Article:
Weichart, Emily, Unger, Layla, King, Nicole et al. (2 more authors) (2024) 'The eyes are the window to the representation':Linking gaze to memory precision and decision weights in object discrimination tasks. Psychological Review. ISSN: 0033-295X

https://doi.org/10.1037/rev0000475

'The eyes are the window to the representation': Linking gaze to memory precision and decision weights in object discrimination tasks

Emily R. Weichart*†, Layla Unger‡, Nicole King‡, Vladimir M. Sloutsky‡, and Brandon M. Turner‡
†Utah State University
‡The Ohio State University

Humans selectively attend to task-relevant information in order to make accurate decisions. However, selective attention incurs consequences if the learning environment changes unexpectedly. This trade-off has been underscored by studies that compare learning behaviors between adults and young children: broad sampling during learning comes with breadth of information in memory, often allowing children to notice details of the environment that are missed by their more selective adult counterparts. The current work extends the generalized context model (Nosofsky, 1986) to account for both the intentional and consequential aspects of selective attention when predicting choice. In a novel direct input approach, we used trial-level eye-tracking data from training (memory precision) and test (decision weights) to replace the otherwise freely-estimated attention dynamics of the model. We demonstrate that only a model imbued with gaze correlates of memory precision in addition to decision weights can accurately predict key behaviors associated with 1) selective attention to a relevant dimension, 2) distributed attention across dimensions, and 3) flexibly shifting strategies between tasks. Although humans engage selective attention with the intention of being accurate in the moment, our findings suggest that its consequences on memory constrain the information that is available for making decisions in the future.

Keywords: attention; categorization; decision-making; encoding and retrieval; eye-tracking; model-based cognitive neuroscience

Introduction

Humans can effortlessly integrate multiple sources of information when making everyday decisions, drawing upon their existing knowledge and cues from the current environment. The way we balance and prioritize information may vary based on a number of factors, including task demands, source salience, and personal goals. Take, for instance, the task of distinguishing between edible and poisonous berries in the wilderness. In this scenario, it can be advantageous to learn and apply highly reliable rules based on a single dimension, such as color (e.g. "White and yellow–kills a fellow. Purple and blue–good for you."). Conversely, when identifying the species of an unfamiliar plant that has sprouted in one's garden, it becomes more useful to use a broader range of information. One can consider dimensions such as flower shape, leaf arrangement, and petal pattern, and make a category inference based on its overall resemblance to a known type of plant. In both examples, the process of object discrimination is influenced by two key factors: 1) the knowledge that one has stored in memory about how features correspond to categories; and 2) how one strategically weights information about the current item in order to make appropriate decisions.

Across theoretical models of human learning and categorization, this strategic weighting of different dimensions is formalized as *attention* (Rescorla & Wagner, 1972; Mackintosh, 1975; Pearce & Hall, 1980; Medin & Shaffer, 1978; Nosofsky, 1986). Under standard assumptions, maximum accuracy may be achieved by selectively allocating attention to dimensions that provide category-diagnostic information (e.g. berry color), and ignoring those that may provide irrelevant, unreliable, or conflicting information. Decades of empirical findings have shown that humans iterate toward an optimal distribution of attention when pursuing accuracy goals (see Weichart et al., 2022, for review). For example, eye-tracking findings have demonstrated that categorization accuracy is commensurate with gaze patterns that prioritize diagnostic over irrelevant dimensions (Rehder & Hoffman, 2005a,b; Meier & Blair, 2013; Blair et al., 2009; Galdo et al., 2022).

Despite apparent intentions to balance information in a way that will yield high accuracy, learners often fail to ad-

*Corresponding author: Emily R. Weichart, Department of Psychology, Utah State University, Logan, UT, USA. Manuscript accepted for publication on January 16, 2024.

just attention accordingly in cases when the task environment suddenly changes. Learners who optimize attention for categorization fail to respond accurately when tested on item recognition (Griffiths & Mitchell, 2008; Deng & Sloutsky, 2016), fail to utilize other features in cases when the prioritized dimension does not contain usable information (Deng & Sloutsky, 2016; Kruschke et al., 2005), and fail to maintain stable accuracy if dimension reliability shifts at some point during the task (Blanco & Sloutsky, 2019; Blanco et al., 2023).

There are two likely interpretations of these findings. The first is that the observer maintains inaccurate beliefs about the continued relevance of a particular dimension despite changes in the task environment (e.g. Rich & Gureckis, 2018). The second is that there are "costs" of engaging selective attention during learning; specifically, limited encoding of information that is not immediately relevant, but may become relevant in the future (Blanco & Sloutsky, 2019; Best et al., 2013; Plebanek & Sloutsky, 2017). The difference is theoretically significant: do failures and biases emerge due to one's inaccurate beliefs about the *current* environment, or as an unintended consequence of storing information that was relevant in a *previous* environment? Within leading mechanistic theories of categorization, however, it is not possible to differentiate between the extent to which features are encoded in memory, and the extent to which dimensions are prioritized during discrete decisions (Medin & Shaffer, 1978; Nosofsky, 1986; Kruschke, 1992).

To address this gap, we present a gaze-based extension to a standard exemplar-similarity categorization model (Generalized Context Model; GCM; Nosofsky, 1986) and investigate the independent contributions of memory for features acquired during learning and the dimension-level weights that govern decisions about new items. Our framework builds upon intuitions from seminal work that established gaze measures as an analogue of decision weights during categorization (Rehder & Hoffman, 2005b), lending two important innovations. First, we incorporated gaze measures as a *direct input* to the GCM specification for attention. While the standard approach is to infer a single average distribution of attention based on a post-learning pattern of behavioral responses, our framework uses trial-level measures of gaze to replace these typically freely-estimated dynamics. We thus gain an advantage of detailed, data-driven insight into the feature information that contributed to each individual choice. Second, we constructed our framework to allow for the possibility that gaze not only provides an analogue to decision weights, but memory precision as well. Although it is often assumed that all information presented to the participant is plausibly stored in memory, we incorporate a simple yet critical intuition: features can only be stored in memory to the extent that they are looked at during initial learning.

By leveraging eye-tracking data within a joint modeling framework for predicting choice, we are uniquely positioned to investigate the hypothesis that attention does not purely represent the observer's moment-to-moment weighting of dimensions, but rather is subject to the constraints of previously-encoded information as well. Our goal is to reconcile traditional views of attention with modern insights, and highlight memory as a critical factor for understanding how even well-intentioned learners can fail to make well-informed decisions.

The exemplar-similarity framework

The *exemplar-similarity* framework has subserved the majority of model-based categorization accounts for the past several decades (Medin & Shaffer, 1978; Estes, 1986; Nosofsky, 1986; Kruschke, 1992, 2001; Love et al., 2004; Galdo et al., 2022). GCM is a prime example of the exemplar-similarity framework, as it is one of the most influential and widely-implemented models in cognitive psychology.

Within the exemplar-similarity framework, categorizing a new stimulus requires the observer to determine its similarity to labeled exemplars of each available category (Figure 1A-B). GCM made a significant contribution by positing that the structure of the psychological space is modified by a latent distribution of attention (Nosofsky, 1986). The observer assigns an attention weight α_j to each dimension j, where $0 \leq \alpha_j \leq 1$ and $\sum_k \alpha_k = 1$. Importantly, however, GCM makes the simplifying assumption that the features of all previously-encountered exemplars are perfectly encoded in memory. The attention weights therefore serve to "stretch" the dimensions of psychological space that are attended, and "shrink" those that are unattended. The consequence is that the observer is more likely to perceive differences between features that occur in attended as opposed to unattended dimensions. Returning to our earlier example, people who selectively attend to the color dimension when categorizing berries as "edible" or "poisonous" would be likely to perceive minor distinctions in the spectrum from white to blue, but unlikely to notice variability in the unattended dimension of leaf shape.

Limitations of free estimation

Using the mechanisms described above, exemplar-similarity models such as GCM can take vectorized versions of stimulus features as input, and generate response probabilities as output. By fitting a model to data, one can identify parameter values that closely approximate the behaviors that participants actually produced. The purpose would be to distill a set of responses collected over the course of an experiment into mechanistic information, such as a dimension-wise distribution of attention.

As an example, Medin & Smith (1981) designed an experiment to investigate the impact of different strategy-targeted task instructions on category learning. Responses differed

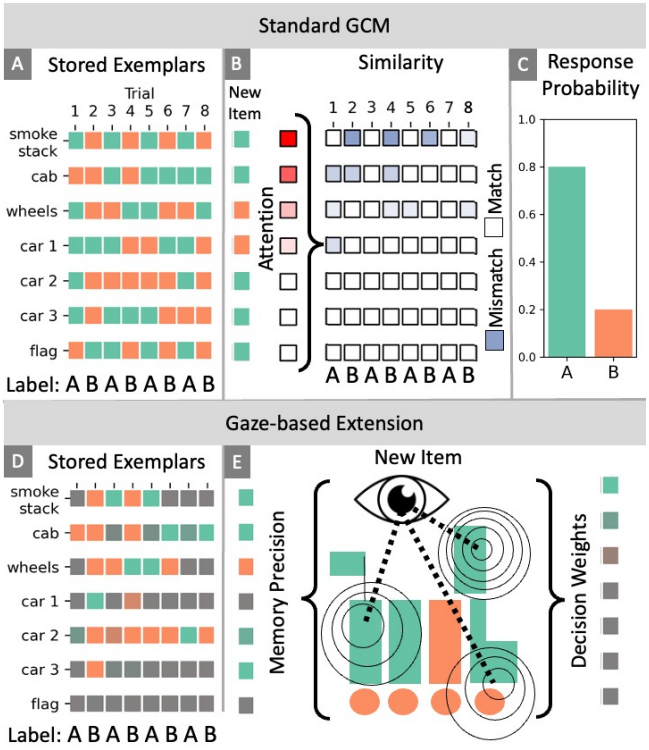

Figure 1

Exemplar-similarity framework. *GCM=Generalized Context Model. (A) Labeled exemplars are stored in memory as vectors of feature information. Here, green and orange squares represent features that were drawn from unseen prototypes of Categories A and B, respectively. (B) The observer compares the features of a new to-be-categorized item to those of the stored exemplars. Feature-level similarity is impacted by a distribution of attention, such that features of highly-attended (deeper hues of red) dimensions result in better discriminability between matching and mismatching features. (C) Exemplars that are perceived to be more similar to the new item are assigned a higher "activation" value. Response probability is a ratio of activation values between Categories A and B. (D) In the proposed gaze-based extension to GCM, exemplar features are stored in memory in proportion to how long they were fixated during learning. Gray hues represent low memory precision. (E) When the observer processes a new item, gaze patters are presumed to provide insight into both which features were plausibly encoded into memory, and how features were weighted during the categorization decision.*

considerably between groups who were instructed to respond based on a rule tied to a specific dimension, or based on overall similarity to category prototypes. Model-based analyses revealed that the former instructions induced a strategy of *selective attention* biased toward the relevant dimension, whereas the latter instructions prompted a strategy of *distributed attention* across dimensions.

