effect of noise added to the same system simulated in Figure 16. Noise introduces a certain amount of behavioural stochasticity by altering the time it takes for action thresholds to be met, for example, in choosing the next feature to look at or the moment to make a category decision.

One-dimensional field equations

The dynamics so far described for a single idealized neuron can be extended to describe neural fields. The activation changes at locations along the field are calculated just like those of single neurons except that there are excitatory and inhibitory interactions with the other locations in the field to include. The differing spatiotemporal contributions of inhibitory and excitatory components of a neuron's receptive field, combine to create a Difference-of-Gaussians (DoG) shaped kernel, depicted in Figure 19.

The expression in Equation 7 for $\dot{u}(x,t)$ defines the rate of change of the activation u at a specific position x. The field equation incorporates the same kinds of inputs as the single neuron defined earlier, including: a linear decay term $-u(x,t)$, a resting level h, an exogenous input $S(x,t)$, and a noise term, $\zeta(x,t)$. The important conceptual difference is that the variable x does not refer to a particular neuron, but rather an abstract location in a feature space relevant to a task. The recurrent term $cf(u(x))$ earlier included in the single neuron equation is substituted in Equation 7 with the

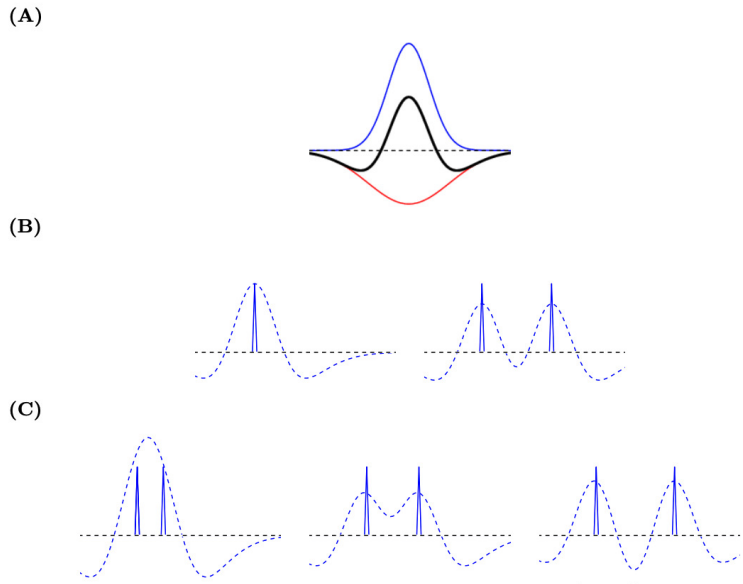


Figure 19. The making of a Difference of Gaussians (DoG) distribution in 1-dimension. A) Blue indicates excitatory input, red indicates inhibitory input and black represents their summation. B) The figure on the left shows a single specific input to a field as the solid blue line, and the result of convolution with the DoG kernel as the dashed blue line. The figure on the right shows the inhibitory effect that multiple inputs can exert on each other once convolved with this kernel. C) The degree of lateral excitation or inhibition between activations on the field depends on the distance between them. Locations close to one another, as on the left, can merge and increase in magnitude through local excitation. The activity at locations where the inhibitory component of the kernel is most pronounced, as in the middle figure, results in particularly reduced activation after convolution. Finally, more distant locations on the field show a more modest reduction of activity as the tails of the kernel become less and less negative, as shown in the rightmost figure.

convolution of a kernel, w, with the sigmoid of u, $f[u]$. This term incorporates the effects of excitation and inhibition on $u(x,t)$ from the other locations on the field, $u(x',t)$.

$$\begin{aligned}\tau\dot{u}(x,t) = &- u(x,t) + h + f(S(x,t)) + \zeta(x,t) \\ &+ \int w_u(x-x')f[u(x',t)]dx'\end{aligned} \tag{7}$$

The kernel is given by the difference of Gaussians (DoG) described in Equation 8.

$$w_u(x-x') = k_e \exp\left[\frac{-\left(x-x'\right)^2}{2{\sigma_e}^2}\right] - k_i \exp\left[\frac{-\left(x-x'\right)^2}{2{\sigma_i}^2}\right] \tag{8}$$

The kernel parameters, k_e and k_i, scale the magnitude of the excitatory and inhibitory contributions depicted in blue and red of Figure 19. The breadth of these excitatory and inhibitory contributions is parameterized by σ_e and σ_i to produce the final integrated weighting.

The net effect of the kernel convolution is similar to that of recurrent input in the single neuron case in that the excitatory distances of the kernel allow for two stable levels of activation on the field even in the absence of exogenous input if the recurrent

term is large enough. A single localized input will generate a bump of activity centered about that input after convolution with the kernel. When multiple points are already active on the field, kernel convolution has varying effects, dependent on the distance of the points relative to the shape of the kernel. Active locations near to one another combine to form a single larger peak. Active locations just a little further from one another may fall in the inhibitory-well of the kernel, yielding distinct but strongly dampened peaks. When the inputs are sufficiently distant, they avoid the well of strong inhibition they suppress each other to a smaller degree. The DoG kernel used within the model has some empirical support from studies of attention [149].

Two-dimensional field equations

Generalizing these equations to higher dimensions is straightforward with the difference being that a higher dimensional integral convolution is required to account for the lateral interactions within the population. This is computationally more demanding but adds little to the complexity of the mathematics. In Figure 20, the continuous version of a two dimensional kernel is depicted, alongside a less computationally intensive variant similar to what is used here, where distant locations of the field are considered to have no effect on each other.

$$\begin{aligned} \tau \dot{u}(x,y,t) = & -u + h_u + S^*(x,y,t) + \zeta(x,y,t) \\ & + \int\int w_u(x-x', y-y') u^*(x',y',t) dx' dy' \end{aligned} \tag{9}$$

The single dimensional field equation (Equation 7) is generalized to two dimensions in Equation 9.

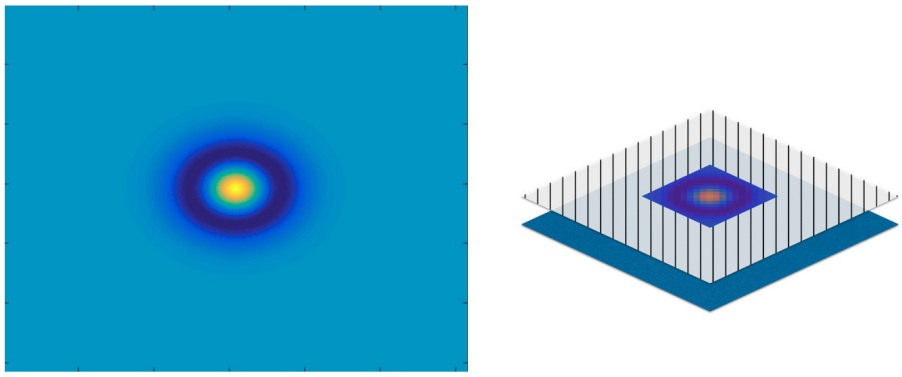
We now can unpack one of the actual equations of the model, using the concepts we have introduced so far. Equation 11 defines the dynamics of the activity of the Visual Field. Remember that the experiment input has 3 dimensions: two spatial, (x, y), and one featural, (z). The expression for the Visual Field however is defined for only the two spatial dimensions. The value of the feature on the third dimension is used only when a location having a feature is actually fixated. The generic input previously referred to as S^* is now replaced with the actual terms for the external inputs to the field: in this case, the stimulus input $I_V(x, y, t)$, the acquired feature expectations $u^*_{\mathrm{ft}_{\mathrm{exp}}}(x, y, t)$ and the appearance of the feedback button $u^*_{\mathrm{fb}_{\mathrm{exp}}}(x, y, t)$.
The linear decay $-u_V(x, y, t)$ term, resting level h_{u_V}. and noise $v_\zeta(x, y, t)$ remain consistent with earlier definitions. The c parameters are simple scaling parameters and are discussed further in the formal model description in the next section.

Noise Part 2

There are two considerations when including noise in these equations that are easily overlooked: noise correlations and noise scaling. Noise across multiple spatial dimensions is correlated by the smoothing effect of the spatial kernel at every step through time whereas we set noise in time as being uncorrelated. This is implemented by starting with complete space-time white noise, having the property that the integral of it over an x-y-t cube of dimensions, δx, δy, δt, is normalized as a Gaussian random variable with mean zero and variance $\delta x \delta y \delta t$ (i.e. the volume of the box in $x - y - t$ space).

$$\zeta(x,y,t) = \tau^{1/2} \int\int w_\xi(x-x', y-y') \xi(x,y,t) dx' dy' \tag{10}$$

In Equation 10, the genuine white noise process, $\xi(x, y, t)$, covering the dimensions of simulation, is smoothed by convolution with the noise kernel w_ξ. The time scale term, $\tau^{1/2}$, ensures the correct scaling in time for different values of τ.



(A) Full, continuous kernel. **(B)** Reduced, discretized kernel.

Figure 20. Two dimensional difference of Gaussian kernels. A) Heat map of an idealized DoG in two dimensions. B) In practice this function is approximated by the sum of a localized function within a small square, and 0 outside of it.

S3 Appendix. Fitting procedure.

In Simulation 1, the target fit vector contained four measures (accuracy, fixation count, the probability of fixating irrelevant information on a trial, and the average fixation durations). The data on each of these measures was summarized by an initial intercept and a slope of change in the measure over the experiment, yielding a target vector with eight elements for each human subject. In Simulation 2, the vector included the individual subject transfer probabilities (for all 16 stimuli), their learning points (the trial number on which the participant began two consecutive error-free blocks) and a measure of the changes in the allocation of attention to stimulus features of high and low diagnosticity. We were unable to precisely reconstruct all of the individual subject data reported in [10] for the attentional allocation component of the fit measure so we instead used the reported averages in three measures of attentional allocation. The first two numbers were the average starting and ending fixation counts to the individual features. The third component was the difference between the least informative (Feature two) and the most informative features (Feature one) on this measure ($\sim$0.3). The target vector for Simulation 2 thus had 19 elements. For both vectors we used a residual sum of squares measure of error to calculate the best model for each subject (for precise details see Table 8).

There are, as indicated in the model's formal description, many parameters that need to be specified in the model (for example, sigmoids and kernels need to be given a precise shape in order to perform the calculations). We fixed these in advance of the simuluations by picking sensible values and then, where necessary, tweaked things till the model had stable performance that roughly resembled human looking and learning. No effort was made at this stage to fit the myriad of findings we address in the simulation; stability was our only concern. Once the model was performing sensibly, everything was fixed and we proceeded with the fitting procedure.

We began the process of finding the best fitting parameter values by running the model with a plausible initial set of low, medium and high levels of the three free parameters: generating a parameter cube with 27 data points. At each of these levels, we ran the model multiple times and took its average behaviour to account for the stochasticity in the model. A measure of fit was generated by comparing every individual human subject with the data of the model at those 27 locations in parameter space. If the best fitting parameters for a particular individual subject was adjacent to unexplored regions of the parameter space, then the cube was extended on that side, and new simulations were run. Simulations were ended only when every individual was surrounded by 26 worse fitting points.

In order to plot the simulations for comparison with the human subjects we needed to collapse the group of simulations at the best fitting set of parameter values for each subject. To do this we simply averaged over the the experiment length data for the measures of interest. We also show the model's variability in the error bars we report. The result is that we have model performance data comparable to that of each human participant, and can plot the model data in exactly the same way as the human data.

S2 Appendix. Companion equations for the formal description of LAG-1 and neurophysiologial context.

Visual Field

$$\begin{aligned}\tau\dot{u}_V(x,y,t) = & -u_V(x,y,t) + h_{V,1} \\ & + \sum_z I_V(x,y,z,t) \\ & + \int\int w_V(x-x',y-y',t)u^*_V(x',y',t)dx'dy' \\ & + c_{13}\sum_z u^*_{\mathrm{ft}_{\mathrm{exp}}}(x,y,z,t) \\ & - c_{14}\sum_z \int\int u^*_{\mathrm{ft}_{\mathrm{exp}}}(x',y',z,t)dx'dy' \\ & + u^*_{\mathrm{fb}_{\mathrm{exp}}}(t) + \zeta_v\end{aligned} \tag{11}$$

The data structure representing the raw stimulus, $I(x,y,z,t)$, specifies x,y,t locations in retintopic space at each moment in trial time, for feature values $j \in z$. The strength of the feature value input is mediated by the saccadic masking: $1 - c_3u^*_r(t)$, where $u_r(t)$ is the value of the Saccade Initiation Neuron (high values of u_r suppress visual input). During the feedback phase of a trial, the correct category is also presented to the field, represented here as an additional feature, $u_{\mathrm{fb}_{\mathrm{exp}}}(t)$ and modulated by a signaling function $\mathbb{1}_{\mathrm{phase=feedback}}$ that indicates whether or not the feedback feature has been fixated.

$$\begin{aligned}I_V(x,y,z,t) = (1 - c_3u^*_r(t))\big(&c_2I(x,y,z,t) \\ &c_4\mathbb{1}_{\mathrm{phase=feedback}}u^*_{\mathrm{fb}_{\mathrm{exp}}}(t)\big)\end{aligned} \tag{12}$$

Spatial Attention Field

$$\begin{aligned}\tau\dot{u}_A(x,y,t) = & -u_A(x,y,t) + h_{A,18} \\ & + \textstyle\sum_z c_{19}u^*_{\mathrm{ft}_{\mathrm{exp}}}(z,t)u^*_V(x,y,z,t) \\ & + c_{25}u^*_{\mathrm{fb}_{\mathrm{exp}}}(t) \\ & + c_{20}u^*_{\mathrm{fix}}(t)\mathcal{G}_{\mathcal{A}}(F,\sigma_{22})\mathbb{1}_{\mathrm{sacc}} \\ & - c_{21}u_{\mathrm{ior}}(x,y,t) \\ & - c_{28}u^*_g(t)\mathcal{G}_A(F,\sigma_{22}) \\ & - c_{29}u^*_r(t)u^*_A(x,y,t) \\ & + c_{45}\mathbb{1}_{\mathrm{sacc}}u_{A\leftarrow V}(x,y,t) \\ & + u^*_{A\leftarrow M}(x,y,t) \\ & + \zeta_A(x,y,t) \\ & + \int\int \big(w_{A,A}(x-x',y-y',t) \\ & \qquad u^*_A(x',y',t)dx'dy'\big) \\ & - c_{38}\int\int u^*_A(x',y',t)dx'dy'\end{aligned} \tag{13}$$

On the Spatial Attention Field, defined in Equation 13, input from the Fixation Neuron, u_{x} (described below) maintains pressure to continue fixating the current feature. This

input is transformed from a scalar to a Gaussian with the same width as the fovea, $\mathcal{G}_A(F, \sigma_{22})$, centered at the extant locus of fixation, F. These fixation forces compete with a polynomial increase in pressure to look away, mediated by a Gaze Change Neuron, $u_g(t)$. A saccade not only reduces the level of input from the Visual Field, denoted by an indicator function $\mathbb{1}_{\text{sacc}}$, but also reduces the total energy of the Spatial Attention Field as a proportion of the activity of the Saccade Initiation Neuron, $u_r^*(t)$. Excitatory input from the Saccade Motor Field $u_{A \leftarrow M}(x, y, t)$ heightens attention at the location of a planned saccade just as observed experimentally [150]. Category-related biases for particular locations emerge from the continuous reentrant activity of Feature Detectors, Category Neurons, and Feature Expectation Neurons, as they project into the Spatial Attention and Visual Fields. Simple associative learning between Feature Detector Neurons (that implement visual working memory) and Category Neurons, is enough to effectively bind information obtained over a series of fixations, enabling a pattern completion process. Correlations between particular feature values and particular categories acquired over the course of many learning trials scale the strength of visual inputs via the top-down projections of the Feature Expectation Neurons, $u_{\text{ft}_{\text{exp}}}(x, y, t)$ (as well as the feedback button, $u^*_{\text{fb}_{\text{exp}}}(t)$, but this is set to a constant as opposed to learned in these simulations). Inhibition of return (IOR) is applied at the previous position of the fovea, making it less likely to again be fixated. This inhibition is defined as:

$$u_{\text{ior}}(x, y, t) = \mathcal{G}_A(F^{-'}, \sigma_{22}) \tag{14}$$

The distribution of IOR defined in Equation 14 is a Gaussian, $\mathcal{G}_A$, centred at the location of the the previous fixation, $F^{-'}$. Attention shifting signals from the Saccade Motor Field, $u_{A \leftarrow M}(x, y, t)$, alter the spatial distribution of attention on the Spatial Attention Field, just as the SC can influence activity of LIP. This projection to the Spatial Attention Field is first convolved with its own kernel, $w_{A,M}$, in Equation 15, prior to being input to the field.

$$\begin{aligned} u_{A \leftarrow M}(x, y, t) = \int \int \big(& w_{A,M}(x - x', y - y', t) \\ & u^*_{\text{M}}(x', y', t) dx' dy' \big) \end{aligned} \tag{15}$$

A similar convolution, using the the kernel $w_{A,V}$, transforms the input to the Spatial Attention Field from the Visual Field in Equation 16.

$$\begin{aligned} u_{V \leftarrow A}(x, y, t) = \int \int \big(& w_{A,V}(x - x', y - y', t) \\ & u^*_{V}(x', y', t) dx' dy' \big) \end{aligned} \tag{16}$$

Saccade Motor Field

$$\begin{aligned} \tau \dot{u}_{\text{M}}(x, y, t) = & - u_{\text{M}}(x, y, t) + h_{M,47} \\ & - c_{48} u^*_{\text{x}}(t) \mathcal{G}_M(F, \sigma_{49}) \\ & + c_{16} u^*_g(t) u^*_{\text{M}}(x, y, t) \\ & + u_{M \leftarrow A}(x, y, t) + \zeta_M(x, y, t) \\ & + \int \int \big(w_{M,M}(x - x', y - y', t) \\ & \qquad u^*_{\text{M}}(x', y') dx' dy' \big) \\ & - c_{62} \sum u^*_{\text{M}}(x, y, t) \end{aligned} \tag{17}$$

Input from the Spatial Attention Field, $u_A(x, y, t)$, is the primary source of excitation to the Saccade Motor Field, the dynamics of which are formalized in Equation 17.

$$\begin{aligned} u_{A \leftarrow M}(x, y, t) = & c_{56}(1 - \mathcal{G}_M(F, \sigma_{55})) \\ & \int\int \big(w_{A \leftarrow M}(x - x', y - y', t) \\ & u^*_{\mathrm{A}}(x', y', t) dx' dy'\big) \end{aligned} \tag{18}$$

The projection from the Spatial Attention Field (Equation 18) is inhibited at the extant position of fixation F, as another way to mitigate continual refixation at the same location.

Inhibitory input from the Fixation Neuron, $u_{\mathrm{x}}(x, y, t)$, and the Gaze Change Neuron, $u_g(x, y, t)$, also project to the translated location of the fovea in order to allow the eye to move and to minimize refixations. Prior to convolving the Spatial Attention Field with its Saccade Motor Field specific kernel, additional Gaussian inhibition of the current foveal region is applied.

The location of a target on the Saccade Motor Field corresponds to the location with the largest activation at the moment a threshold level of activity in the Saccade Initiation Neuron is surpassed: $F' = \arg\max_{x,y}\{u_{\mathrm{M}}(x, y, t) | u^*_r \geq \theta_r\}$. As the Saccade Motor Field is reoriented during the saccade, it will start to relax to its resting level h_M.

In this version of LAG-1, the various forces that can affect the horizontal and vertical calculations of saccade amplitude are simplified to the Euclidean distance between the current point of fixation and the saccade target: $D = \|(F, F^{+'})\|$. An estimation of the time τ_D, needed to complete the saccade based on a predefined constant for velocity, υ_D is then calculated by $\tau_D = D/\upsilon_D$. In the brain, a multitude of factors, such as distance and direction of the movement, are known to influence the distinct pulse generators for horizontal and vertical movement of the eye. Once a saccade is initiated in Lag-1, the foveal position is updated at each time step along the path of the saccade.

Gaze Change Neuron

$$\begin{aligned} \tau \dot{u}_g(t) = & -u_g(t) + h_{g,63} - c_{64} u^*_r(t) \\ & + c_{65} u^*_{\mathrm{x}}(t) + c_{66} u^*_g(t) + c_{67} t_F + \zeta_g \end{aligned} \tag{19}$$

The Gaze Change Neuron, $u_g(t)$, defined in Equation 19, has three exogenous sources of input: it is inhibited by the Saccade Initiation Neuron $u_r(t)$, excited by the Fixation Neuron, $u_{\mathrm{x}}(t)$, and also excited by the Fixation Impatience, $t_F = (t - F_t^{-'})^{\lambda_{P2}}$, where $F_t^{-'}$ is the time since the last saccade was initiated. Larger values of λ_{P2} translate into steeper growth in input to the Gaze Change neuron, leading to shorter fixations.

Fixation Neuron

$$\begin{aligned} \tau \dot{u}_x(t) = & -u_x(t) + h_{x,68} - c_{69} u^*_g(t) - c_{70} u^*_r(t) \\ & + c_{71} u^*_{\mathrm{x}}(t) + \zeta_x \\ & + c_{72} u^*_{\mathrm{ft}_{\mathrm{det}}}(F_j, t) \end{aligned} \tag{20}$$

Changes in the activity of the Fixation Neuron, $u_{\mathrm{x}}(t)$, are described in Equation 20. Inhibitory input from the Saccade Initiation Neuron, $u_r(t)$, strongly suppresses the Fixation Neuron during a saccade. The Gaze Change Neuron, $u_g(t)$, also inhibits the Fixation Neuron in order to allow alternative locations to better attract attention prior to the initiation of a saccade. Finally, the stimulus input to the active Feature Detection Neuron, $u^*_{\mathrm{ft}_{det}}(j = F, t)$ also feeds into the Fixation Neuron allowing ongoing information processing to more directly affect the activity of the Fixation Neuron.

Saccade Initiation Neuron

$$\begin{aligned}\tau\dot{u}_r(t) = & -u_r(t) + h_r, 75 - c_{76}u_x^*(t) \\ & + c_{77}u_g^*(t) + c_{79}u_r^*(t) \\ & + c_{78}\max_{x,y} u_{\mathrm{M}}(x,y,t) + \zeta_r\end{aligned} \tag{21}$$

In Equation 21, the Saccade Initiation (**r**eset) Neuron, $u_r(t)$, is inhibited by the Fixation Neuron, $u_{\mathrm{x}}(t)$, excited by the Saccade Motor Field, $\max_{x,y} u_{\mathrm{M}}(x,y,t)$, and the Gaze Change Neuron, $u_g(t)$.

Feature Detection Neurons

$$\begin{aligned}\tau\dot{u_{\mathrm{ft_{det}}}}(j,t) = & -u_{\mathrm{ft_{det}}}(j,t) + h_{\mathrm{ft_{det}},80} \\ & + \zeta_{\mathrm{ft_{det}}}(t) \\ & + c_{82}u^*_{\mathrm{ft_{det}}}(j,t) \\ & + c_{83}u^*_{\mathrm{ft_{det}}}(j,t) - c_{81}u^*_{\mathrm{ft_{det}}}(j',t) \\ & + c_{84}\sum_j u^*_{\mathrm{ft_{det}}}(j,t)\end{aligned} \tag{22}$$

The activity of the Feature Detection Neurons, $u_{\mathrm{ft_{det}}}(j,t)$, is defined in Equation 22. The neuron representing the complementary feature value for the same spatial location is indexed by j'. Transduction of the stimulus attributes is defined in Equation 23 as a fovea sized integration over the jth layer of the the input to the Visual Field, where F_{Mask} is a Gaussian with parameters specified in the appendix and the active Feature Detection index j corresponds with the same index z_j along the feature dimension of the Visual Field.

$$u_{\mathrm{ft_{det}}}(j,t) = \int\int u^*_V(x,y,z_j)F_{Mask}(x,y,t)dx'dy' \tag{23}$$

Self-excitatory input, $u_{\mathrm{ft_{det}}}(j,t)$, is what provides the capacity for the neurons to act as a working memory that is constrained by the total capacity constrained activity of the other Feature Detection Neurons. As features are fixated, the neurons representing the previously viewed features decay down to their self-sustaining level, or are even forced out of their excited state. The activity of all the other Feature Detection Neurons is summed and subtracted from each detector as global inhibition.