According to GCM, being explicitly told which dimension was most category-diagnostic would impose no decrements upon the observer's motivation to encode the features of the *other* dimensions. Instead, any differences in behavior between groups in Medin & Smith (1981)'s design would be attributed to "stretching" and "shrinking" the dimensions of a perfectly-encoded store of exemplars. Here, we propose a more nuanced explanation whereby a latent distribution of attention indeed describes the observer's ongoing weighting of information, but also impacts the precision with which features are stored in memory over the course of learning. As shown in Figure 1E, our extension to GCM determines the psychological similarity between the trial stimulus and past exemplars based on both 1) the availability of exemplar features in memory; and 2) the distribution of decision weights applied to the current stimulus.

Although the standard implementation of GCM has been criticized as being overly simplifying due to its assumption of perfect exemplar encoding (Murphy, 2002; Griffiths & Mitchell, 2008), this assumption has continued to be presented as a necessity for computational constraint since the earliest introduction of the exemplar-similarity framework (Medin & Shaffer, 1978; Nosofsky, 1986). The current work is the first modeling effort to investigate the forces of memory precision and decision weights independently, using gaze to disentangle what has previously been viewed as inextricable.

Memory and decision subcomponents of attention

Several empirical findings suggest that memory and decision weights bear dissociable impacts on object discrimination judgements. First, a vast literature on *blocking* has advanced our knowledge about the impacts of selective attention on the processing of information from unattended dimensions (Beesley & Le Pelley, 2011; Le Pelley et al., 2007; Kruschke & Blair, 2000). In a typical blocking design, participants are pre-trained to associate a cue, A, with an outcome. In a second phase, a compound cue AB is then associated with the same outcome. Because participants learn to associate A with the outcome during pre-training, they fail to learn the relationship between B and the outcome during the second phase. Acquisition of knowledge about B is thus said to be "blocked" by the more predictive cue A (Kamin, 1968).

In one relevant study, Griffiths & Mitchell (2008) tested participants on recognition and perceived outcome causality of each cue presented during a blocking procedure. In

addition to blocking effects (B rated less causally related to the outcome than A), the authors noted significantly reduced recognition performance for B in comparison to A, even after controlling for cue frequency. Another study conducted by Easdale et al. (2019) administered a similar multi-cue design alongside eye-tracking measures, and manipulated how reliably each cue predicted the outcome. The authors found that by reducing causal certainty with probabilistically- rather than deterministically-predictive cues, participants fixated to a wider breadth of cues before making decisions. These findings call simplifying assumptions of a perfectly-encoded memory store into question, instead suggesting that selective attention bears a significant impact on the very formation of the representation itself.

Perhaps the most revealing findings on the dichotomy of memory precision and decision weights has come from work on the development of selective attention from young childhood to adulthood (Best et al., 2013; Deng & Sloutsky, 2015a,b, 2016; Plebanek & Sloutsky, 2017). It is well-documented that adults optimize attention in a goal-directed manner, and tend toward a strategy of selectively attending to dimensions that most reliably result in accurate responses (Shepard et al., 1961; Treisman, 1969; Duncan, 1984). Young children, however, are less equipped to ignore irrelevant information and are therefore more likely than adults to use a strategy of distributed attention during learning (Smith & Kemler, 1977; Blanco et al., 2023).

In a study by Deng & Sloutsky (2016), stimuli were composed of one dimension that was perfectly deterministic of category membership, and six dimensions that were each probabilistically predictive. The authors observed that after training, adults systematically categorized new items according to the deterministic dimension and remembered its features substantially better than features in the other dimensions. By contrast, young children were more likely to utilize multiple dimensions to categorize new items and showed good memory for all features, even outperforming adults on recognition of features in probabilistic dimensions (Experiment 3).

Blanco et al. (2023) further investigated the costs of selective attention by collecting behavioral and eye-tracking data from adults and children while they completed a two-phase learning task, using stimuli similar to Deng & Sloutsky (2016). After learning to categorize stimuli during Phase 1, the most- and least-informative dimensions suddenly swapped roles to mark the onset of Phase 2. Analyses of eye-tracking data showed that adults primarily fixated to the deterministic dimension during Phase 1, then rapidly shifted attention to the probabilistic dimensions at the onset of Phase 2. By contrast, children were more likely to fixate to a broader range of features during Phase 1, and tended to maintain this strategy during Phase 2. Importantly, the authors identified a subset of children who responded more consistently with the deterministic dimension than adults in Phase 2. This finding suggests that encoding a broad range of information may offer an advantage when determining the most category-diagnostic dimension in a changing environment.

These developmental findings point to intriguing nuances that are missed by the standard definition of attention (Nosofsky, 1986, 1991; McKinley & Nosofsky, 1996). If strategic allocation of attention occurs only after all information is already encoded, one cannot explain why optimality-seeking adults would incur robust costs to accuracy that children are demonstrably able to avoid (Deng & Sloutsky, 2016; Blanco et al., 2023). It is perhaps the case that relative to engaging selective attention during learning, a strategy of broad feature sampling and encoding information in memory uniquely supports flexible adaption to a changing task environment.

Experiment

We administered an eye-tracking version of a paradigm developed by Deng & Sloutsky (2015a, 2016). In the task, participants are first trained to map stimulus features to categories via feedback-based learning. Participants then complete a *recognition* test phase (i.e. "Have you seen this exact item before?") and a *categorization* test phase (i.e. "Does this item belong to [Category A] or [Category B]?"). Importantly, feedback is not provided during the test phases, so participants must rely on the information they learned during training to make judgements about the new stimuli they encounter during test.

We selected this paradigm in light of robust empirical evidence that humans naturally distribute attention differently to serve recognition and categorization goals (Ashby & Lee, 1991; Maddox & Ashby, 1996; Little & Lewandowsky, 2009; Greene & Oliva, 2009). In rule-based categorization, optimal performance is achieved by selectively attending to relevant dimensions. On the other hand, recognizing an item among highly similar exemplars requires attention to be broadly distributed across many or all dimensions. The inclusion of both recognition and categorization phases in our study was therefore meant to provoke strategic differences within-subjects.

We additionally wanted to study the impacts of feature encoding during training on subsequent strategic flexibility between-subjects. Some ways to induce sampling variability would be to include manipulations of task instructions (Medin & Smith, 1981), feature salience (Liu et al., 2015), predictive certainty (Easdale et al., 2019; Beesley & Le Pelley, 2011), or mode of feedback (Meier & Blair, 2013; Little & Lewandowsky, 2009), which have been shown to impact the extent to which adults engage selective attention during learning. However, the benefit of variability would arguably come at the cost of interpretability in the current work, given that these interventions would interact with the strategies that

participants would naturally use in pursuit of optimal responding. We therefore took an alternative approach, and selected a participant population that has demonstrated both effective learning across paradigms and widespread strategic variability: preschool-aged children.

Previous work has shown that 4- to 5-year-old children are more likely than adults to use a strategy of distributed attention when behavioral effects are aggregated across subjects (Deng & Sloutsky, 2016; Plebanek & Sloutsky, 2017). Subject-level analyses, however, suggest that these effects can be attributed to higher strategic variability among children compared to adults, rather than children being unilaterally unable to engage selective attention.

Blanco & Sloutsky (2019), for example, classified individual adults and children in terms of attention strategies used during a learning task. The distribution of strategy usage for adults was 66% selective, 19% distributed, and 16% intermediate. For children, the strategy distribution was more even with 29% selective, 32% distributed, and 38% intermediate (Experiment 1). In the current study, we hoped that this variability among children would provide the opportunity to identify robust strategy groups that were suitable for between-subject comparisons of gaze patterns.

Methods

Participants

Participants were 219 children who were recruited from preschools and childcare centers in the suburbs of Columbus, Ohio (mean age: 52.0 months, range: 44.7-58.1 months). All research activities were approved by the Institutional Review Board at The Ohio State University (Protocol 2004B0422). Written informed consent was acquired from a parent or guardian of each participant in advance of the study, and the children themselves consented verbally. We used a larger sample size than what is typical for the selected paradigm (N=25-35; Blanco & Sloutsky, 2019; Deng & Sloutsky, 2012, 2015a, 2016). This was done in consideration of recommended sample sizes in excess of 100 for model comparison (Myung & Pitt, 2004), and in an effort to observe individual differences in strategy within the population of interest.

This study was not pre-registered. Data and model code are publicly available and are hosted by the Open Science Foundation (OSF; https://osf.io/9r7k8/?view_only=ca4010ffb8aa4cbea6b02be3ff8ad80f).

Materials

Training stimuli were colorful drawings of trains that were divided into categories that we denote "A" (Category A) and "B" (Category B). As shown in Table 1, each category was represented by an un-presented prototype. Prototypes contained seven features that were distinct in shape and color: smoke stack, cab, wheels, car 1, car 2, car 3, and flag. A majority of the features were drawn from the prototypes *probabilistically* so that they would collectively represent the overall similarity among category exemplars (henceforth referred to as "P" features). One feature, however, was perfectly *deterministic* of category membership (henceforth referred to as the "D" feature). The D dimension was selected among three options for each participant (cab, wheels, and flag), and selections were counterbalanced between-subjects.

Four stimulus types that were presented during the experiment will be discussed.* The stimulus structure of each item type is shown in Table 1. Each item type configuration discussed below resulted in 30 possible stimuli, 15 from each category as determined by the D feature (*high-match*, *conflict*, and *one-new-P*) or by the majority of P features (*new-D*).

1. The majority of stimuli were a *high-match* to one of the two category prototypes, meaning the D feature and 4 out of 6 P features were drawn from a consistent prototype. The remaining 2 P features were drawn from the opposite prototype. High-match items were presented both during category training (with labels), and in subsequent recognition and categorization test phases (without labels).

2. *Conflict* items contained the D feature and 2 out of 6 P features from one category prototype, and the majority (4 out of 6) of the P features from the other. These items were only presented during tests of memory and categorization, and were never paired with labels.

3. *New-D* items contained a novel feature in the D dimension, which was never explicitly paired with a label during category training. 4 out of 6 P features were drawn from one category prototype, and the remaining 2 features were drawn from the other. These items were only presented during tests of memory and categorization, and were never paired with labels.

4. *One-new-P* items contained a novel feature in a randomly-selected P dimension. The D feature and 4 out of 6 P features were drawn from one category prototype, and the remaining P feature was drawn from the other. These items were only presented during tests of memory and categorization, and were never paired with labels.

*A fifth *all-new-P* item type was presented to participants as well, which contained novel features in all 6 P dimensions. The D feature was drawn from one of the available category prototypes. These items were not related to the effects of interest in the current study, and were therefore excluded from analysis.