Category Neurons

$$\begin{aligned}\tau\dot{u}_c(i,t) = & -u_c(i,t) + h_{c,97} + c_{99}u_c^*(i,t) \\ & + \sum_j c_{98}u_g(i,j,t)u^*_{\mathrm{ft_{det}}}(j,t) \\ & - c_{100}\sum_i u_c^*(i,t) \\ & + u^*_{\mathrm{fb_{det}}}(t)\mathbb{1}_{\mathrm{phase=feedback}}\mathbb{1}_{\mathrm{boost}\to u_c(i,t)} + \zeta_c(t)\end{aligned} \tag{24}$$

Activity from the Feature Detection Neurons, $u_{\mathrm{ft_{det}}}(j,t)$, passes through gain modulated connection weights, $u_g(i,j,t)$, defined in Equation 25, to selectively activate categories. Gain is calculated by taking the absolute difference in the connecting weights between the feature values connected to a particular category at a particular location in space. During the feedback phase, the correct category is boosted by multiplying together the activity of the Feedback Button Detector, $u^*_{\mathrm{fb_{det}}}$, and a constant $\mathbb{1}_{\mathrm{boost}\to u_c(i,t)}$.

$$u_g(i,j,t) = W(i,j,t) + c_{105}\left|W(i,j,t) - W(i,j',t)\right| \tag{25}$$

Feature to category association

On each time step of trial feedback, the weights connecting features and categories will undergo at least 2 of 3 possible types of associative learning (where the third type only occurs on error trials) as defined in Equation 26.

$$\tau \dot{W}(i,j,t) = dW_1(i,j,t) + dW_2(i,j,t) + dW_3(i,j,t) \tag{26}$$

On correct trials, only the first two terms of Equation 27 result in weight changes. The first weight change term, dW_1, moderately strengthens only those weights linking the correct category to active features in working memory according to the learning rate λ_{P1}, where the supervisory feedback signal for the correct category is given as input to the category neurons, according to Equation 24 at the start of the feedback phase.

$$\begin{aligned} dW_1(i,j,t) \quad &= \lambda_{P1} W(i > \theta_{131}, j > \theta_{w,c,130}, t) \\ &\quad (1 - c(i > \theta_{w,c,130}, t) \\ &\quad + c(i > \theta_{130}, t) u_{\mathrm{ft}_{\mathrm{det}}}(j > \theta_{131}, t)) \end{aligned} \tag{27}$$

Second, the entire weight matrix is incremented by a proportion of the learning rate in dW_2, reflecting a token amount of association between the task elements in general in Equation 27, and any of the responses. Although we set c_{126} and c_{127} to nominal values in these simulations, association like this would be needed to explain the intra category reversal learning advantage when compared to exra dimensional reversal learning [16].

$$dW_2(i,j,t) \quad = c_{126} + c_{127} \lambda_{P1} W(i,j,t) \tag{28}$$

The third term, dW_3 only applies to error trials, $\mathbb{1}_{\mathrm{error}}$, where a sharp anti-Hebbian decoupling between the response, $\mathbb{1}_{\mathrm{response}_i}$ and the active features is applied.

$$dW_3(i,j,t) = \begin{cases} -(c_{128} + c_{129}\lambda_{P1}) \\ \quad W(\mathbb{1}_{\mathrm{response}_i}, j > \theta_{131}, t), & \text{if } \mathbb{1}_{\mathrm{error}} \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

By subtracting off a proportion of the active weights from the weight change in Equation 27, Hebbian association slows down slightly as the weights get larger, similar again to earlier category learning models that annealed learning rates (e.g., [108, 151]).

Feature Expectation Neurons

$$\begin{aligned} \tau \dot{u_{\mathrm{ft}_{\mathrm{exp}}}}(j,t) = \; & - u_{\mathrm{ft}_{\mathrm{exp}}}(j,t) + h_{\mathrm{ft}_{\mathrm{exp}},106} \\ & - u^*_{\mathrm{ft}_{\mathrm{exp}}}(j',t) - c_{108} u^*_{\mathrm{ft}_{\mathrm{det}}}(j,t) \\ & + c_{109} \sum_i u_c(i,t) c_{107} u_g(i,j,t) \\ & + c_{110} u^*_{\mathrm{ft}_{\mathrm{exp}}}(j,t) - \sum_{m \neq j} u^*_{\mathrm{ft}_{\mathrm{exp}}}(m,t) \\ & + \zeta_{\mathrm{ft}^*_{\mathrm{exp}}}(t) \end{aligned} \tag{30}$$

Activation of the Feature Expectation Neurons, $u_{\mathrm{ft}_{\mathrm{exp}}}(j,t)$, as defined in Equation 30, is primarily driven by input from the Category Neurons, $u_c(i,t)$, after scaling the weights using the gain function, $u_g(i,j,t)$. The Feature Detection Neurons $u_{\mathrm{ft}_{\mathrm{det}}}(j,t)$ inhibit the Feature Expectation Neurons representing the same values in order to distinguish what is known so far in the trial from what has yet to be looked at. As with the Feature Detection Neurons, complementary feature values, denoted by $u_{\mathrm{ft}_{\mathrm{exp}}}(j',t)$, inhibit one another. Finally, global inhibition suppresses each Feature Expectation Neuron according the total activation energy of all the others.

Click Decision Neuron

$$\begin{aligned}\tau \dot{u}_d(t) = & -u_d(t) + h_{d,107} - (c_{114} u_c^*(t))^2 \\ & + c_{115} {u_{\text{impatience}_{\text{trial}}}}^*(t) - c_{116} u_{\text{x}}^*(t) + \zeta_d(t)\end{aligned} \tag{31}$$

The Click Decision Neuron, $u_d(t)$, is increasingly excited by impatience, $u_{\text{impatience}_{\text{trial}}}(t)$, over the course of the trial. Larger maximum Category Neuron activation is a much stronger influence on the decision to respond than smaller levels. This is modelled in Equation 31, by squaring the maximum value of the Category Neurons. An inhibitory input from the Fixation Neuron $u_{\text{x}}(t)$, was also needed in order to have LAG-1 complete fixations prior to responding. Trial Impatience is modelled by its own differential equation in Equation 32.

$$\begin{aligned}\tau \dot{u}_{\text{impatience}_{\text{trial}}}(t) = & -u_{\text{impatience}_{\text{trial}}}(t) + h_{\text{impatience}_{\text{trial}},120} \\ & + c_{123}(t - (\mathbb{1}_{\text{phase=feedback}_t}))^{\lambda_{P3}} \\ & + \zeta_{\text{impatience}_{\text{trial}}}(t)\end{aligned} \tag{32}$$

The indicator function $\mathbb{1}_{\text{phase=feedback}}$ is set to 1 if the feedback phase has been initiated, effectively resetting the model's impatience relative to the start time of the feedback phase, $\mathbb{1}_{\text{phase=feedback}_t}$.

The selection of a particular category i, as a final response when the Click Decision Neuron crosses its decision threshold, is the result of a stochastic decision, given by Equation 33:

$$P(i) = \frac{e^{u_{c_i}/T}}{\sum_i e^{u_{c_i}/T}} \tag{33}$$

The use of the Softmax rule specifies one way to make choices under conditions of uncertainty. Neurologically this calculation relies on prefrontal circuits that can differentially weight sensorimotor competitions for different body parts mapped parietally [99]. A choice of strategy, such as the decision to exploit current knowledge or explore the space, is itself a kind of category decision. The Softmax temperature parameter, T, scales the effect that differences in activation will have on the probability of choice. When the temperature is high, the decision choices move toward being equally likely. When the temperature is lower, the choice with the highest activation becomes increasingly favoured.

S4 Appendix. Supplementary equation parameters.

Table 5. Supplementary equation parameters

Parameter	Arguments
Noise Kernel Extras	$w_\zeta = [c_5, 1]$
Visual Field extras	$u_V^* = f_{\beta_{10},\mu_{0,11}}(u_V(x,y,t))$ $w_V = [c_{6,\sigma 7,c8,\sigma_9}]$ $u_{\mathrm{ft}_{\mathrm{exp}}}^* = f_{\beta_{15},\mu_{0,16}}(u_{\mathrm{ft}_{\mathrm{exp}}}(z,t))$ $\zeta_V = [c_5, 1]$ $u_{\mathrm{fb}_{\mathrm{exp}}}^* = f_{\beta_{23},\mu_{0,24}}(u_{\mathrm{fb}_{\mathrm{exp}}}(t))$
Spatial Attention Field extras	$u_A^* = f_{\beta_{32},\mu_{0,33}}(u_A(x,y,t))$ $\zeta_A = [c_{39}, c_{40}]$ $w_{A,A} = [c_{34}, \sigma_{35}, c_{36}, \sigma_{37}]$ $w_{A,M} = [c_{41}, \sigma_{42}]$
Saccade Motor Field extras	$w_{M,M} = [c_{53}, \sigma_{54}]$ $u_M^* = f_{\beta_{43},\mu_{0,44}}(u_{\mathrm{M}}(x,y,t))$ $w_{M,M} = [c_{59}, \sigma_{60}, c_{61}, \sigma_{62}]$ $\zeta_{M \leftarrow A} = [c_{39}, c_{40}]$
Gaze Change Neuron extras	$\zeta_g = [c_{89}, c_{90}]$
Saccade Initiation Neuron extras	$\zeta_r = [c_{93}, c_{94}]$ $u_r^* = f_{\beta_{26},\mu_{0,27}}(u_r(t))$
Fixation Neuron extras	$u_{\mathrm{x}_{\mathrm{ft}_{\mathrm{det}}}}^* = f_{\beta_{73},\mu_{0,74}}(\mathrm{x}_{u_{\mathrm{ft}_{\mathrm{det}}}}(t))$ $\zeta_x = [c_{91}, c_{92}]$ $u_{\mathrm{x}}^*(t) = f_{\beta_{51},\mu_{0,51}}(\mathrm{x}^*(t))$
Feature Detection Extras	$u_{\mathrm{ft}_{det}}^*(j,t) = f_{\beta_{85},\mu_{0,86}}(u_{\mathrm{ft}_{det}}(t))$ $F_{Mask} = \mathcal{G}[\mu = (0,0), \sigma = (3,3)]$ $u_{\mathrm{ft}_{det},f,V}^*(j,t) = f_{\beta_{87},\mu_{0,88}}(u_{\mathrm{ft}_{det},f,V}(j,t))$ $\zeta_{u_{\mathrm{ft}_{\mathrm{det}}}} = [c_{95}, c_{96}]$
Category Neuron extras	$u_c^*(i,t) = f_{\beta_{101},\mu_{0,102}}(u_c(i,t))$ $\zeta_c = [c_{103}, c_{104}]$
Feature Expectation Neuron extras	$\zeta_{\mathrm{ft}_{\mathrm{exp}}} = [c_{111}, c_{112}]$
Decision Neuron extras	$u_d^*(t) = f_{\beta_{116},\mu_0,117}(u_c(i,t))$ $\zeta_d = c_{118}, c_{119}$
Trial Impatience extras	$u_{\mathrm{impatience}_{\mathrm{trial}}}(t)^* = f_{\beta_{120},\mu_0,121}(u_{\mathrm{impatience}_{\mathrm{trial}}}(t))$ $\zeta_{\mathrm{impatience}_{\mathrm{trial}}} = [c_{124}, c_{125}]$

S5 Appendix. Parameter tables and best fits.

Table 6. Parameter changes between Simulation 1 and Simulation 2.

Parameter	Simulation 1	Simulation 2
$c_{c,\mathrm{ft}_{\mathrm{det}},98}$	1	4.6
$c_{M,+A,53}$	7	12
$c_{d,k,113}$	4	2
$c_{c,g,105}$	1	0.2
$\mathbb{1}_{\mathrm{boost}}$	0.3	0.7

Table 7. Simulation 1: Covariance parameters.

Measure	Covariance
Average accuracy	0.00001
Accuracy slope	0.0001
Average fixation count	1
Fixation count slope	1
Average probability of irrelevant fixation	1
Probability of irrelevant fixation slope	1
Average fixation duration	1
Fixation duration slope	1

Table 8. Simulation 2: Covariance parameters.

Measure	Covariance
Exemplar A1 at Transfer	0.0001
Exemplar A2 at Transfer	0.0001
Exemplar A3 at Transfer	1
Exemplar A4 at Transfer	1
Exemplar A5 at Transfer	1
Exemplar B1 at Transfer	0.0001
Exemplar B2 at Transfer	0.001
Exemplar B3 at Transfer	0.1
Exemplar B4 at Transfer	1
Exemplar T1 at Transfer	1
Exemplar T2 at Transfer	1
Exemplar T3 at Transfer	1
Exemplar T4 at Transfer	1
Exemplar T5 at Transfer	1
Exemplar T6 at Transfer	1
Exemplar T7 at Transfer	1
Criterion point	1
Feature 1 first half fixation count	0.001
Feature 1 and Feature 2 second half difference	0.001

Table 9. Specifies the experiment parameters for Simulation 1 and Simulation 2.

Parameter	Simulation 1	Simulation 2
Feature placement	F1:(51,68), F2:(66,43), F3:(36,43), Feedback:(33,71)	F1:(39,63), F2:(63,63), F3:(63,39), F4:(39,39), Feedback:(33,71)
Fovea size, (visual angle °)	$\sigma^2 = 3 = 1.7°$	$\sigma^2 = 3 = 1.6°$
Feature diameter	6.25	6.25
Visual Field (u_V) size (px)	101x101	101x101
Spatial Attention Field (u_A) size (px)	51x51	51x51
Saccade Motor Field (u_M) size (px)	51x51	51x51
Timescale, τ	10	10
Characteristic time, δt (5ms)	2	2
Temperature	0.07	0.07
Response threshold	0.8	0.8
Feedback threshold	0.98	0.98
Saccade threshold, θ_r	0.2	0.2

Table 10. All fixed parameter values.

Name	Value	Name	Value	Name	Value	Name	Value
$h_{V,1}$	-4	$c_{A,+A,34}$	5	$c_{g,\text{imp}_{\text{fixation}},67}$	0.0001	$c_{c,-c,100}$	3.5
$c_{V,I,2}$	4	$\sigma_{A,+A,35}$	5	$h_{\text{x},68}$	-3.5	$\beta_{c,101}$	1.37
$c_{V,r,3}$	3	$c_{A,-A,36}$	8	$c_{\text{x},g,69}$	-0.4	$\mu_{0,c,102}$	0.15
$c_{V,\text{fb}_{\text{nov}},4}$	10,1	$\sigma_{A,-A,37}$	7.5	$c_{\text{x},r,70}$	-3	$\zeta_{c,T,103}$	0.02
$\zeta_{v,T,5}$	0.005	$c_{-A_{gi},38}$	0.2	$c_{\text{x,x},71}$	3	$\zeta_{c,\delta,104}$	0.2
$c_{V,+V,6}$	1	$\zeta_{A,T,39}$	0.1	$c_{\text{x},\text{ft}_{\text{det}},72}$	5	$c_{c,g,105}$	1
$\sigma_{V,+V,7}$	1.75	$\zeta_{A,\delta,40}$	15	$\beta_{\text{x},\text{ft}_{\text{det}},73}$	1	$h_{\text{ft}_{\text{exp}},106}$	-3
$c_{V,-V,8}$	1	$c_{A,+M,41}$	8	$\mu_{0,u_{\text{x}},\text{ft}_{\text{det}},74}$	1	$h_{d,107}$	-5
$\sigma_{V,-V,9}$	3.75	$\sigma_{A,M,42}$	2.5	$h_{r,75}$	-5	$c_{\text{ft}_{\text{exp}},f,108}$	0.01
$\beta_{V,10}$	2	$\beta_{M,43}$	4	$c_{r,\text{x},76}$	-10	$c_{\text{ft}_{\text{exp}},c,109}$	2
$\mu_{0,V,11}$	2	$\mu_{0,M,44}$	0	$c_{r,g,77}$	4	$c_{\text{ft}_{\text{exp}},+\text{ft}_{\text{exp}},110}$	1.5
$c_{V,-A,12}$	0.001	$c_{A,+V,45}$	70	$c_{r,M,78}$	8	$\zeta_{\text{ft}_{\text{exp}},T,111}$	0.1
$c_{v,+\text{ft}_{\text{exp}},13}$	3	$\sigma_{A,+V,46}$	1.25	$c_{r,r,79}$	1.4	$\zeta_{\text{ft}_{\text{exp}},\delta,112}$	1
$c_{V,-\text{ft}_{\text{exp}},14}$	0.04	$h_{M,47}$	-2	$h_{\text{ft}_{\text{det}},80}$	-2	$c_{d,k,113}$	4
$\beta_{\text{ft}_{\text{exp}},15}$	0.5	$c_{M,\text{x},48}$	-50	$c_{\text{ft}_{\text{det}},-\text{ft}_{\text{det}},81}$	4	$c_{d,\text{imp}_{\text{trial}},114}$	1.6
$\sigma_{0,\text{ft}_{\text{exp}},16}$	0	$\sigma_{M,-M,49}$	5	$c_{\text{ft}_{\text{det}},v,82}$	500	$c_{d,x,115}$	-1.2
$\sigma_{\zeta,17}$	5	$\beta_{x,50}$	0.5	$c_{\text{ft}_{\text{det}},\text{ft}_{\text{det}},83}$	4.6	$\beta_{d,116}$	2.5
$h_{A,18}$	-0.5	$\mu_{0,\text{x},51}$	0	$c_{\text{ft}_{\text{det}},gi,84}$	0.1	$\mu_{0,d,117}$	1
$c_{A,+\text{ft}_{\text{exp}},19}$	0.2	$c_{M,g,52}$	30	$\beta_{\text{ft}_{\text{det}},85}$	1	$\zeta_{d,T,118}$	0.1
$c_{A,+\text{x},20}$	5	$c_{M,+A,53}$	7	$\mu_{0,\text{ft}_{\text{det}},86}$	0	$\zeta_{d,\delta,119}$	0.1
$c_{A,-i,21}$	0.3	$\sigma_{M,+A,54}$	3.75	$\beta_{\text{ft}_{\text{det}}I,87}$	1	$\beta_{\text{imp}_{\text{trial}},120}$	0.01
$\sigma_{A,I,22}$	2.5	$\sigma_{M,-F,55}$	4	$\mu_{\text{ft}_{\text{det}}I,88}$	0	$\mu_{0,\text{imp}_{\text{trial}},121}$	300
$\beta_{\text{fb}_{\text{exp}},23}$	2	$c_{M,-F,56}$	3	$\zeta_{g,T,89}$	0.001	$h_{\text{imp}_{\text{trial}},122}$	1
$\mu_{0,\text{fb}_{\text{exp}},24}$	2	$\zeta_{M,T,57}$	0.2	$\zeta_{g,\delta,90}$	0.1	$c_{\text{imp}_{\text{trial}},T,123}$	0.001
$c_{A,\text{fb}_{\text{Button}},25}$	5	$\zeta_{M,\delta,58}$	1	$\zeta_{u_{\text{x}},T,91}$	0.01	$\zeta_{\text{imp}_{\text{trial}},T,124}$	0.01
$\beta_{r,26}$	2.5	$c_{M,+M,59}$	5	$\zeta_{\text{x},\delta,92}$	0.1	$\zeta_{\text{imp}_{\text{trial}},\delta,125}$	1
$\mu_{0,r,27}$	2.5	$c_{M,-M,60}$	8	$\zeta_{r,T,93}$	0.005	$c_{w,126}$	0
$c_{A,+g,28}$	-5	$\sigma_{M,-M,61}$	7.5	$\zeta_{r,\delta,94}$	0.1	$c_{w,127}$	0.2
$c_{A,+r,29}$	-10	$c_{-M,62}$	5	$\zeta_{\text{ft}_{\text{det}},T,95}$	0.02	$c_{w,128}$	10
$\beta_{g,30}$	3	$h_{g,63}$	-1	$\zeta_{\text{ft}_{\text{det}},\delta,96}$	0.2	$c_{w,129}$	0.33
$\mu_{0,g,31}$	0	$c_{g,r,64}$	-1	$h_{c,97}$	-0.5	$\theta_{w,c,130}$	0
$\beta_{A,32}$	1	$c_{g,\text{x},65}$	0.2	$c_{c,\text{ft}_{\text{det}},98}$	4.6	$\theta_{w,\text{ft}_{\text{exp}},131}$	-1
$\mu_{0,A,33}$	0	$c_{g,g,66}$	0.5	$c_{c,+c,99}$	1.2		

Each parameter appears in the text with only the numerical subscript.

Table 11. Best fitting parameters: Simulation 1.

Subject	Learning Rate	Fixation Impatience	Trial Impatience	RSS
Subject-1	1.40e-05	1.75	1.65	-13411
Subject-2	1.65e-05	1.70	1.65	-11553
Subject-3	2.50e-06	1.65	1.85	-16749
Subject-4	1.65e-05	1.65	1.60	-12057
Subject-5	4.50e-06	1.50	1.75	-45454
Subject-6	1.25e-05	1.65	1.85	-11144
Subject-7	1.10e-05	1.90	1.65	-21914
Subject-8	1.40e-05	1.70	1.90	-11146
Subject-9	1.15e-05	1.55	1.85	-23273
Subject-10	1.65e-05	1.70	1.65	-11553
Subject-11	1.45e-05	1.80	1.90	-16196
Subject-12	6.00e-06	1.85	2.00	-22008
Subject-13	1.20e-05	1.50	1.80	-42564
Subject-14	7.50e-06	1.75	1.75	-12927
Subject-15	2.50e-06	1.65	1.85	-16749
Subject-16	7.50e-06	1.75	1.65	-12753
Subject-17	1.50e-05	1.80	1.65	-16005
Subject-18	1.30e-05	1.75	1.95	-13570
Subject-19	5.50e-06	1.70	1.65	-11025
Subject-20	6.00e-06	1.80	1.65	-15396
Subject-21	1.25e-05	1.50	1.75	-42805
Subject-22	1.40e-05	1.70	1.90	-11146
Subject-23	6.00e-06	1.70	1.85	-13640
Subject-24	1.55e-05	1.65	1.65	-11506
Subject-25	1.60e-05	1.70	1.65	-11700
Subject-26	1.25e-05	1.60	1.80	-14213
Subject-27	1.45e-05	1.85	1.85	-18727
Subject-28	1.25e-05	1.65	1.85	-11144
Subject-29	1.25e-05	1.60	1.85	-15218
Subject-30	1.45e-05	1.75	1.75	-12969
Subject-31	1.40e-05	1.75	1.65	-13411
Subject-32	1.30e-05	1.75	1.95	-13570
Subject-33	1.65e-05	1.65	1.60	-12057
Subject-34	1.25e-05	1.50	1.75	-42805
Subject-35	1.40e-05	1.75	1.75	-12938
Subject-36	8.00e-06	1.80	2.05	-20832
Subject-37	7.50e-06	1.65	1.75	-12477
Subject-38	1.30e-05	1.70	1.75	-11241
Subject-39	1.25e-05	1.60	1.85	-15218
Subject-40	6.50e-06	1.70	1.65	-11058
Subject-41	1.25e-05	1.70	1.65	-11119
Subject-42	1.40e-05	1.75	1.65	-13411

Table 12. Best fitting parameters: Simulation 2.

Subject	Learning Rate	Fixation Impatience	Trial Impatience	RSS
Subject-1	1.75E-05	1.45	1.65	-3491
Subject-2	4.50E-06	2.30	2.20	-3845
Subject-3	2.00E-06	1.85	1.90	-4224
Subject-4	6.00E-06	1.65	1.85	-2536
Subject-5	1.05E-05	1.65	1.85	-2352
Subject-6	9.50E-06	1.55	1.75	-4799
Subject-7	9.50E-06	1.60	1.80	-2614
Subject-8	1.75E-05	1.45	1.65	-3491
Subject-9	5.50E-06	1.60	1.80	-2921
Subject-10	1.75E-05	1.45	1.65	-3491
Subject-11	9.50E-06	1.60	1.80	-2614
Subject-12	5.50E-06	1.60	1.80	-2921
Subject-13	8.00E-06	2.25	2.15	-3456
Subject-14	5.50E-06	1.60	1.80	-2921
Subject-15	5.00E-06	1.65	1.90	-5188
Subject-16	1.05E-05	1.65	1.85	-2352
Subject-17	5.50E-06	1.60	1.80	-2921
Subject-18	9.50E-06	1.55	1.75	-4799
Subject-19	1.10E-05	1.70	1.85	-2201
Subject-20	3.00E-06	2.30	2.20	-5019
Subject-21	1.75E-05	1.45	1.65	-3491
Subject-22	5.50E-06	1.60	1.80	-2921
Subject-23	1.75E-05	1.45	1.65	-3491
Subject-24	6.50E-06	1.95	2	-3996
Subject-25	1.75E-05	1.45	1.65	-3491
Subject-26	6.50E-06	1.55	1.75	-2647
Subject-27	2.50E-06	2.15	2.20	-4219
Subject-28	6.50E-06	1.55	1.75	-2647
Subject-29	1.05E-05	1.65	1.85	-2352
Subject-30	1.75E-05	1.45	1.65	-3491
Subject-31	1.10E-05	1.70	1.85	-2201
Subject-32	1.75E-05	1.45	1.65	-3491
Subject-33	6.50E-06	1.55	1.75	-2647
Subject-34	1.75E-05	1.45	1.65	-3491
Subject-35	1.75E-05	1.45	1.65	-3491
Subject-36	6.50E-06	1.95	2	-3996
Subject-37	1.75E-05	1.45	1.65	-3491
Subject-38	1.75E-05	1.45	1.65	-3491
Subject-39	1.75E-05	1.45	1.65	-3491
Subject-40	4.50E-06	1.55	1.75	-2750
Subject-41	1.75E-05	1.45	1.65	-3491
Subject-42	6.50E-06	1.55	1.75	-2647
Subject-43	1.05E-05	1.65	1.85	-2352
Subject-44	1.05E-05	1.65	1.85	-2352
Subject-45	8.50E-06	1.60	1.80	-3887
Subject-46	1.75E-05	1.45	1.65	-3491
Subject-47	8.00E-06	1.80	1.95	-2855