								Recognition: Correct Response	**Categorization:** Dimension-Consistent Response	
Item Type	**D**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**		**D**	**P**
Prototypes	0	0	0	0	0	0	0	-	A	A
	1	1	1	1	1	1	1	-	B	B
High-match	0	0	0	0	0	1	1	Old	A	A
	1	1	1	1	1	0	0	Old	B	B
Conflict	0	1	1	1	1	0	0	New	A	B
	1	0	0	0	0	1	1	New	B	A
One-New-P	0	0	0	0	0	1	N	New	A	A
	1	1	1	1	1	0	N	New	B	B
New-D	N	0	0	0	0	1	1	New	-	A
	N	1	1	1	1	0	0	New	-	B

Table 1

Category Structure. *D=deterministic; P=probabilistic; N=novel. Rows provide examples of feature configurations for each type of item presented during the task. Values correspond to unique features in each dimension. 0s correspond to an unseen prototype from Category A, and 1s correspond to a prototype from Category B. "D" and "P" headings refer to the reliability of feature information in the corresponding dimension. The three right-most columns indicate expected responses, considering the feature information provided by the relevant item type.*

Procedure

The experiment was similar to that of Deng & Sloutsky (2016), and was comprised of four phases: instructions, training, recognition test, and categorization test. Instructions and prompts that were specific to each trial were read aloud by a trained experimenter, and participants responded verbally. The experimenter then pressed the corresponding key on the keyboard to log the response. The experiment lasted approximately 20 minutes in total.

During the *instructions*, participants were told that they would see different trains and that they would have to decide which ones belonged to Categories A and B. Features drawn from each category prototype were presented on the screen in isolation, and the experimenter verbally indicated the appropriate category association. In particular, P features were displayed alongside a message in the form: "Most of the [A / B] trains have this type of [e.g. smoke stack / car / cab / wheels]." D features were accompanied by the message: "All [A / B] trains have this type of [e.g. flag]." Across two categories, 14 features and their associated category mappings were displayed to participants during the instructions. The experimenter read the following message aloud before the experiment began: "There are two parts in this game. This is the first part. In this part of the game, you will see many trains. Some of them are A trains and some are B trains. You will tell me whether it's an A train or a B train."

The *training* phase consisted of 30 high-match items (15 per category). During each trial, stimuli were presented in the center of the screen and participants were asked the question: "What is this? A or B?" After the the experimenter logged the participant's response, corrective feedback was provided in the form of "Correct! This is a/n [A / B] train" or "Oops! This is actually a/n [A / B] train." Additionally, feedback highlighted the D feature and similarity to category prototypes with a message in the form of "It looks like a/n [A / B] train and has the [A / B] [e.g. flag]." Feedback was presented as text on the screen and read aloud by the experimenter. The order of stimulus presentation was randomized across participants.

Training was followed by recognition and categorization test phases. At the point of transition between training and test, the experimenter read the following message aloud: "Now, it's the second part of the game. In this part, you will see more trains. You saw some of them in the first part of the game, but some of the trains are new. You did not see them in the first part. You will tell me whether it's an A train, or B train. Also, you will tell me whether you saw exactly the same train in the first part, or if it's new." Each of the two test phases contained 40 trials (20 per category). Eight items were presented from each of the four types shown in Table 1.

During each trial of the *recognition test* phase, participants were presented with a stimulus and were asked, "Did you see exactly the same train in the first part of the game?" Participants responded "yes" if they believed the stimulus had been presented during training, or "no" if they believed the stimulus was new. No feedback was provided after the experimenter logged the participant's response; the experiment

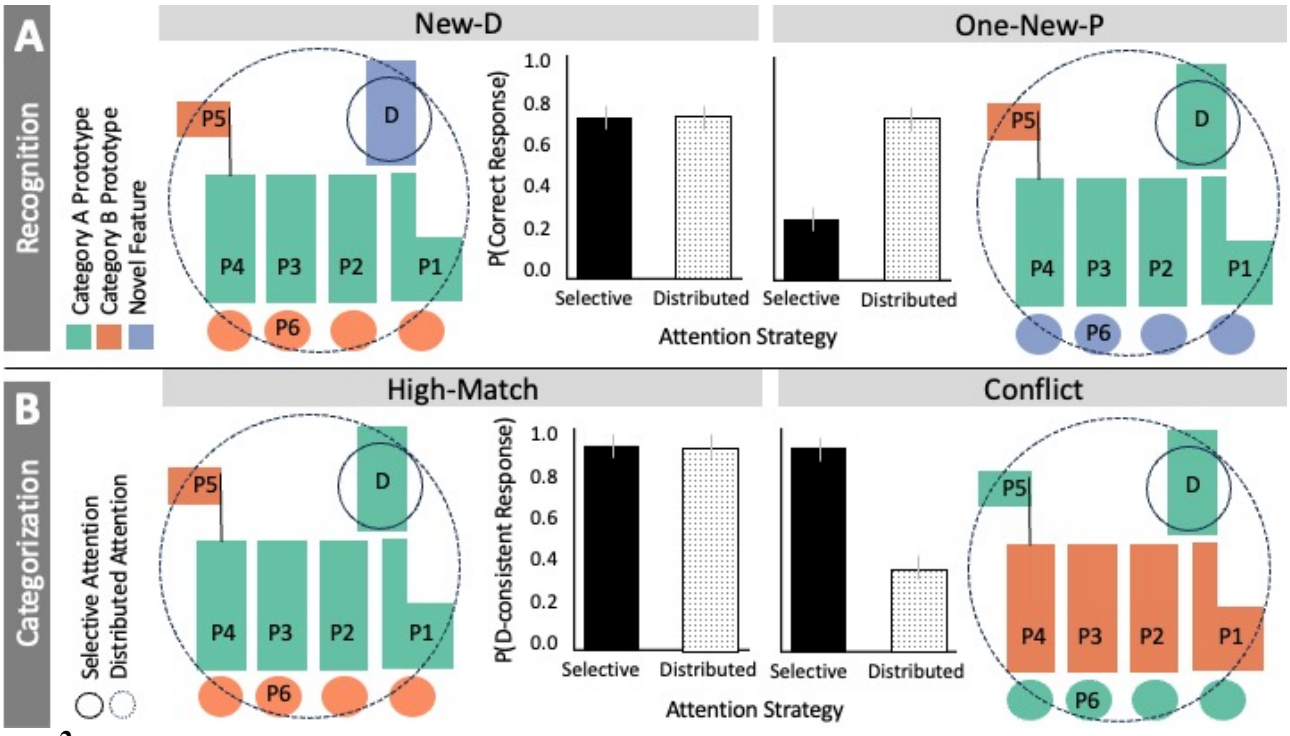

Figure 2

Predictions of key behavioral effects. *D=deterministic; P=probabilistic. (A) During recognition, a strategy of distributed attention should result in correct rejections of both new-D and one-new-P items as "new". A Strategy of selective attention should the D dimension should result in a reduced ability to correctly reject one-new-P items. (B) During categorization, a strategy of selective attention should result in high proportions of responses consistent with the D feature during both high-match and conflict items. A strategy of distributed attention should result in a lower proportion of D-consistent responses during conflict items.*

simply proceeded to the next trial. As shown in Table 1, only high-match items were correctly considered to be old, whereas items drawn from conflict, new-D, and one-new-P types were new.

During the *categorization test* phase, participants were presented with a stimulus and were asked, "What is this? A or B?" As in the recognition test phase, no feedback was provided after the experimenter logged the participant's response.

Eye-tracking

Throughout the experiment, monocular gaze fixations were recorded using an EyeLink 1000 eye tracker (SR Research, Ontario, Canada) at a sampling rate of 500 Hz with a manufacturer-reported accuracy of 0.5°. Participants were seated 60 cm from the eye tracker, facing a 1280 x 1024-pixel display monitor. To analyze the data, we defined seven rectangular areas of interest (AOIs) that were centered at the spatial location of each dimension. AOIs varied in size from 2° x 2° (flag) to 4.3° x 4.7° (cab). When preprocessing the data, we calculated the total time that a participant's gaze overlapped with a particular AOI at the level of each trial (Blanco et al., 2023).

Analysis

Deng & Sloutsky (2016) defined key behavioral effects for evaluating how adults, 7-year-olds, and 4-year-olds allocate attention during the task described above. Here, we conducted analyses for identifying subgroups of participants who demonstrate these key behaviors. Because Deng & Sloutsky's key effects pertain to aggregate group-level behaviors rather than individual subjects, we first classify participants into groups using individual-level, model-based cognitive assessment techniques (Weichart et al., 2021; Wiecki et al., 2015). We then conduct comparisons between groups to verify that the contrasting behavioral correlates of attention described by Deng & Sloutsky (2016) are indeed observable within the current participant pool, despite controlling for age.

Model	Recognition	Categorization	N free α parameters
$Sel_R - Sel_C$	$\alpha_D > \alpha_P$ (Selective)	$\alpha_D > \alpha_P$ (Selective)	2
$Sel_R - Dist_C$	$\alpha_D > \alpha_P$ (Selective)	$\alpha_D = \alpha_P$ (Distributed)	1
$Dist_R - Sel_C$	$\alpha_D = \alpha_P$ (Distributed)	$\alpha_D > \alpha_P$ (Selective)	1
$Dist_R - Dist_C$	$\alpha_D = \alpha_P$ (Distributed)	$\alpha_D = \alpha_P$ (Distributed)	0

Table 2

Comparing models with freely-estimated attention. α=*freely-estimated attention parameters in the Generalized Context Model. The table provides parameterizations of attention from four candidate models. Model comparison was used to identify which attention strategies each participant used during recognition and categorization.*

Effects of interest

During the recognition test, it is of particular interest to compare correct rejections of *new-D* and *one-new-P* items. As illustrated in Figure 2A, participants who use either a strategy of selective or distributed attention during the recognition test should be equipped to notice when a novel feature appears in the D dimension. Although participants who selectively attend to D may be *more* sensitive to novel D features than participants who distribute attention broadly, all participants are expected to be plausibly adept at correctly rejecting new-D items as "new." Participants who selectively attend to the D dimension, however, should fail to notice if a novel feature appeared in one of the unattended P dimensions.

During the categorization test, it useful to compare responses between *high-match* and *conflict* items. While all participants who learned the task should be expected to accurately categorize high-match items, conflict items should yield different response profiles between strategies (Figure 2B). Because conflict items contain a D feature from one category prototype and the majority of P features from the other, participants who distribute attention evenly across dimensions should respond close to chance, while those who selectively attend to D should respond consistently with the D dimension.

The key effects shown in Figure 2 will serve as an essential benchmark for model evaluation.

Identifying strategy groups

We applied a suite of GCM variants (Nosofsky, 1986) with separate freely-estimated distributions of α for recognition and categorization, and used a switchboard analysis to characterize individual-level attention (Turner et al., 2018). Our approach follows from relatively recent efforts in *model-based cognitive assessment*, in which well-established cognitive models are used to delineate participants according to the latent mechanisms that plausibly underlie their behaviors (Weichart et al., 2021; Weichart & Sederberg, 2021; Darby & Sederberg, 2022). Here, the relevant mechanism for delineation is attention, and the four variants of interest are summarized in Table 2. To instantiate selective attention, we freely-estimated the value of attention corresponding to the D dimension (α_D) with constraints $\frac{1}{7} < \alpha_D < 1$ and calculated attention to each P dimension as $\alpha_{P_n} = \frac{1-\alpha_D}{6}$. For distributed attention, α_k values corresponding to all seven dimensions were fixed to $\frac{1}{7}$. In both cases,

$$\sum_k \alpha_k = 1$$

per convention (Nosofsky, 1986). After fitting the models to data, we identified a preferred model for each subject via comparison of Akaike Information Criterion (AIC; Akaike, 1974) values. Because comparisons via AIC favor parsimonious models, participants were only determined to use selective attention if the addition of a free α_D was justified by a sufficient improvement in model fit.