Continued on next page

Table 12 – continued from previous page

Subject	Learning Rate	Fixation Impatience	Trial Impatience	RSS
Subject-48	8.00E-06	1.80	1.95	-2855
Subject-49	5.50E-06	1.50	1.75	-3752
Subject-50	1.75E-05	1.45	1.65	-3491
Subject-51	3.50E-06	1.55	1.75	-2229
Subject-52	4.50E-06	2.25	2.15	-4997
Subject-53	2.00E-06	1.70	1.90	-5253
Subject-54	3.00E-06	2	2.10	-4547
Subject-55	6.00E-06	2.15	2.10	-4743
Subject-56	6.50E-06	1.55	1.75	-2647
Subject-57	1.10E-05	1.70	1.85	-2201
Subject-58	4.50E-06	1.90	2	-3674
Subject-59	6.50E-06	1.55	1.75	-2647
Subject-60	9.50E-06	1.55	1.75	-4799
Subject-61	6.50E-06	1.55	1.75	-2647
Subject-62	9.50E-06	1.55	1.75	-4799
Subject-63	3.50E-06	2.45	2.25	-6432
Subject-64	1.05E-05	1.65	1.85	-2352

S6 Appendix. Individual fit visualizations.

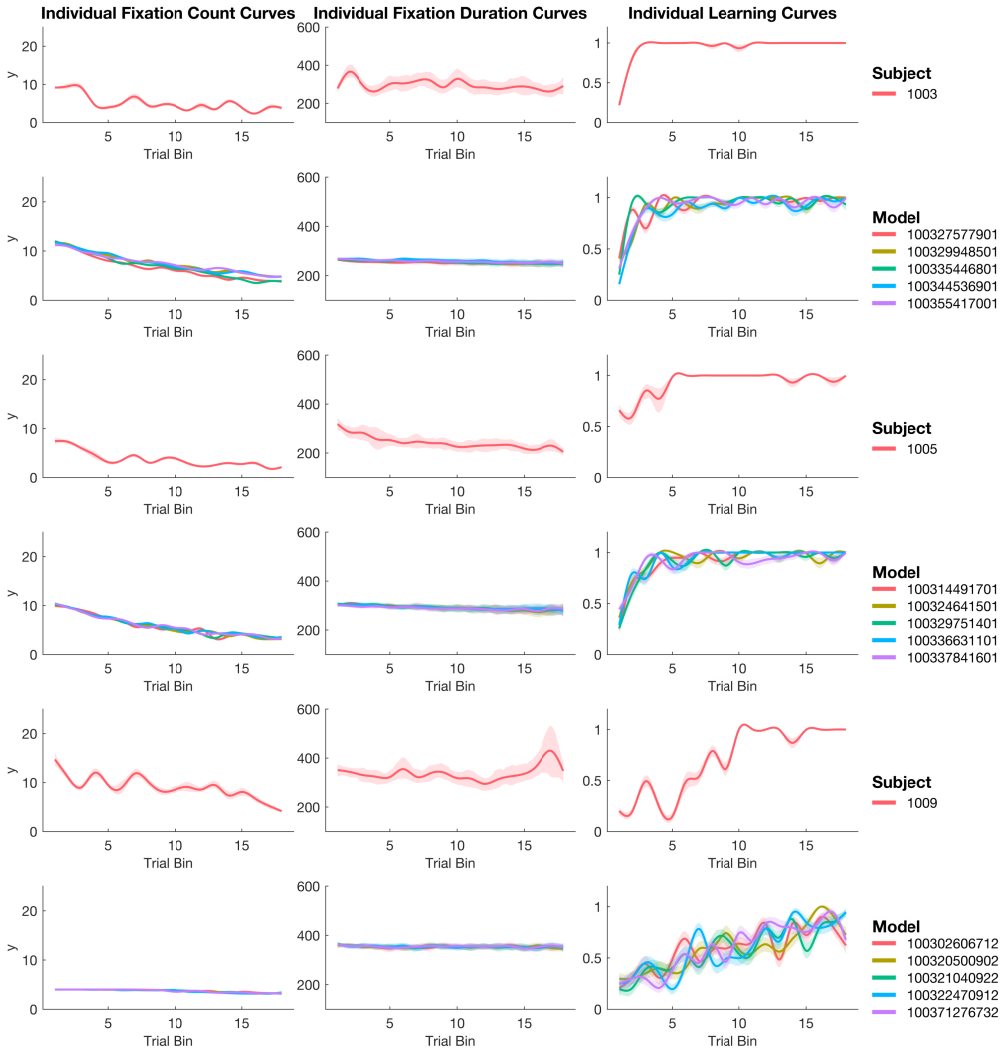


Figure 21. Subjects: 1003, 1005, 1008. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

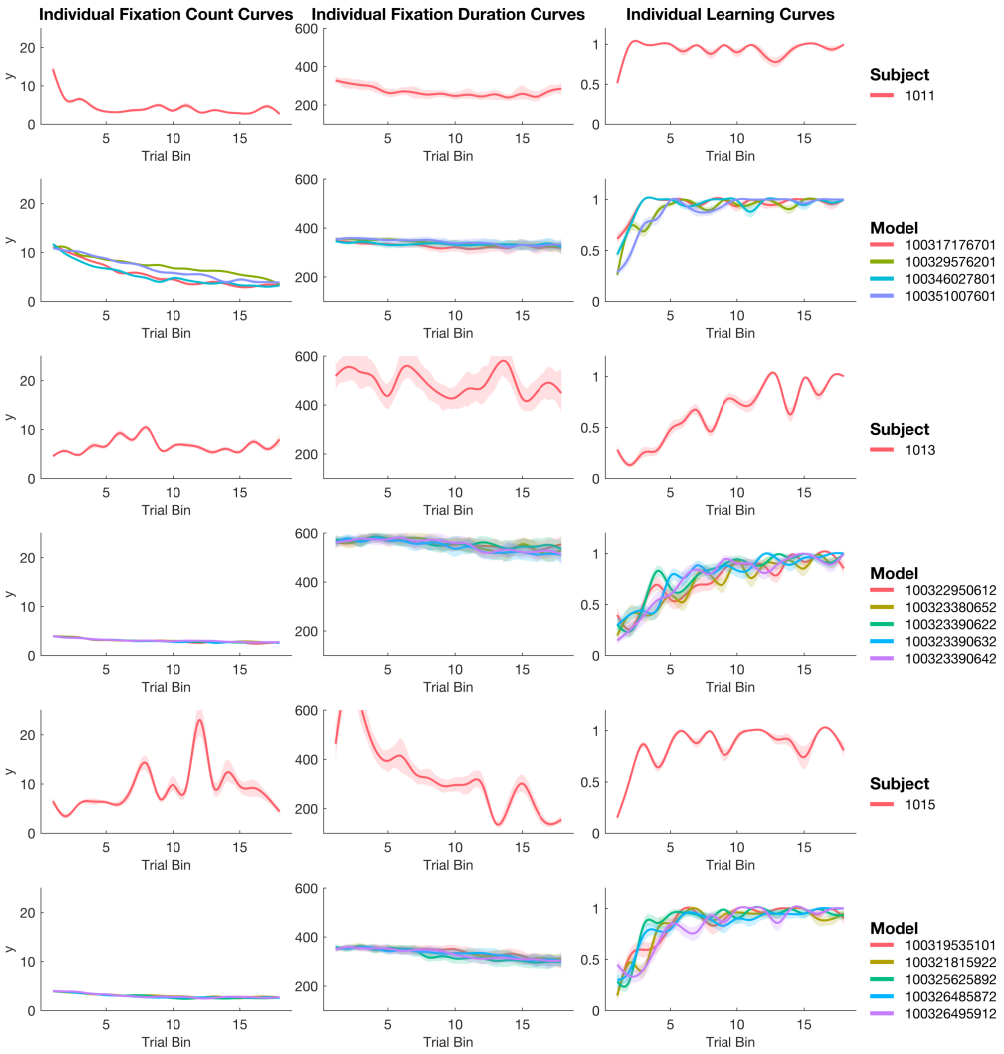


Figure 22. Subjects: 1011, 1013, 1015. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

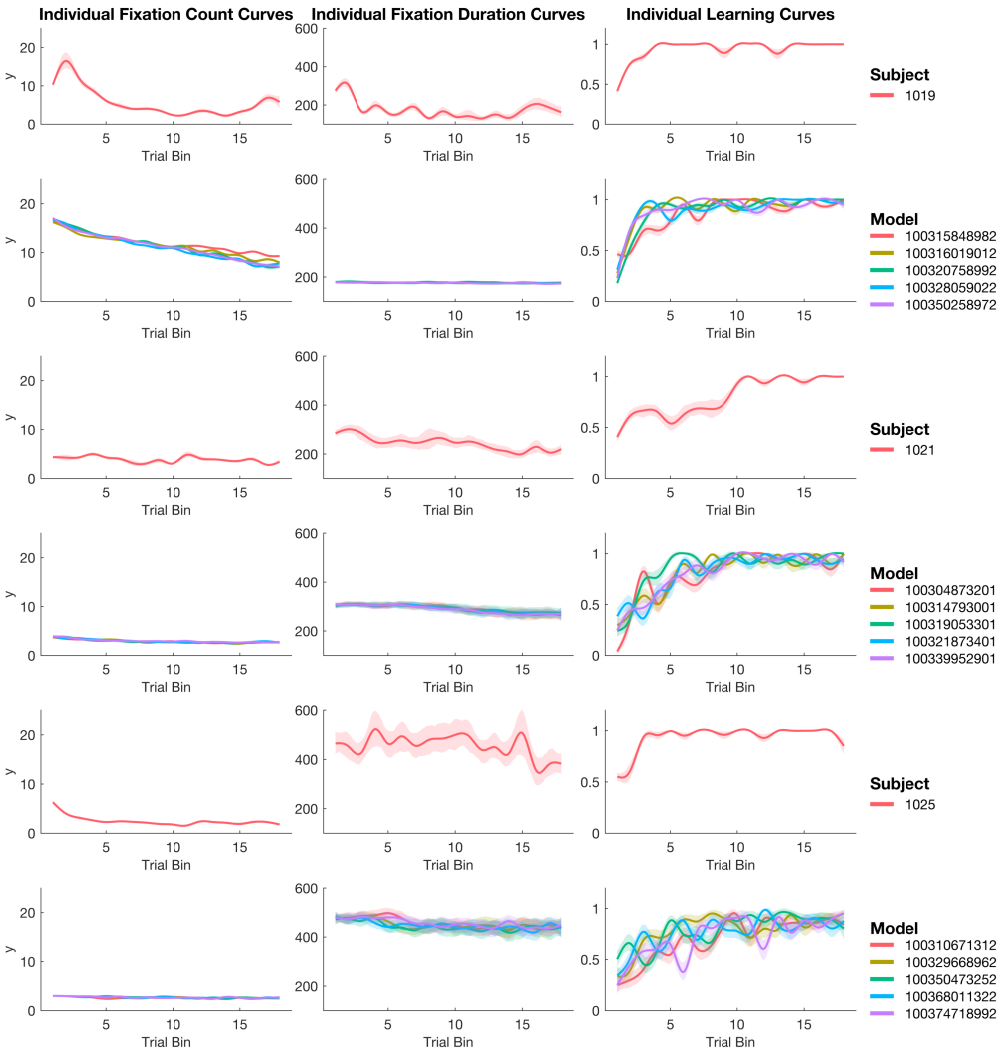


Figure 23. Subjects: 1019, 1021, 1025. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

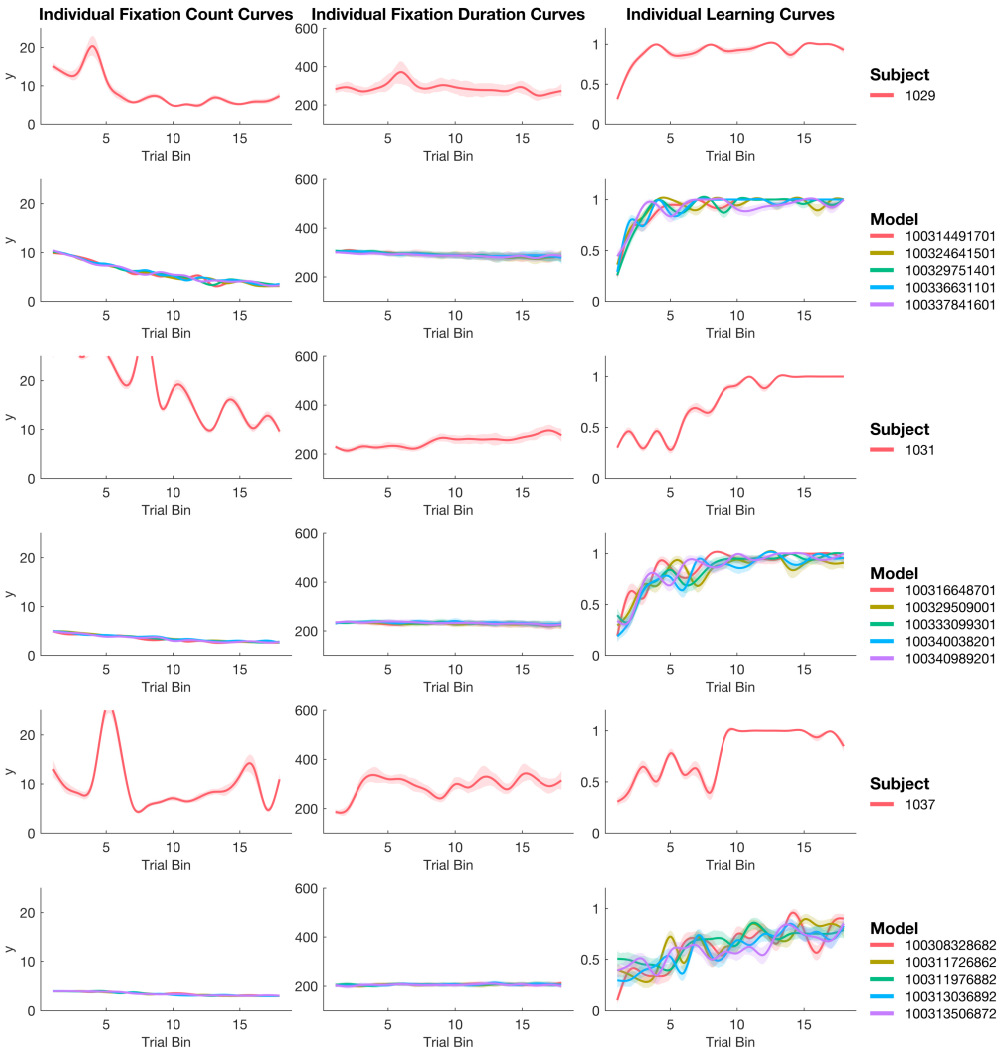


Figure 24. Subjects: 1029, 1031, 1037. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

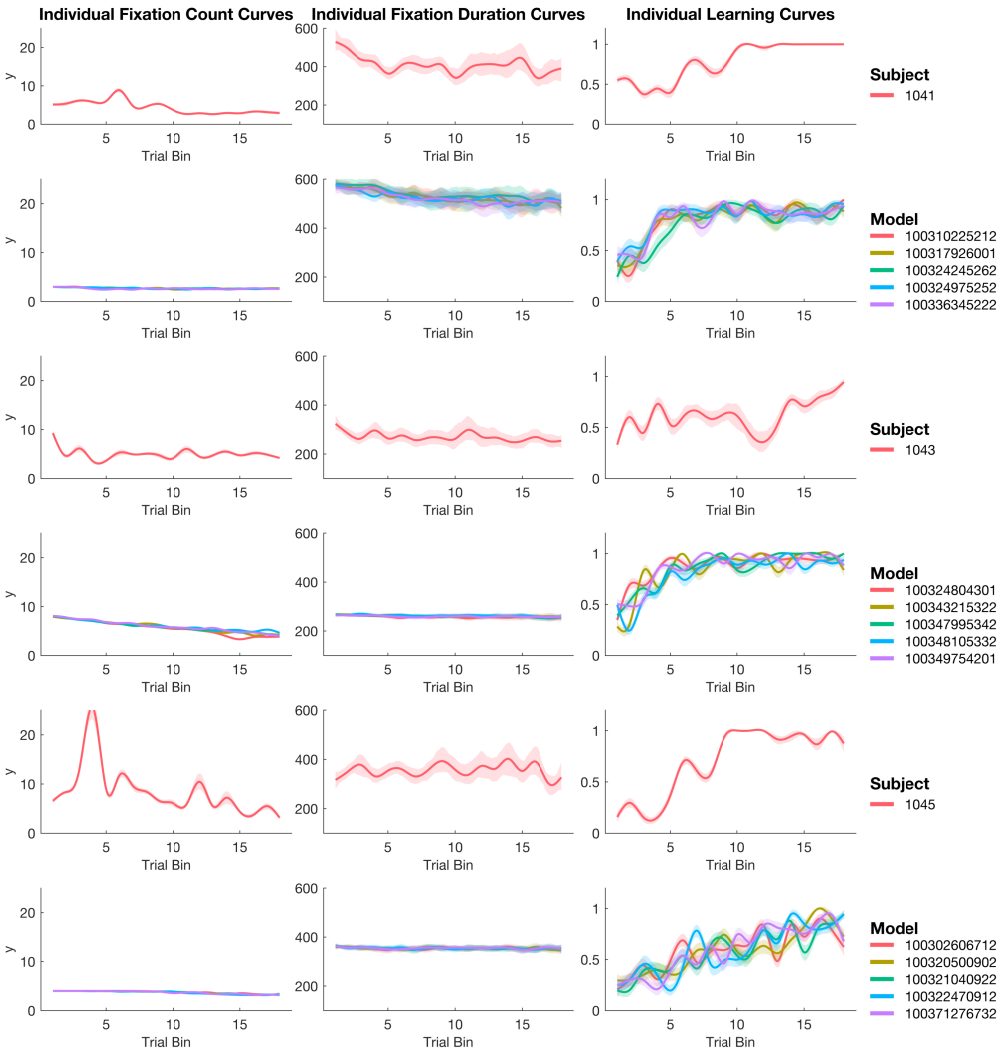


Figure 25. Subjects: 1041, 1043, 1045. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

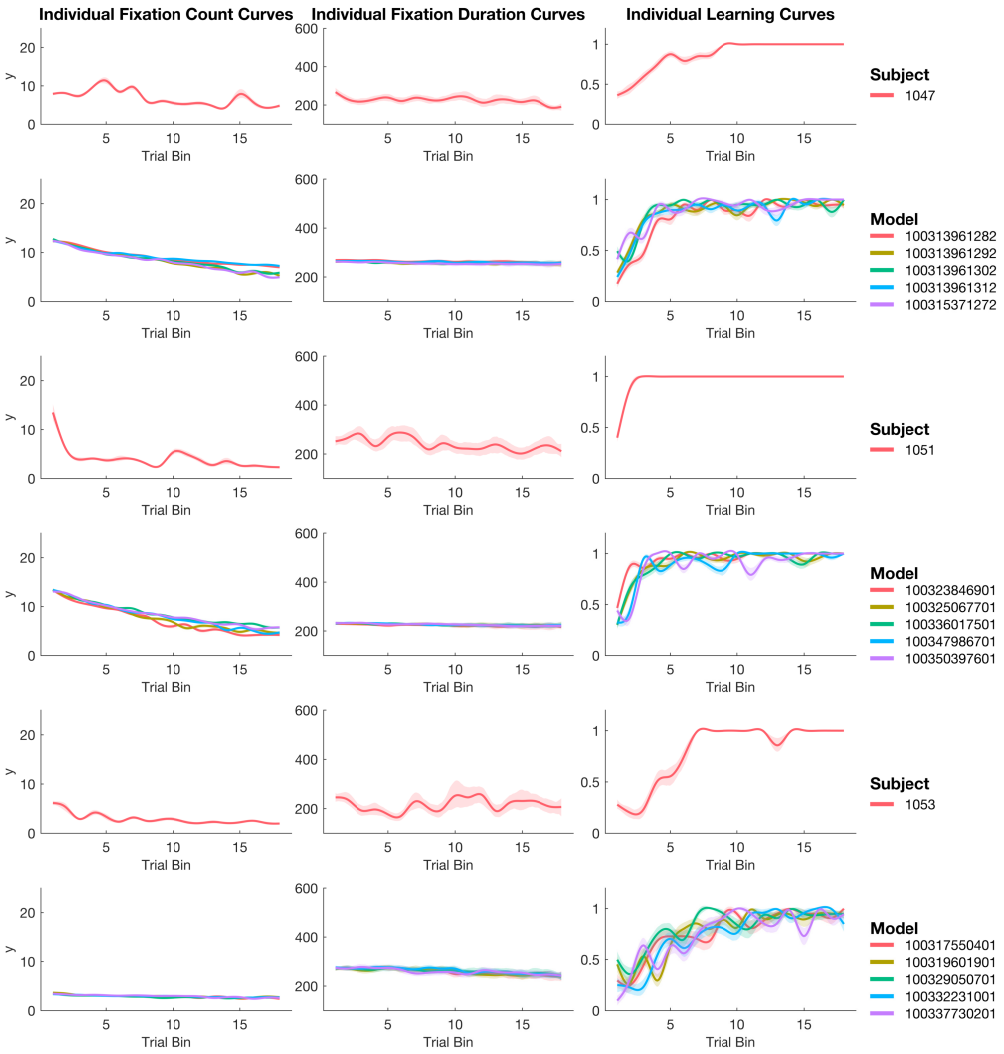


Figure 26. Subjects: 1047, 1051, 1053. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

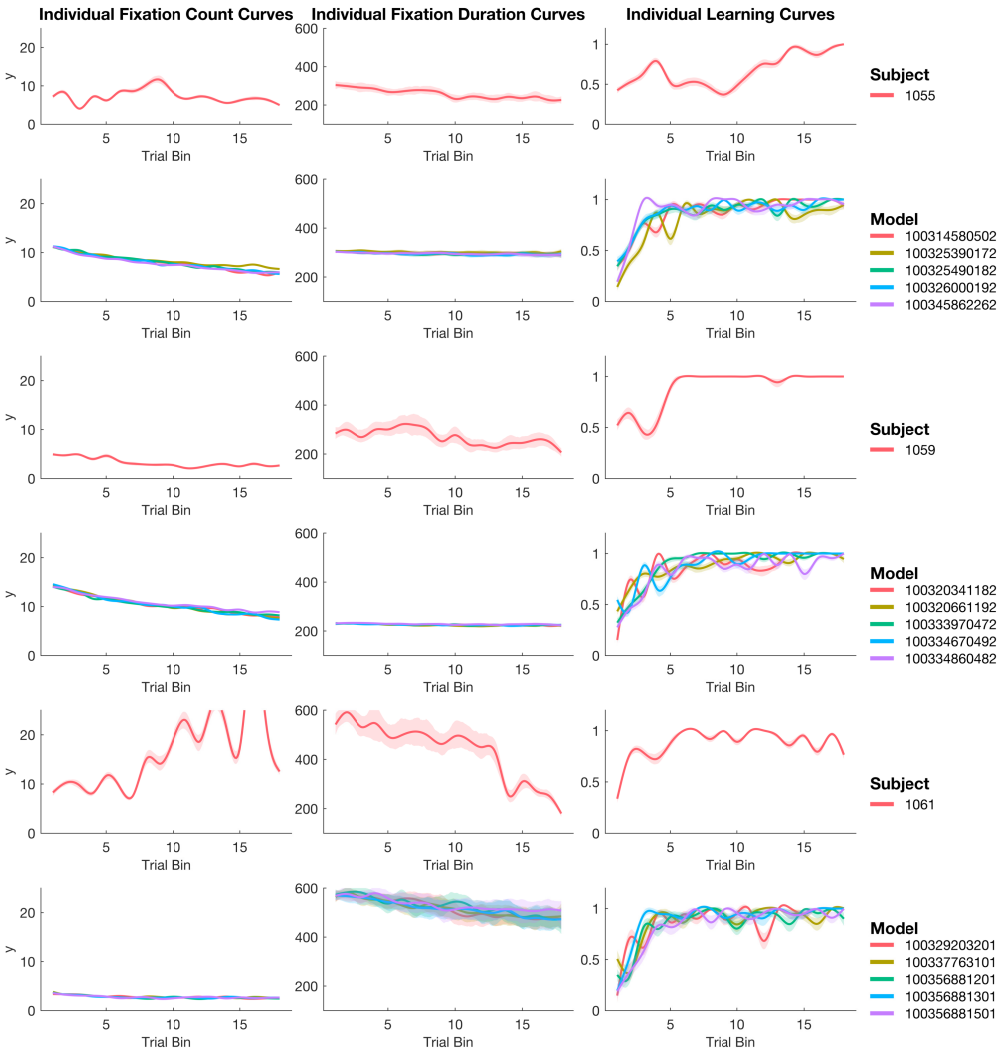


Figure 27. Subjects: 1055, 1059, 1061. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