For our purposes, it was essential to identify participants who utilized some kind of discernible strategy (as opposed to random responding or making the same response on every trial) during both recognition and categorization tests. We therefore opted to exclude participants who appeared not to perform one or both of the tasks. Specifically, participants were excluded if they failed to exceed *a priori* criteria of 60% correct responses to high-match items during the recognition test (failed recognition: N=23), and categorization test (failed categorization: N=73; failure to meet either criteria: N=13). The results to follow are based on the remaining 110 participants (50.2% of the full sample). Similar criteria were imposed in a previous dual-test study on selective attention, which also resulted in a high exclusion rate of 32.7% (67.3% inclusion) despite using adult participants (Griffiths & Mitchell, 2008).

By-subject model comparisons among candidate GCM variants (Table 2) identified the following 4 strategy groups within our participant pool:

1. Selective attention during both recognition and categorization, henceforth denoted $Sel_R - Sel_C$ (N=43);
2. Selective attention during recognition, distributed attention during categorization, henceforth denoted $Sel_R - Dist_C$ (N=30);

3. Distributed attention during recognition, selective attention during categorization, henceforth denoted $Dist_R - Sel_C$ (N=20);

4. Distributed attention during both recognition and categorization, henceforth denoted $Dist_R - Dist_C$ (N=17).

Analyses of behavior within- and between- strategy groups replicated the results of Deng & Sloutsky (2016), and were consistent with the key behavioral effects shown in Figure 2. Detailed results are provided in Appendix A.

It is important to note that according to these results, participants did not necessarily use a consistent attention strategy across the recognition and categorization test phases. Specifically, Groups $Sel_R - Dist_C$ and $Dist_R - Sel_C$ used selective attention during one phase and distributed attention during the other. Although adult participants have historically shown effects consistent with selective attention across both phases of Deng & Sloutsky's design (similarly to Group $Sel_R - Sel_C$ here), the selection of young children as our population of interest provided the opportunity to additionally observe instances of distributed attention (Group $Dist_R - Dist_C$) and strategic flexibility (Groups $Sel_R - Dist_C$ and $Dist_R - Sel_C$). Given our goal of using gaze to dissociate memory precision and decision weight components of attention, analyzing the information sampling behaviors of these groups will provide uniquely rigorous theoretical constraint.

Gaze as a direct input for components of attention

As emphasized by Turner et al. (2017), developments in model-based cognitive neuroscience have provided new opportunities to link model mechanisms to neurophysiological measures for the purposes of theoretical constraint, adjudication, and elaboration (Turner et al., 2017; Palmeri et al., 2017; Turner, Palestro, et al., 2019; Turner, Forstmann, & Steyvers, 2019). In the domain of categorization, seminal work by Rehder & Hoffman (2005b,a) noted a correspondence between gaze and attention weights estimated by the exemplar-similarity framework. The authors used what would be classified by Turner et al. (2017) as a "two-stage approach" for linking behavioral and neurophysiological data, whereby the relationship between two independently-analyzed modalities is assessed through a second stage of correlation or regression analyses. Here, we go a step further and present a "direct input approach" whereby gaze *itself* serves as a mechanism for feature encoding and predicting choice.

Following previous work, we assume features that are fixated longer during training are more likely to be encoded into memory (Loftus, 1972; Peterson et al., 2001; Foulsham & Underwood, 2008), and features that are fixated longer during test reflect prioritization during decisions (Rehder & Hoffman, 2005b,a; Blair et al., 2009; Meier & Blair, 2013). We examine four linking functions for converting dwell times into correlates of memory precision and decision weights, where outputs are bound by 0 and 1 per convention (Nosofsky, 1986; Medin & Shaffer, 1978). A conceptual overview and simulation study of our approach in contrast to the conventional unitary view of attention (Medin & Shaffer, 1978; Nosofsky, 1986) are provided in Appendix B.

Methods

Modeling framework

To represent the stimulus on trial n of the training phase, we denote a vector $x^{(n)} = [x_{n,1}\ x_{n,2}\ \ldots\ x_{n,J}]$ where each element corresponds to the feature value in dimension j. After completing all N trials of the training phase, feature information about all exemplars is stored in matrix $X = [x^{(1)}\ \ldots\ x^{(N)}]^\intercal$ and associated feedback is stored in vector $F = [f^{(1)}\ \ldots\ f^{(N)}]$. During each trial i of test, the observer is presented with a stimulus probe $e^{(i)} = [e_{i,1}\ e_{i,2}\ \ldots\ e_{i,J}]$ and is expected to make an informed judgement (i.e. recognition or categorization). The probe acts as a retrieval cue to access information associated with similar stimuli that were encountered during training. To this end, the observer first computes the *feature similarity* between the probe and exemplar $x^{(n)}$ along each dimension j:

$$s_j\left(e^{(i)}, x^{(n)}\right) = exp\left(-\delta d_j\left(e^{(i)}, x^{(n)}\right)\alpha_j\right). \qquad (1)$$

Values of feature similarity range between 0 and 1, where 1 indicates that features $e_j^{(i)}$ and $x_j^{(n)}$ are perceived to be identical. In Equation 1, δ modulates the specificity of the similarity kernel. Separate values δ_R and δ_C were used for recognition and categorization. d_j represents the simple distance between values corresponding to the relevant features. Values of α represent attention, which modify the perceived distance between mismatching features. Although a single α_j is typically estimated across trials, we hypothesize that α_j should involve information specific to both the probe and the exemplar components of the comparison. We therefore specified

$$\alpha_j = \eta_j^{(n)}\zeta_j^{(i)} \qquad (2)$$

where $\eta_j^{(n)}$ represents memory precision for the feature presented in dimension j on training trial n, and $\zeta_j^{(i)}$ represents the decision weight allocated to dimension j on test trial i. By using a multiplicative rule, we ensure that usage of information during the choice is only possible if the relevant exemplar feature had a non-zero memory precision *and* the relevant probe feature had a non-zero decision weight.

The observer next computes the *overall similarity* between the probe and each exemplar, combining feature similarity across dimensions:

$$a\left(e^{(i)},x^{(n)}\right) := \prod_j s_j\left(e^{(i)},x^{(n)}\right). \tag{3}$$

Overall similarity is analogous to the *activation* of the relevant exemplar in memory, given the presence of the current probe.

The choice rules used here follow iterations of GCM that incorporated assumptions about the determinism of responding (Ashby & Maddox, 1993; Navarro, 2007). When making a choice during the recognition test phase, the relevant feature comparison d in Equation 1 is:

$$d_j\left(e^{(i)},x^{(n)}\right) = \begin{cases} 1 & \text{if } e_j^{(i)} \neq x_j^{(n)} \;\forall\; n \in \{1,2,\ldots,N\} \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

which determines whether or not a test feature was presented during training. We then determine the activation of a "new" response based on the total activation across all exemplars:

$$A(\text{“}new\text{”}) = exp\left[\phi \sum_{n=1}^{N}\left(1 - a\left(e^{(i)},x^{(n)}\right)\right)\right] \tag{5}$$

The probability of making a "new" response is given by

$$P(\text{“}new\text{”}) = \frac{A\left(\text{“}new''\right)}{A\left(\text{“}new''\right)+\beta}, \tag{6}$$

and $P(\text{“}old\text{”}) = 1 - P(\text{“}new\text{”})$. Here, β represents a baseline bias for responding "old" and ϕ is a temperature parameter for scaling the activations.

During categorization test, the relevant feature comparison d in Equation 1 is:

$$d_j\left(e^{(i)},x^{(n)}\right) = \begin{cases} 1 & \text{if } {e_j}^{(i)} \neq {x_j}^{(n)} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The activation of a given category l is given by

$$A(\text{“}l\text{”}) = exp\left[\phi \sum_{n=1}^{N} a\left(e^{(i)},x^{(n)}\right) I(f^{(n)} = \text{“}l\text{”})\right], \tag{8}$$

where $I(q)$ is an indicator function returning 1 if the condition q is true and 0 otherwise.

The probability of making a response consistent with category "A" is the ratio of activation for category "A" relative to the total activation across all available categories (which in this case is just A and B):

$$P(\text{“}A\text{”}) = \frac{A\left(\text{“}A\text{”}\right)}{A\left(\text{“}A\text{”}\right)+A\left(\text{“}B\text{”}\right)}. \tag{9}$$

Linking functions

We selected a set of increasing functions that returned outputs bound between 0 and 1 (inclusive) for converting feature-level dwell times to elements of attention within our modeling framework. The goal was to ascertain if any transformation of gaze was sufficient for predicting strategy-relevant behaviors between groups, and whether it was necessary to account for the features that were fixated during training to make accurate test predictions.

In the equations below, the input $dwell_j^{(t)}$ refers to the total time spent looking at the feature in dimension j on Trial t, and output is denoted $v_j^{(t)}$. When a given function is applied to fixations during training, output $v_j^{(t)}$ is used as $\eta_j^{(n)}$ in Equation 2 to represent memory precision for exemplar feature $x_j^{(n)}$. When a function is instead applied to fixations during test, $v_j^{(t)}$ is used as $\zeta_j^{(i)}$ in Equation 2 to represent the decision weight applied to probe feature $e_j^{(i)}$. Examples of each function are shown in Figure 3.

A: Binary step function. This function has a free threshold parameter $\theta \in (0,\infty)$, and returns 0 or 1 according to the following conditional:

$$v_j^{(t)} = \begin{cases} 0 & \text{if } dwell_j^{(t)} \leq \theta \\ 1 & \text{otherwise.} \end{cases} \tag{10}$$

B: Piecewise linear function. This function has a free threshold parameter $\theta \in (0,\infty)$, and returns an attention value as a proportion of θ. If the input exceeds the threshold, the function returns 1.

$$v_j^{(t)} = min\left(\frac{dwell_j^{(t)}}{\theta}, 1\right) \tag{11}$$

C: Softmax function. The softmax function is often used in multi-class classification problems, where the goal is to assign an input to one of several mutually exclusive classes. The function calculates the exponential of each input element and then normalizes the results by dividing each element by the sum of all exponentials. This normalization ensures that the output values sum to 1, forming a valid probability distribution. This function has a free temperature parameter $\theta \in (0,\infty)$ that scales the element-wise activations.

$$v_j^{(t)} = \frac{exp\left(\theta dwell_j^{(t)}\right)}{\sum_k exp\left(\theta dwell_k^{(t)}\right)} \tag{12}$$

D: Logistic function. The logistic function is commonly used as an activation function in neural networks because it produces non-linear transformations of the input, enabling the model to learn complex relationships between input and output variables. This function has two free parameters $\theta \in (0,\infty)$ and $\omega \in (0,\infty)$ that control the steepness and inflection point of the function, respectively.