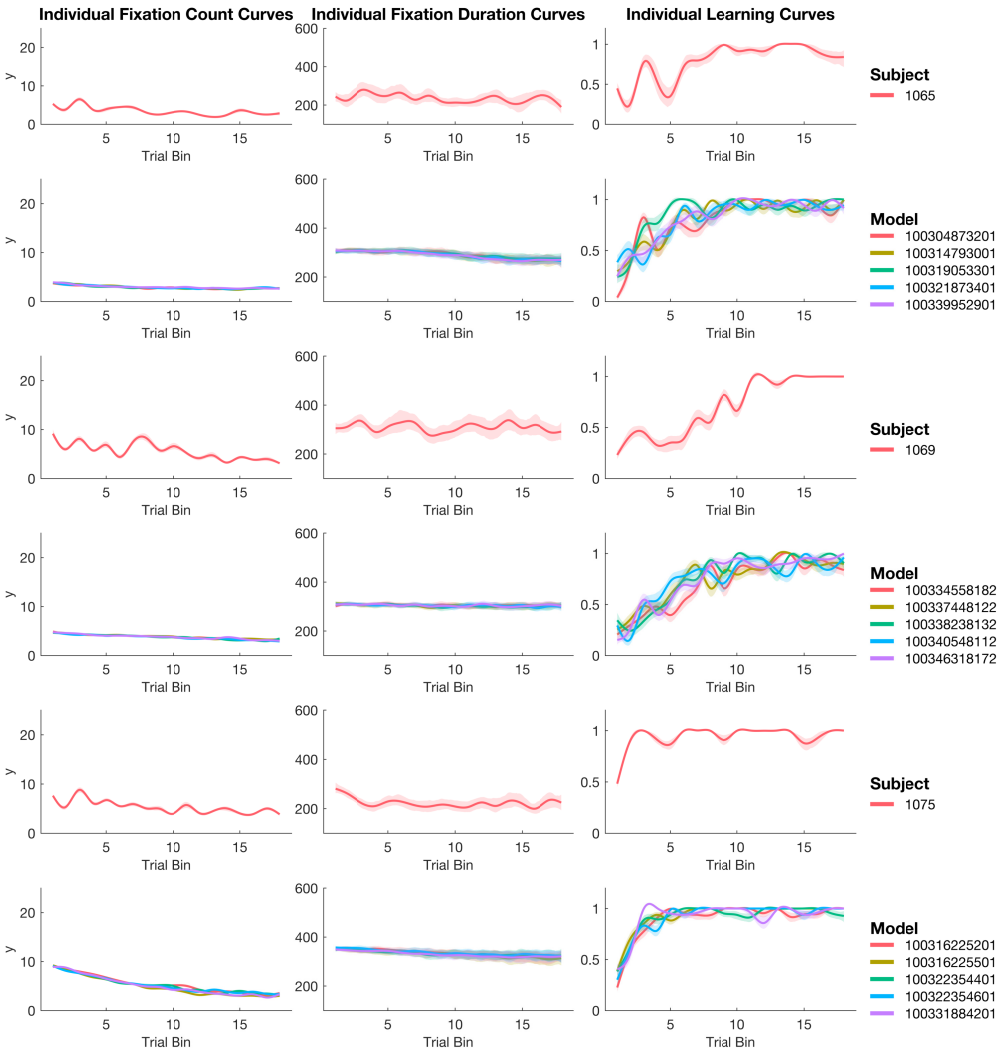


Figure 28. Subjects: 1065, 1069, 1075. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

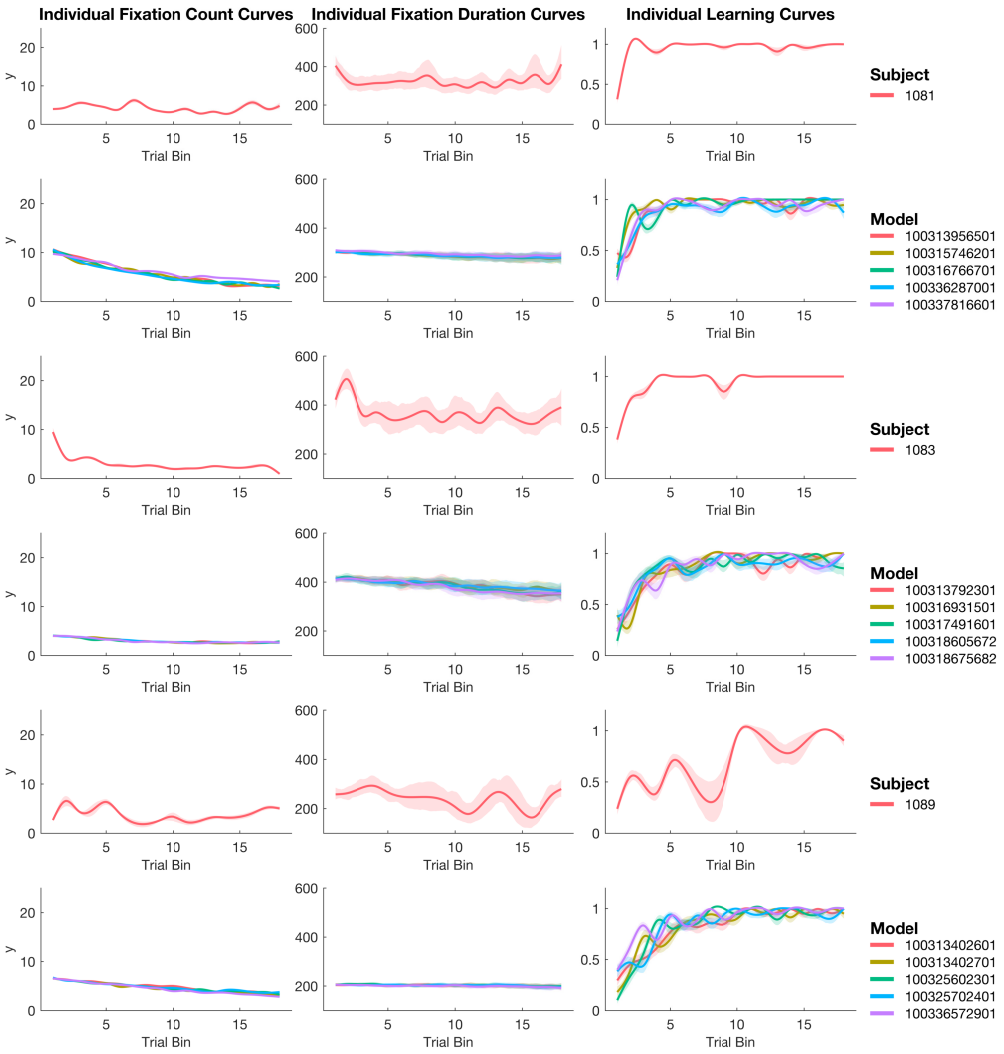


Figure 29. Subjects: 1081, 1083, 1089. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

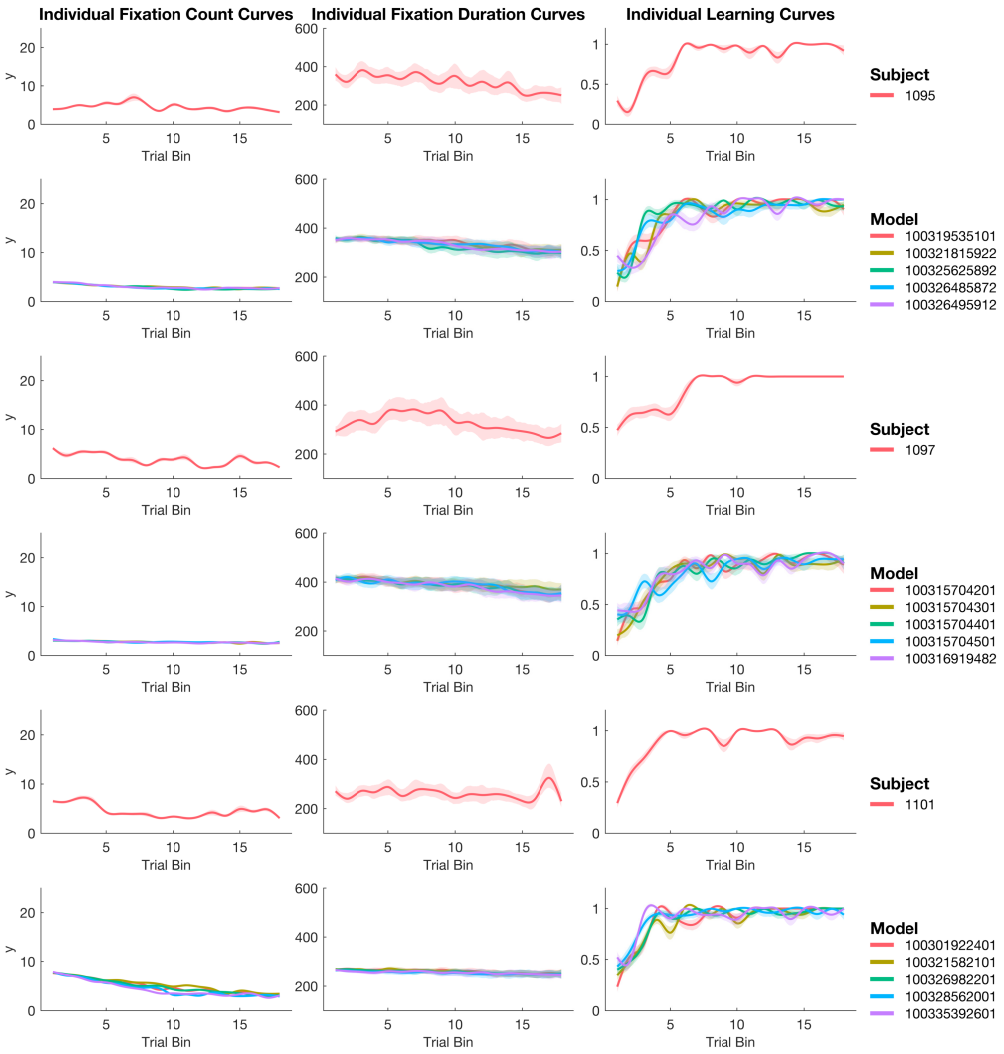


Figure 30. Subjects: 1095, 1097, 1101. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