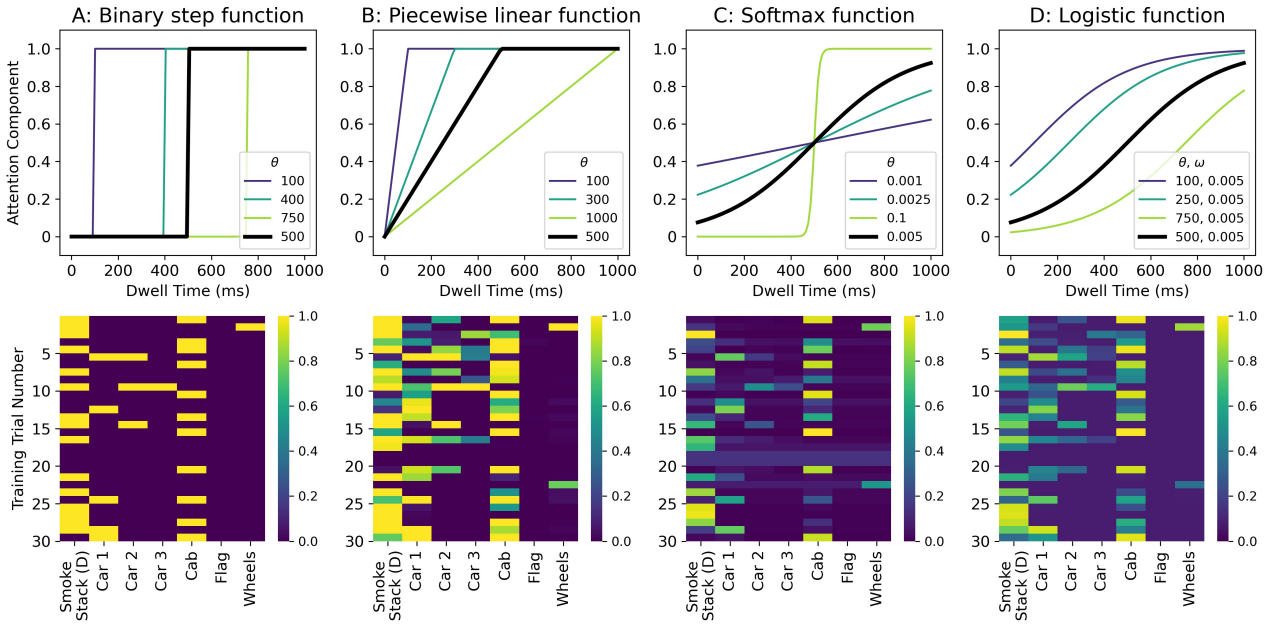

Figure 3

Linking functions. *D=deterministic,* θ *and* ω*=linking parameters. (Top) Candidate linking functions used in our investigation. X values show dwell time inputs, and Y values show outputs representing memory precision or decision weight components of attention. Colored lines illustrate changes to the function that result from modulation of free parameters* θ *and* ω*. (Bottom) Heatmaps show examples of attention outputs (Z values; colors) when applying the candidate functions to one subject's gaze data. The X axis shows stimulus dimensions. The Y axis indexes training trials.*

$$v_j^{(t)} = \frac{1}{1 + exp\left(-\theta\left(dwell_j^{(t)} - \omega\right)\right)} \tag{13}$$

Candidate Models

For our main model comparison, we identified every pairwise combination of functions for converting gaze to memory precision and decision weights. This resulted in a core set of 16 candidate models, which we refer to in the format "$X-Y$." "X" refers to a function A, B, C, or D that was applied to dwell times during training to calculate a matrix η. "Y" similarly refers to a function that was applied to dwell times during *both* test phases (recognition and categorization) to calculate a matrix ζ. Linking parameters (e.g. θ and ω) were estimated independently for memory precision and decision weights.

As specified by Equation 2, attention is calculated as the product of memory precision $(\eta_j^{(n)})$ and decision weight elements $(\zeta_j^{(i)})$. This defies the GCM convention of an attention vector that sums to a constant quantity of 1 (Nosofsky, 1986; but see Galdo et al., 2022; Weichart et al., 2022, for contradictory arguments). We therefore included a model variant "$C-C^+$" that follows the approach of Lamberts (1995) to ensure that attention varies from trial-to-trial, but is still constrained to sum to 1. As in model C-C, both $\eta^{(n)}$ and $\zeta^{(i)}$ are softmax ratios of trial-level dwell times. Instead of calculating attention as a product of the two vectors as in Equation 2, however, model $C-C^+$ uses the specification $\alpha_j = \gamma\eta_j^{(n)} + \left((1-\gamma)\,\zeta_j^{(i)}\right)$ where $\gamma \in [0,1]$.

Finally, we included 4 model variants that assumed perfect encoding of all feature information presented during training such that all $\eta_j^{(t)} = 1$. Values of $\zeta_j^{(t)}$ for decision weights were functions of each candidate linking function. These models are denoted $1-A$, $1-B$, $1-C$, and $1-D$ in the results. In total, 21 model variants were fit to data and evaluated. Details of the model-fitting procedures are provided in Appendix C.

Results

Beyond comparing the fits of the candidate gaze-based models via fit statistics, we took an additional step of evaluating each model based on its ability to predict behavioral markers of selective and distributed attention (Figure 2). To be consistent with observed key differences between groups (Tables 3-4), a successful model had to be able to predict the following:

1. Groups $Sel_R - Sel_C$ and $Sel_R - Dist_C$ (selective attention during recognition) made more false alarm "old"

Model	**Interaction**	$Sel_R - Sel_C$ $Dist_R - Sel_C$	$Sel_R - Sel_C$ $Dist_R - Dist_C$	$Sel_R - Dist_C$ $Dist_R - Sel_C$	$Sel_R - Dist_C$ $Dist_R - Dist_C$	$Sel_R - Sel_C$ $Sel_R - Dist_C$	$Dist_R - Sel_C$ $Dist_R - Dist_C$
Obs.	$F = 35.71$	$t = 9.61$	$t = 9.92$	$t = 7.03$	$t = 7.34$	$t = 1.18$	$t = 0.49$
	p<.001*	**p<.001***	**p<.001***	**p<.001***	**p<.001***	$p = .81$	$p = .99$
$B-B$	$F = 9.76$	$t = 5.08$	$t = 7.00$	$t = 2.66$	$t = 4.10$	$t = 1.66$	$t = 2.10$
	p<.001*	**p<.001***	**p<.001***	$p = .06$	**p<.01***	$p = .47$	$p = .24$
$C-A$	$F = 15.46$	$t = 6.50$	$t = 8.14$	$t = 2.94$	$t = 4.20$	$t = 2.90$	$t = 1.51$
	p<.001*	**p<.001***	**p<.001***	**p<.05***	**p<.05***	$p < .05*$	$p = .61$
$C-B$	$F = 17.36$	$t = 6.61$	$t = 8.72$	$t = 2.78$	$t = 4.45$	$t = 3.08$	$t = 2.23$
	p<.001*	**p<.001***	**p<.001***	**p<.05***	**p<.001***	$p = .08$	$p = .83$
$C-D$	$F = 15.36$	$t = 6.51$	$t = 8.52$	$t = 2.99$	$t = 4.44$	$t = 2.36$	$t = 2.27$
	p<.001*	**p<.001***	**p<.001***	**p<.05***	**p<.01***	$p = .12$	$p = .17$
$D-B$	$F = 14.85$	$t = 5.89$	$t = 7.74$	$t = 2.37$	$t = 4.14$	$t = 3.36$	$t = 1.97$
	p<.001*	**p<.001***	**p<.001***	$p = .13$	**p<.01***	$p < .01*$	$p = .30$
$D-D$	$F = 11.34$	$t = 5.87$	$t = 7.22$	$t = 3.45$	$t = 4.37$	$t = 1.49$	$t = 1.22$
	p<.001*	**p<.001***	**p<.001***	**p<.01***	**p<.01***	$p = .60$	$p = .80$
$C-C^+$	$F = 8.87$	$t = 4.53$	$t = 6.55$	$t = 2.40$	$t = 4.04$	$t = 1.73$	$t = 2.02$
	p<.001*	**p<.001***	**p<.001***	$p = .12$	**p<.01***	$p = .42$	$p = .29$

Table 3

Recognition phase: Pairwise key effects predicted by gaze-informed models. *Obs.=observed, R=recognition, C=categorization, Sel=selective, Dist=distributed. Statistical output from a 2 (feature type: D vs. P) by 4 (group) ANOVA and post hoc pairwise tests. Matching analyses were performed on observed and model-generated response data. Bold text indicates significant effects among comparisons of model-generated responses that are consistent with the observed key effects.*

Model	**Interaction**	$Sel_R - Dist_C$ $Sel_R - Sel_C$	$Sel_R - Dist_C$ $Dist_R - Sel_C$	$Dist_R - Dist_C$ $Sel_R - Sel_C$	$Dist_R - Dist_C$ $Dist_R - Sel_C$	$Dist_R - Dist_C$ $Sel_R - Dist_C$	$Sel_R - Sel_C$ $Dist_R - Sel_C$
Obs.	$F = 19.48$	$t = 4.03$	$t = 5.69$	$t = 4.75$	$t = 6.15$	$t = 1.22$	$t = 2.72$
	p<.001*	**p<.001***	**p<.001***	**p<.001***	**p<.001***	$p = .79$	$p = .06$
$B-B$	$F = 4.36$	$t = 1.38$	$t = 3.70$	$t = 0.73$	$t = 3.23$	$t = -0.69$	$t = 2.57$
	p<.01*	$p = .68$	**p<.01***	$p = .98$	**p<.05***	$p = .98$	$p = .08$
$C-A$	$F = 3.28$	$t = 2.43$	$t = 1.99$	$t = 2.34$	$t = 1.96$	$t = 0.01$	$t = 0.14$
	p<.05*	$p = .10$	$p = .28$	$p = .14$	$p = .30$	$p = .99$	$p = .99$
$C-B$	$F = 6.08$	$t = 2.75$	$t = 3.19$	$t = 3.31$	$t = 3.70$	$t = 0.42$	$t = 0.75$
	p<.001*	**p<.05***	**p<.05***	**p<.05***	**p<.01***	$p = .99$	$p = .97$
$C-D$	$F = 10.13$	$t = 4.01$	$t = 4.02$	$t = 3.81$	$t = 3.84$	$t = -0.26$	$t = 0.77$
	p<.001*	**p<.01***	**p<.01***	**p<.01***	**p<.01***	$p = .99$	$p = .97$
$D-B$	$F = 3.19$	$t = 2.33$	$t = 2.18$	$t = 2.07$	$t = 1.92$	$t = -0.49$	$t = 0.49$
	p<.05*	$p = .13$	$p = .20$	$p = .24$	$p = .32$	$p = .99$	$p = .99$
$D-D$	$F = 2.92$	$t = 1.89$	$t = 2.46$	$t = 1.41$	$t = 2.09$	$t = -0.37$	$t = 1.03$
	p<.05*	$p = .33$	$p = .11$	$p = .66$	$p = .23$	$p = .99$	$p = .89$
$C-C^+$	$F = 7.18$	$t = 3.89$	$t = 3.89$	$t = 0.83$	$t = 1.43$	$t = -2.22$	$t = 0.93$
	p<.001*	**p<.01***	**p<.01***	$p = .96$	$p = .64$	$p = .18$	$p = .93$