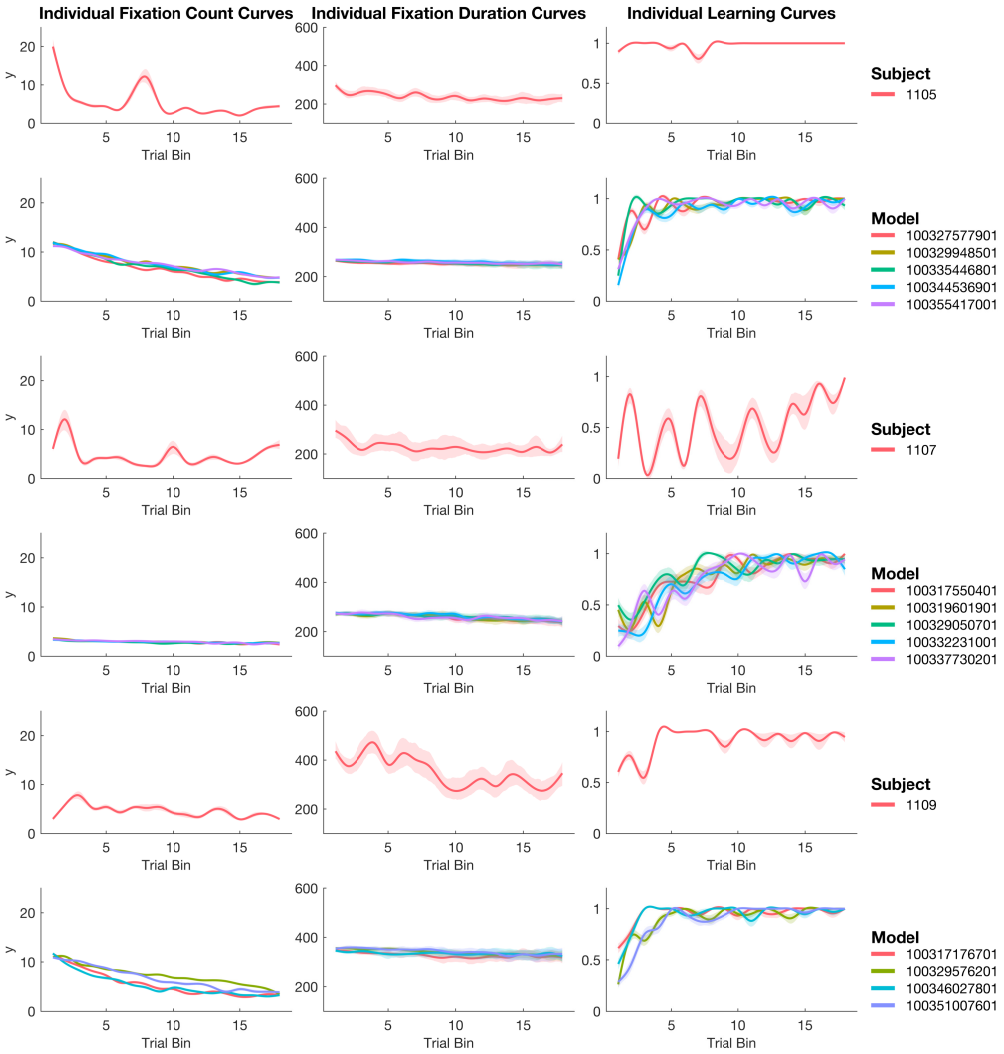


Figure 31. Subjects: 1105, 1107, 1109. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

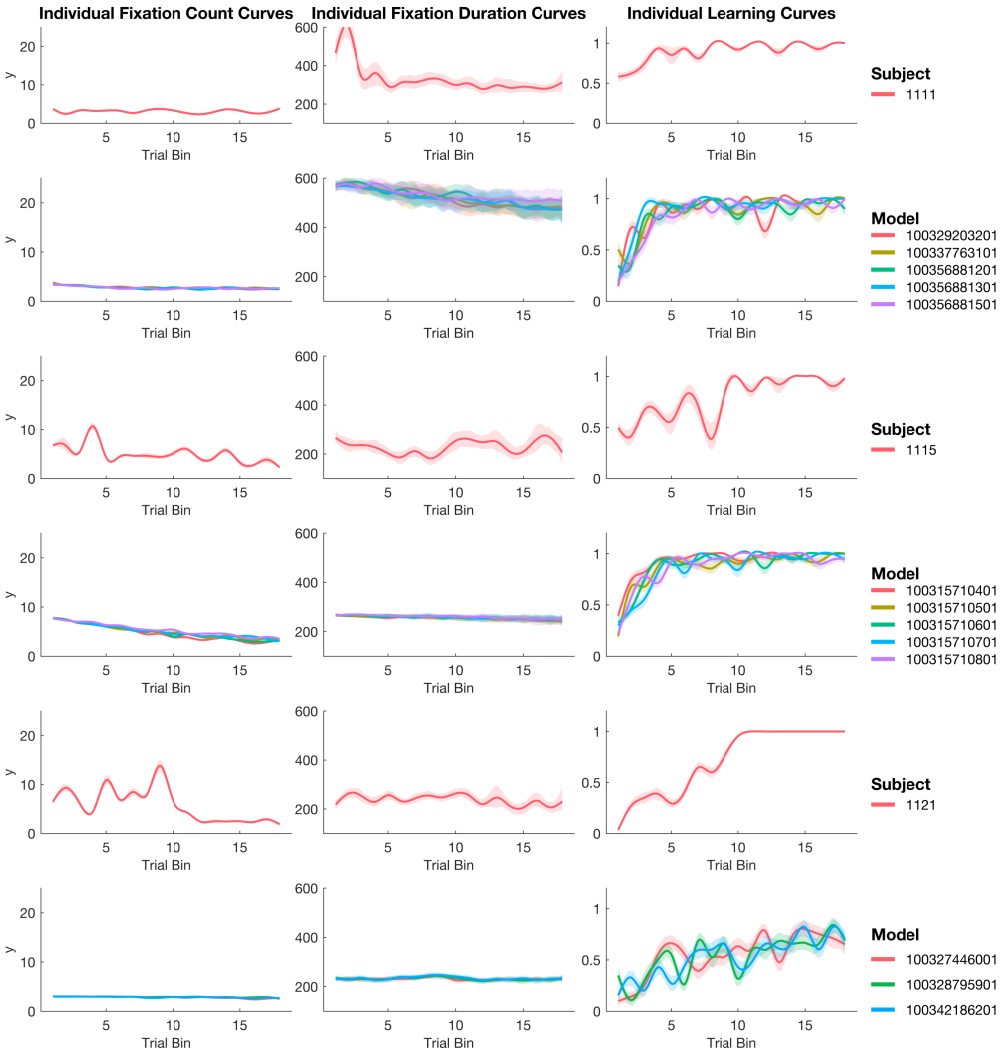


Figure 32. Subjects: 1111, 1115, 1121. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

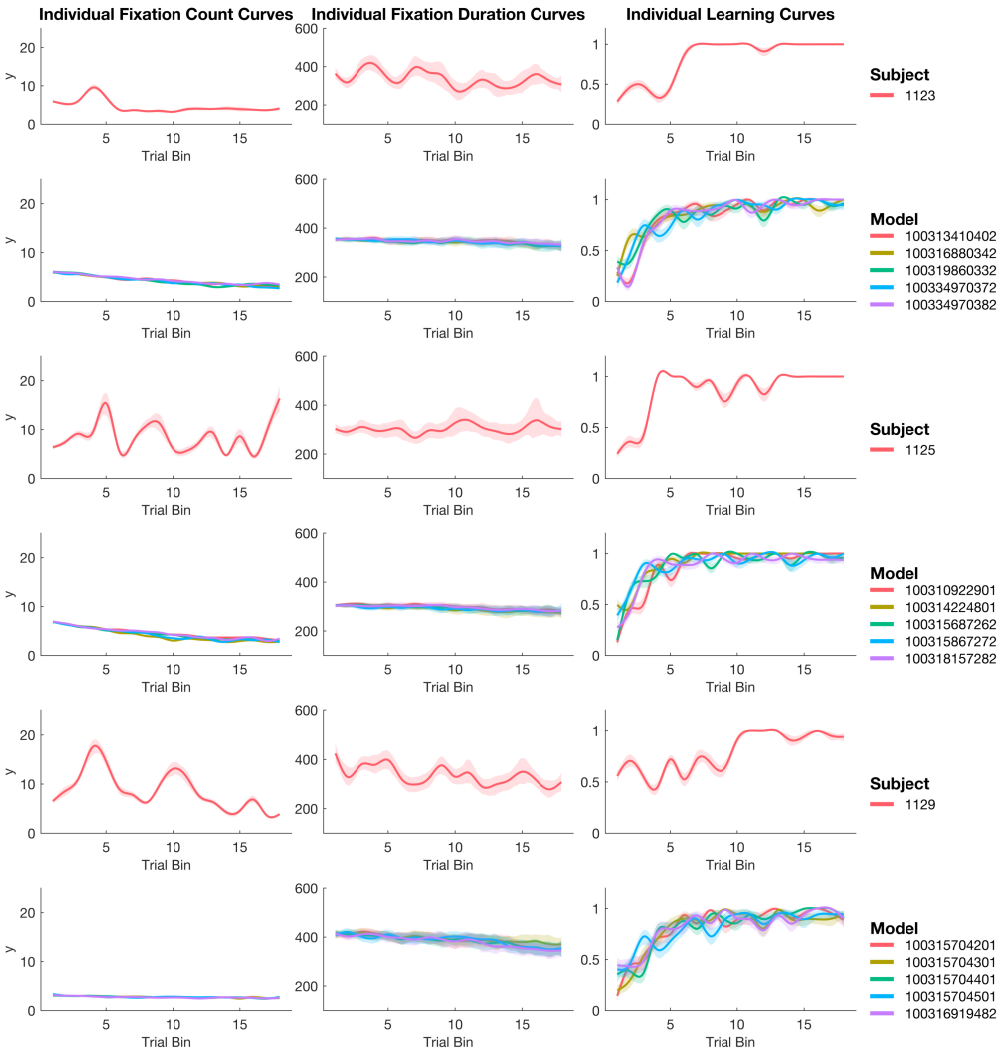


Figure 33. Subjects: 1123, 1125, 1129. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.

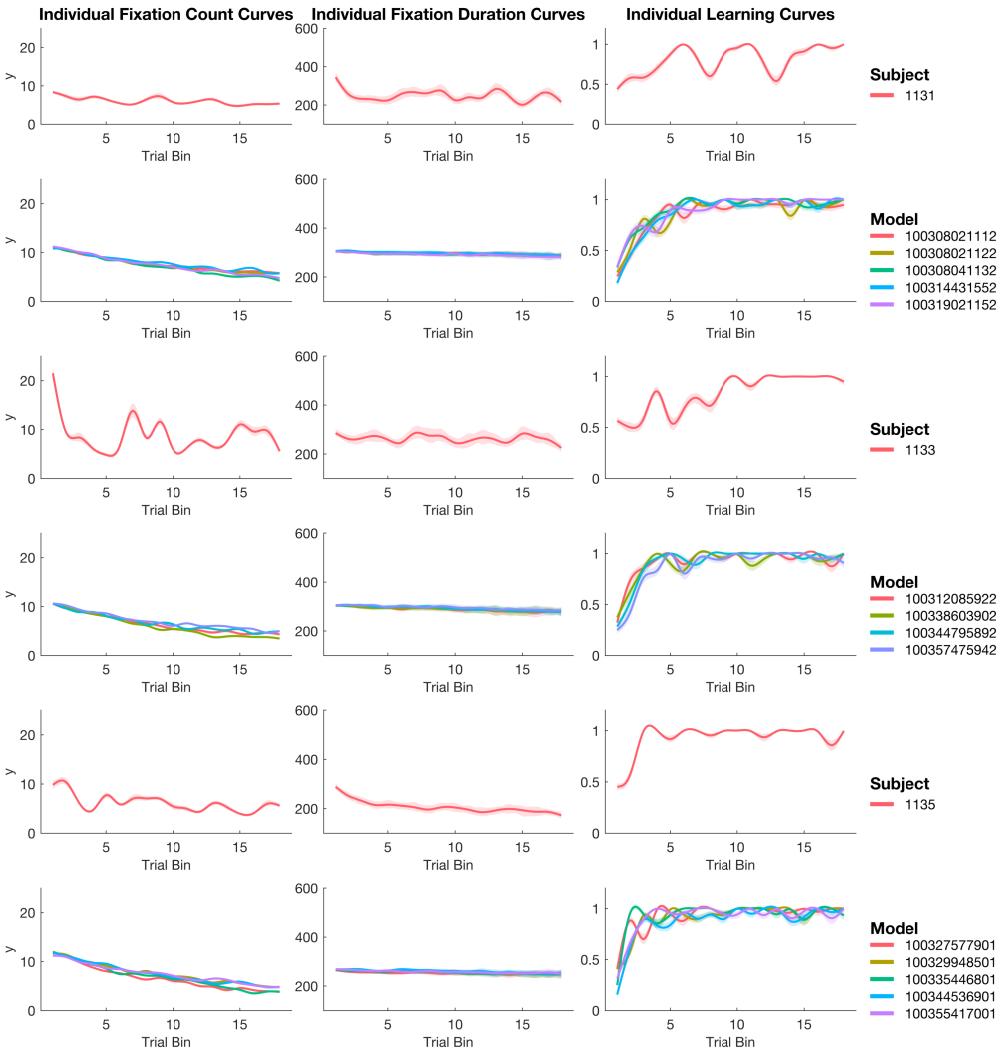


Figure 34. Subjects: 1131, 1133, 1135. Subject fixation counts, fixation durations, and learning curves are contrasted with their best fitting sample of models in the next row.