Table 4

Categorization phase: Pairwise key effects predicted by gaze-informed models. *Obs.=observed, R=recognition, C=categorization, Sel=selective, Dist=distributed. Statistical output from a 2 (item type: high-match vs. conflict) by 4 (group) ANOVA and post hoc pairwise tests. Matching analyses were performed on observed and model-generated response data. Bold text indicates significant effects among comparisons of model-generated responses that are consistent with the observed key effects.*

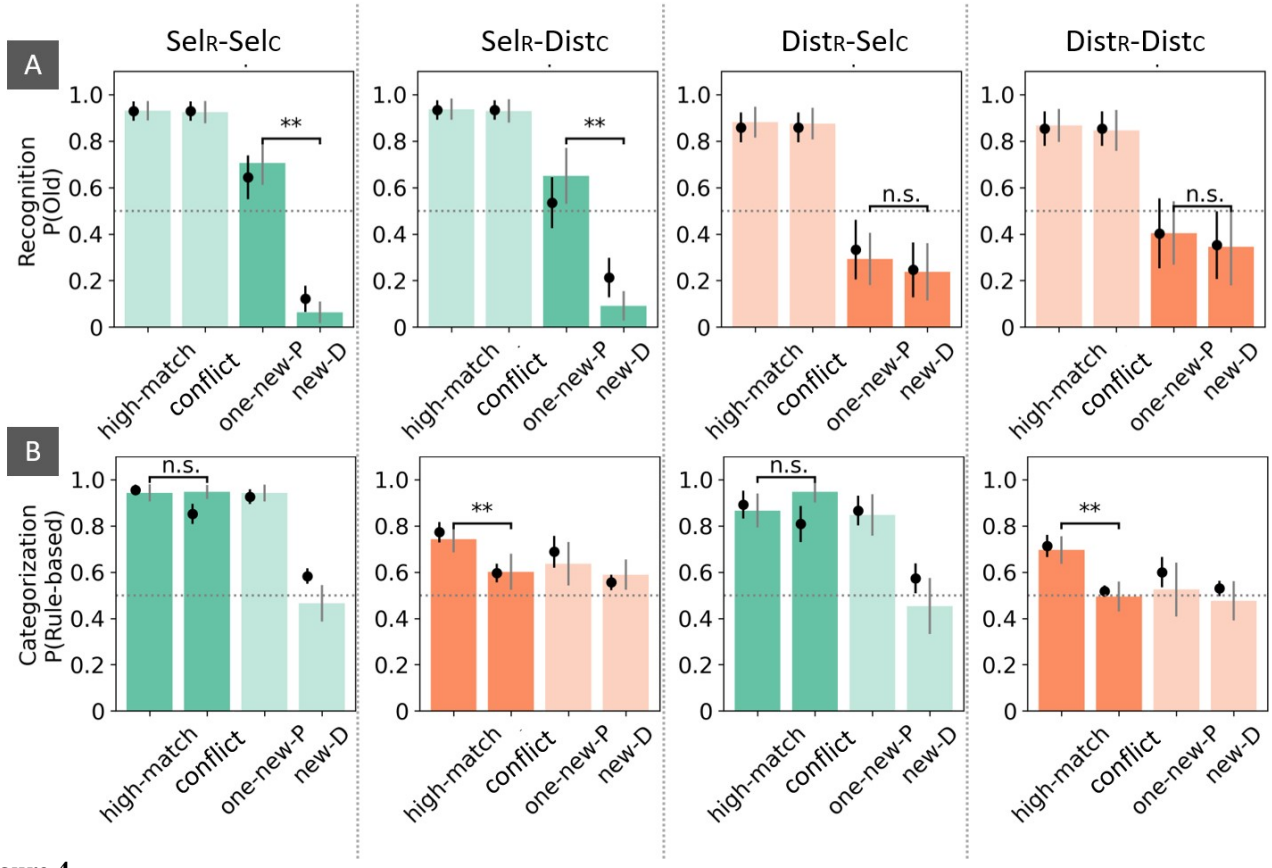

Figure 4

Gaze-predicted behavioral correlates of selective and distributed attention. *D=deterministic, P=probabilistic, P(X)=proportion of X, **=p<0.001, n.s.=not significant. Green bars represent patterns of behavior consistent with a selective strategy of attention, and orange bars correspond to distributed attention. Bold bars and significance markers denote key effects in the observed behavior. (A) Bars show mean probabilities of making an "old" response to each item type during the recognition test phase. Points show aggregate simulations using best-fitting parameters from Model C-B. (B) Bars for high-match, conflict, and one-new-P items reflect mean probabilities of responding consistently with the D feature. Bars for new-D reflect probabilities of responding consistently with the majority of P features. Points show aggregate simulations using best-fitting parameters from Model C-B.*

responses than Groups $Dist_R - Sel_C$ and $Dist_R - Dist_C$ (distributed attention during recognition) to one-new-P compared to new-D items.

2. Groups $Sel_R - Sel_C$ and $Dist_R - Sel_C$ (selective attention during categorization) made more D-consistent responses than Groups $Sel_R - Dist_C$ and $Dist_R - Dist_C$ (distributed attention during categorization) to conflict compared to high-match items.

We first identified best-fitting parameters and gaze transformation values for each participant and model, using procedures provided in Appendix C. We then used the models to generate simulated trial-level response probabilities for each participant, using their best-fitting parameters, observed gaze data, and the sequence of stimuli that the relevant participant experienced during the task. We then determined average response proportions within item type and test phase for each participant and model. As such, model predictions and observed data could be subjected to identical statistical analyses.

Model evaluation

We evaluated each model by its ability to predict the interaction effects that were relevant to each test phase (Figure 2). For the recognition test, we calculated d' for D and P features and submitted the values to a 2 (feature type: D, P) by 4 (group) mixed ANOVA with feature type as a within-subjects factor and group as a between-subjects factor. For the categorization test, we calculated proportions of D-consistent responses and submitted the values to an analogous 2 (item

type: high-match, conflict) by 4 (group) mixed ANOVA. Out of 21 candidate models, simulations from 7 models replicated both critical interaction effects. F statistics (degrees of freedom: 3, 106) and p values are reported in Tables 3 and 4 as they pertain to recognition and categorization, respectively. Predictions from models that predicted the appropriate interaction effects in both phases (models $B-B$, $C-A$, $C-B$, $C-D$, $D-B$, $D-D$, and $C-C^{+}$) were submitted to additional post hoc evaluation with pairwise independent-samples t-tests.

To summarize the observed effects of interest during recognition, Groups Sel_R-Sel_C and Sel_R-Dist_C were less sensitive to novel P features than Groups $Dist_R-Sel_C$ and $Dist_R-Dist_C$ ($ps < 0.001^{*}$). Groups who used a common attention strategy did not differ in sensitivity between one another (Sel_R-Sel_C vs. Sel_R-Dist_C: $p=0.81$; $Dist_R-Sel_C$ vs. $Dist_R-Dist_C$: $p=0.99$). As shown in Table 3, 4 gaze-informed models (indicated by bold text) appropriately predicted all 4 pairwise effects of interest: models $C-A$, $C-B$, $C-D$, and $D-D$.

During the categorization test, analyses of observed data showed that Groups Sel_R-Sel_C and $Dist_R-Sel_C$ responded more consistently with the D feature during high-match and conflict items than Groups Sel_R-Dist_C and $Dist_R-Dist_C$ ($ps < 0.001^{*}$). Groups who used a common attention strategy did not significantly differ in D-consistent responding between one another (Sel_R-Sel_C vs. $Dist_R-Sel_C$: $p=0.06^{\cdot}$; Sel_R-Dist_C vs. $Dist_R-Dist_C$: $p=0.79$). As shown in Table 4, only 2 gaze-informed models (indicated by bold text) appropriately predicted all 4 pairwise effects of interest: models $C-B$ and $C-D$.

Table 5 shows total AIC values for the selection of 7 models that effectively simulated key interaction effects for recognition and categorization. Model $C-B$ provided the best fits to data from Groups Sel_R-Sel_C and Sel_R-Dist_C, and Model $B-B$ provided the best fits to data from Groups $Dist_R-Sel_C$ and $Dist_R-Dist_C$. However, Model $B-B$ proved to be ineffective for predicting behavioral differences between selective and distributed attention strategy groups during both recognition and categorization (Tables 3 and 4).

Although Models $C-B$ and $C-D$ both predicted all pairwise behavioral effects of interest, Model $C-B$ presumably attained more favorable AIC values on the basis of parsimony (one fewer free parameter). Considering all results together, we selected Model $C-B$ as the most effective model overall out of 21 candidates. Aggregate predictions using each subject's best-fitting parameters from Model $C-B$ are shown in Figure 4 (points).

Examining conventions

Due to their theoretical significance (Nosofsky, 1986), statistics for evaluating the predictions of the perfect encoding models ($1-A$, $1-B$, $1-C$, and $1-D$) are provided in Table 6. None of these models were able to predict key interaction effects during categorization, however, and were therefore not subjected to additional post hoc evaluation. Considering total AIC, all four perfect encoding models performed worse than every model listed in Table 5, which included allowances for memory precision. From these results, we note that simply accounting for sparsity in the feature encoding has profound effects on behavioral predictions.

Within our direct input approach, we made the choice to calculate attention at the level of each probe-exemplar comparison as a product of memory precision and decision weights (Equation 2). Because this specification contradicts the standard GCM constraint where

$$\sum_{k}\alpha_k = 1,$$

we included a gaze-informed model $C-C^{+}$ that satisfies the constraint on total attention by-trial. To reiterate, Model $C-C^{+}$ calculates attention as a mixture of softmax-transformed η and ζ, similar to how freely-estimated perceptual and decisional components of attention are combined in EGCM (Lamberts, 1995). Although Model $C-C^{+}$ predicted the item type by group interaction effects relevant to both recognition and categorization phases, it failed to predict several key pairwise effects for distinguishing selective and distributed attention strategies (Tables 3 and 4) and was unremarkable compared to the other candidate models in terms of AIC (Table 5). Consistent with the findings of previous work, these results suggest that attention allocation is highly flexible and variable within- and between-trials, and may not be adequately summarized with hardline summation constraints in place (Galdo et al., 2022; Weichart et al., 2022).

Eye-tracking

Mean proportions of raw dwell times to the D feature during training (in sets of 10 trials), recognition test, and categorization test are provided in Table 7. We identified a significant group difference in proportions of gaze allocated to the D feature during the latter trials of training ($F(3,106)=5.03, p<0.01$), with Group Sel_R-Sel_C showing longer relative dwell times to the D feature compared to Group $Dist_R-Dist_C$. No other group-wise comparisons of dwell time during training reached statistical significance.

Figure 5 shows aggregate softmax-transformed dwell times during training, using best-fitting θ_{train} values from our winning model, $C-B$. These transformed gaze maps serve as a way of visualizing memory precision of the features presented during training, and provide uniquely nuanced information that is constrained by both gaze and choices during subsequent test. Group Sel_R-Sel_C shows high memory precision for the D dimension in particular, while Groups Sel_R-Dist_C, $Dist_R-Sel_C$, and $Dist_R-Dist_C$ show more evenly-distributed precision among the P dimensions. Group

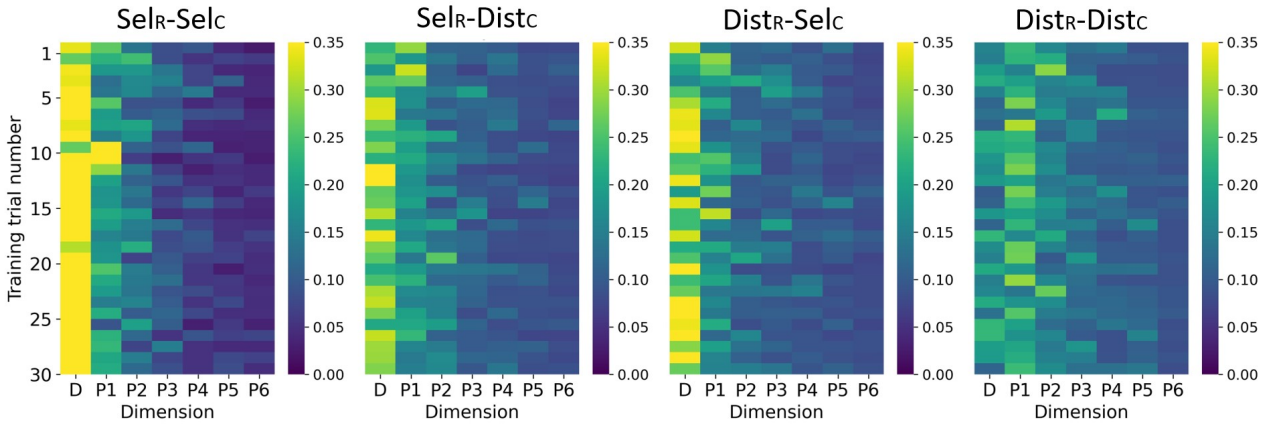

Figure 5

Gaze-based memory precision for training features. *D=deterministic, P=probabilistic. Heatmaps show aggregate memory precision across subjects. X-ticks indicate stimulus dimensions, where P dimensions were rank-ordered within-subject according to gaze preference. Y-ticks show trial numbers. Subject-wise memory precision maps were calculated by subjecting raw dwell time data to best-fitting model-based transformations.*

Model	N free par.	$Sel_R - Sel_C$	$Sel_R - Dist_C$	$Dist_R - Sel_C$	$Dist_R - Dist_C$	Total
$B-B$	6	2131	2012	**1162**	**1326**	6631
$C-A$	6	2009	2036	1196	1341	6582
$C-B$	6	**1995**	**1993**	1170	1335	**6493**
$C-D$	7	2056	2059	1210	1355	6680
$D-B$	7	2157	2075	1223	1357	6812
$D-D$	8	2186	2091	1233	1382	6892
$C-C^+$	7	2240	2070	1215	1345	6870

Table 5

AIC comparison: Selected gaze-informed models *Values are total AICs across subjects in the indicated groups. Bold text indicates the lowest (i.e. preferred) AIC value within each column.*

$Dist_R - Dist_C$ appeared to not preferentially encode D features at all, and instead slightly favored one of the P dimensions. Extended analyses of model-transformed gaze during recognition and categorization test are provided in Appendix D.

General Discussion

This study explored the hypothesis that memory precision for features encountered during learning influences decisions in subsequent test contexts. To examine this relationship, we leveraged eye-tracking data as dissociable components of attention in an exemplar-similarity model. We found compelling evidence that the availability of information stored in memory during training plays a pivotal role in accurately predicting observed choices during test. The following sections offer interpretations of these findings within the established literature on selective attention as an indicator of the observer's beliefs and intentions within the task environment.

Intentions and consequences

Our study stands apart from previous investigations on the consequences of selective attention (Blanco et al., 2023; Plebanek & Sloutsky, 2017; Best et al., 2013) due to its novel usage of the exemplar-similarity framework. This framework is extremely influential in cognitive psychology, and yet has historically been woefully non-committal in its treatment of memory precision in predictions of choice. We used gaze correlates of attention as direct inputs to an exemplar-similarity model for predicting recognition and categorization decisions. In one set of models, training features were presumed to be perfectly encoded, and gaze during test was the sole determinant of choice (Table 6). In another set of models, gaze was used to constrain estimates of memory precision for features encountered during training, as well

Model	Recog.	Cat.	Total AIC
$C-B$	F=17.36	F=6.08	6493
	p<.001*	**p<0.001***	
$1-A$	F=3.31	F=1.68	7205
	p<.05*	p=.17	
$1-B$	F=3.88	F=2.12	6960
	p<.05*	p=.10	
$1-C$	F=3.71	F=0.20	6925
	p<.05*	p=.90	
$1-D$	F=4.89	F=0.22	7086
	p<.01*	p=.88	

Table 6

Model comparison: Perfect encoding models. *Recog.=recognition, Cat.=categorization. F statistics and p values evaluate key interaction effects of item type and group in a 2 by 4 mixed ANOVA.*

as estimates of decision weights among features of the test stimuli (Table 5). These two sets of models represented competing hypotheses concerning the relevant determinants of attention: the former representing attention as decision weights (e.g. GCM), and the latter representing attention as decision weights constrained by memory.

The results of model evaluation and comparison favored the latter account: models that included gaze correlates of memory precision outperformed those that did not, as determined by AIC. In addition, a subset of models that accounted for memory could predict nuanced behavioral correlates of selective and distributed attention that were defined in an independent investigation (Deng & Sloutsky, 2016). Models that assumed perfect encoding of information presented during training, by contrast, failed to even predict basic interaction effects between groups defined by contrasting attention strategies.

When considering these results, it is important to note that the standard implementation of GCM with freely-estimated attention parameters was able to predict key behaviors perfectly well (Appendix A). The successes of the memory-informed models therefore do not denote an incremental improvement in model fits. Instead, these successes redress a theoretical opacity in prior modeling frameworks. Instead of attributing behavior to a nebulous construct of "attention," we find that accounting for the effects of memory provided a significantly better approximation of freely-estimated attention parameters than if we account for decision weights at test alone. Our results support the hypothesis that behavioral correlates of attention reflect the strategic weighting of the information that was *encoded* by the individual participant, not of all information that was *presented*.

Implications for human learning

It is not our intention to admonish early presentations of the exemplar-similarity framework (Medin & Shaffer, 1978; Nosofsky, 1986) for making simplifying assumptions. The assumption of perfect encoding is clearly computationally necessary for constraining model estimates of attention when behavior alone is the output. It has not, to our knowledge, been asserted by users of the framework as a genuine theory that humans perfectly and equally store all information that exists in the learning environment. Our findings should instead be interpreted as a cause for theoretical reevaluation of attention as it is specified in contemporary accounts of human learning.

One influential class of *adaptive attention models* has built upon the GCM framework to explore how attention updates from trial-to-trial to support learning (e.g. Kruschke, 1992; Love et al., 2004; Galdo et al., 2022). Within these models, the observer uses trial-level category feedback to update their distribution of attention in a way that is intended to reduce the probability of future errors. Through iterative attention optimization, these models have been shown to predict trajectories of attention and accuracy that mirror the trajectories of dimension-level gaze preferences observed by Rehder & Hoffman (2005a) (Galdo et al., 2022; Nosofsky et al., 1994; Kurz et al., 2013). Importantly, attention is calculated to optimally weight *all* information that was presented on prior trials. This policy, however, may not accurately reflect the information that is actually available to participants–unless of course we can reasonably conclude that humans store features equally well whether they look at them or not.

While most model instantiations of attention interpret failures of accuracy as an inappropriate weighting of irrelevant information, our findings suggest that failures to behave optimally can also be attributed to sparse encoding. Recent findings from Wan & Sloutsky (2023) provided important insight into this distinction using a version of the same experiment presented here (Deng & Sloutsky, 2016). All stimulus features were occluded at the onset of each trial, and participants revealed the desired feature information by tapping occlusion bubbles on a touch-screen. By contrast to gaze measures as an index of attention, Wan & Sloutsky's innovative approach offers the advantage of providing insight into which features were plausibly encoded into memory during training and, importantly, which features could not have possibly entered the representation.

The results showed that adult participants tended to selectively reveal the feature in the most category-diagnostic (D) dimension, and behaviors at test denoted a strategy of selective attention (e.g. Figure 2). Interestingly, participants revealed significantly *more* features when they encountered new-D items at test compared to the other item types. This behavior is potentially indicative of an attempt to optimally redistribute attention upon encountering unusable in-

Group	Training 1-10	Training 11-20	Training 21-30	Recog. Test	Cat. Test
$Sel_R - Sel_C$	0.32(0.14)	0.37(0.17)	0.38(0.17)	0.26(0.12)	0.43(0.13)
$Sel_R - Dist_C$	0.25(0.14)	0.30(0.18)	0.30(0.16)	0.25(0.14)	0.30(0.17)
$Dist_R - Sel_C$	0.27(0.13)	0.26(0.15)	0.31(0.19)	0.18(0.07)	0.35(0.13)
$Dist_R - Dist_C$	0.17(0.11)	0.22(0.17)	0.20(0.14)	0.15(0.05)	0.23(0.12)

Table 7

Observed gaze preference for deterministic (D) dimension. *R=recognition, C=categorization, Sel=selective, Dist=distributed. Table entries show proportions of fixations to the D dimension in the format: [mean]([standard deviation]).*

formation in the D position. Indeed, other eye-tracking work demonstrated that by increasing the uncertainty of choice via requiring reliance on probabilistic cues, participants were provoked to sample more sources of information before making a response (Easdale et al., 2019; Beesley et al., 2015).

We consider these findings to be consistent with our own: regardless of whether participants use strategies that can be attributed to past or present optimality, the information stored during training is immutable in its impact on future decisions. A participant who manages to encode all information presented during training (à la GCM) would presumably be able to weight information at test in a way that best serves their goals, whether characterized by accuracy, efficiency, novelty preference, information gain, or otherwise (Matsuka & Corter, 2008). In the more likely case of imperfect memory storage, we posit that information that is not encoded cannot be retrospectively reclaimed as needed.

Although most adaptive attention models propose pure accuracy optimization as a mechanism for updating, the *adaptive attention representation model* was developed to explore other secondary goals that humans could plausibly pursue during learning (AARM; Galdo et al., 2022). In one study, the authors fit model-predicted quantities of choice probability and attention directly to simultaneous streams of behavioral and gaze data. With the constraint provided by gaze data, the authors were equipped to evaluate contrasting theories about the goals that contribute to attention allocation. Results across five experiments supported the conclusion that the pursuit of accuracy goals alone was insufficient for explaining observed patterns of attention during learning. Instead, the authors advocated for efficiency considerations as well, instantiated as active suppression of redundant information in memory.

The findings from the current study, however, provide an alternative explanation. It may be the case that humans indeed proceed with the *intention* of being as accurate as possible, but are constrained by the *consequences* of partial encoding. If one accounts for partial encoding as a natural consequence of limited information sampling, simple rules for updating attention in an effort to be accurate may prevail (Kruschke, 2001, 1992). We suggest that a complete theory of attention optimization will need to consider information sampling and decision weights as dissociable contributors to common goals.

Limitations and future directions

Our winning model $C - B$ is characterized by a softmax function to convert gaze during training into estimates of memory precision, and a piecewise linear function to convert gaze during test into estimates of decision weights. This difference in transformations implies that when a participant encounters a stimulus, the way they weight feature information when making an object discrimination judgement may be incongruous to the contents of the memory trace that they store.

This finding is tentatively consistent with the concept of evidence accumulation dynamics. Evidence accumulation models posit that decisions are made by considering multiple sources of information, and allowing processes of competition and inhibition to ultimately favor one choice option over another (Ratcliff, 1978; Usher, 2001). It may be the case that information is stored in proportion to low-level perceptual processes (i.e. such that all information that is fixated to some extent is plausibly stored), but additional dynamics that occur during the decision may result in high fixations to dimensions with conflicting information even though the choice only reflects the “winning" source (Krajbich & Rangel, 2011). Future work will need to investigate the plausibility of this conjecture, and determine whether the evidence accumulation dynamics that impact choice additionally impact how the memory trace corresponding to the stimulus is formed.

The current study took a foreseeably controversial approach by using data from young children to investigate a general theory about the impact of memory precision on attention strategies. We argue, however, that the use of children in our current investigation is more of a strength than a weakness. One can reason that the typical child-like policy of broad information sampling during training (Blanco & Sloutsky, 2020; Blanco et al., 2023) is consistent with the GCM description of an unabridged memory store that is manipulated by attention at test. It is therefore notable that a group who has an even *better* chance of favoring the

conventional account than adults still favored models with encoding biases as a determinant of decision making. We nevertheless acknowledge that vast developmental changes to the attention, memory, and decision-making faculties of interest occur after age 5, and it is therefore essential to validate our framework with data from adults as well. Although we believe it is useful to evaluate the relationship between gaze, memory, and decision processes within a population who naturally exhibited variability in sampling and decision strategies, future work with adults will focus on strategic manipulations involving uncertainty (Easdale et al., 2019) and feedback reliability (Little & Lewandowsky, 2009).

Additional limitations to our study relate to the selection of training stimuli and the simplicity of our model specifications for memory precision. Given that features were repeated multiple times during training, we cannot draw strong conclusions about which specific features were best-represented in memory. Although we were able to effectively predict behavior using a simple transformation of gaze data to represent memory precision, we assume that additional forces of lag-based decay, context effects, and repetition effects are at play as well (Kahana, 2012). Future work will, for example, utilize paradigms that manipulate the sequence of items presented during learning (Carvalho & Goldstone, 2017; Kim & Rehder, 2011) in the hopes of providing more precise measurements for relating gaze to memory for individual features.

Conclusions

There are two main takeaways from the current work, one methodological and one theoretical. First, we provided a novel model-based method for leveraging eye-tracking data to observe the contents of memory, which underlie the malleable object representations that are used to make decisions. Second, we provided model comparison results that support the theory that engaging selective attention during learning incurs costs to breadth of information storage in memory, which in turn imposes unintended limitations on future decision-making.

We assert that our findings using a data-driven approach that considers dissociable components of attention have important implications for ongoing theoretical developments in human learning. The field continues to push the boundaries of the exemplar-similarity framework for unraveling the intricacies of learning, most often instantiating dynamic mechanisms of attention as the locus of innovation (e.g. Kruschke, 1992; Love et al., 2004; Galdo et al., 2022; Carvalho & Goldstone, 2022). Without further scrutiny of attention's core principles, however, venturing into new frontiers becomes an exercise in futility. The current article therefore takes an important step toward understanding the component operations of attention that are essential to contemporary theories of learning, yet are rarely explored.

Declaration of Conflicting Interests

The authors declare no conflict of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by the National Institutes of Health (grant number RO1HD078545; awarded to VMS) and the National Science Foundation (grant number 1847603; awarded to BMT).

Author Contributions

ERW: conceptualization, methodology, formal analysis, visualization, writing-original draft; **LU:** writing-review and editing; **NK:** writing-review and editing; **VMS:** data curation; **BMT:** conceptualization, writing-review and editing.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. (DOI: https://doi.org/10.1109/TAC.1974.1100705)

Ashby, F., & Lee, W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*(2), 150–172. (DOI: https://doi.org/10.1037/0096-3445.120.2.150)

Ashby, F., & Maddox, W. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*(3), 372–400. (DOI: https://doi.org/10.1006/jmps.1993.1023)

Beesley, T., & Le Pelley, M. (2011). The influence of blocking on overt attention and associability in human learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(1), 114. (DOI: https://doi.org/10.1037/a0019526)

Beesley, T., Nguyen, K., pearson, D., & Le Pelley, M. (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. *Quarterly Journal of Experimental Psychology*, *68*(11), 2175-2199. (DOI: https://doi.org/10.1080/17470218.2015.1009919)

Best, C., Yim, H., & Sloutsky, V. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119. (DOI: https://doi.org/10.1016/j.jecp.2013.05.002)

Blair, M., Watson, M., Walsche, R., & Maj, F. (2009). 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*, *35*(5), 1196–1206. (DOI: https://doi.org/10.1037/a0016272)

Blanco, N., & Sloutsky, V. (2019). Adaptive flexibility in category learning? young children exhibit smaller costs of selective attention than adults. *Developmental Psychology*, *55*(10), 2060. (DOI: https://doi.org/10.1037/dev0000777)

Blanco, N., & Sloutsky, V. (2020). Attentional mechanisms drive systematic exploration in young children. *Cognition*, *101*, 104327. (DOI: https://doi.org/10.1016/j.cognition.2020.104327)

Blanco, N., Turner, B., & Sloutsky, V. (2023). The benefits of immature cognitive control: How distributed attention guards against learning traps. *Journal of Experimental Child Psychology*, *226*, 105548. (DOI: https://doi.org/10.1016/j.jecp.2022.105548)

Brest, J., Greiner, S., Boskovic, B., Merrik, M., & Zumer, V. (2006). Self-adapting control parameters in differential evolution. *IEEE Transactions on Evolutionary Computation*, *10*(6), 646–657. (DOI: https://doi.org/10.1109/TEVC.2006.872133)

Carvalho, P., & Goldstone, R. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699–1719. (DOI: https://doi.org/10.1037/xlm0000406)

Carvalho, P., & Goldstone, R. (2022). A computational model of context-dependent encodings during category learning. *Cognitive Science*, *46*(4), e13128. (DOI: https://doi.org/10.1111/cogs.13128)

Darby, K., & Sederberg, P. (2022). Transparency, replicability, and discovery in cognitive aging research: A computational modeling approach. *Psychology and Aging*, *37*(1), 10. (DOI: https://doi.org/10.1037/pag0000665)

Deng, W., & Sloutsky, V. (2012). Carrot eaters or moving heads: Inductive inference is better supported by salient features than by category labels. *Psychological Science*, *23*(2), 178-186. (DOI: https://doi.org/10.1177/0956797611429133)

Deng, W., & Sloutsky, V. (2015a). The development of categorization: Effects of classification and inference training on category representation. *Developmental Psychology*, *51*(3), 392–405. (DOI: https://doi.org/10.1037/a0038749)

Deng, W., & Sloutsky, V. (2015b). Linguistic labels, visual features, and attention in infant category learning. *Journal of Experimental Child Psychology*, *134*, 62–77. (DOI: https://doi.org/10.1016/j.jecp.2015.01.012)

Deng, W., & Sloutsky, V. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62. (DOI: https://doi.org/10.1016/j.cogpsych.2016.09.002)

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*(4), 501–517. (DOI: https://doi.org/10.1037/0096-3445.113.4.501)

Easdale, L., Le Pelley, M., & Beesley, T. (2019). The onset of uncertainty facilitates the learning of new associations by increasing attention to cues. *Quarterly Journal of Experimental Psychology*, *72*(2), 193-208. (DOI: https://doi.org/10.1080/17470218.2017.13632)

Estes, W. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. (DOI: https://doi.org/10.1016/0010-0285(86)90008-3)

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 6. (DOI: https://doi.org/10.1167/8.2.6)

Galdo, M., Weichart, E., Sloutsky, V., & Turner, B. (2022). The quest for simplicity in human learning: Identifyig the constraints on attention. *Cognitive Psychology*, *138*(5), 101508. (DOI: https://doi.org/10.1016/j.cogpsych.2022.101508)

Greene, M., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137-176. (DOI: https://doi.org/10.1016/j.cogpsych.2008.06.001)

Griffiths, O., & Mitchell, C. (2008). Selective attention in human associative learning and recognition memory. *Journal of Experimental Psychology: General*, *137*(4), 626–648. (DOI: https://doi.org/10.1037/a0013685)

Kahana, M. (2012). *Foundations of human memory*. Oxford University Press.

Kamin, L. (1968). Predictability, surprise, attention, and conditioning. In B. Campbell & R. Church (Eds.), *Punishment and aversive behavior.* New York: Appleton-Century-Crofts.

Kim, S., & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition*, *39*(4), 649–665. (DOI: https://doi.org/10.3758/s13421-010-0050-3)

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857. (DOI: https://doi.org/10.1073/pnas.1101328108)

Kruschke, J. (1992). ALCOVE:An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. (DOI: https://doi.org/10.1037/0033-295X.99.1.22)

Kruschke, J. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*(6), 812–863. (DOI: https://doi.org/10.1006/jmps.2000.1354)

Kruschke, J., & Blair, N. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636–645. (DOI: https://doi.org/10.3758/BF03213001)

Kruschke, J., Kappenman, E., & Hetrick, W. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(5), 830–845. (DOI: https://doi.org/10.1037/0278-7393.31.5.830)

Kurz, K., Levering, K., Stanton, R., Romero, J., & Morris, S. (2013). Human learning of elemental category structures: Revising the classic result of Shepard, Hovland, and Jenkins (1961). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 552–369. (DOI: https://doi.org/10.1037/a0029178)

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*(2), 161–180. (DOI: https://doi.org/10.1037/0096-3445.124.2.161)

Le Pelley, M., Beesley, T., & Suret, M. (2007). Blocking of human causal learning involves learned changes in stimulus processing. *Quarterly Journal of Experimental Psychology*, *60*(11), 1468–1476. (DOI: https://doi.org/10.1080/17470210701564)

Little, D., & Lewandowsky, S. (2009). Better learning with more error: Probabilistic feedback increases sensitivity to correlated cues in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1041. (DOI: https://doi.org/10.1037/a0015902)

Liu, Z., Song, X., & Seger, C. (2015). An eye-tracking study of multiple feature value category structure learning: The role of unique features. *Plos one*, *10*(8), e0135729. (DOI: https://doi.org/10.1371/journal.pone.0135729)

Loftus, G. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*(4), 525–551. (DOI: https://doi.org/10.1016/0010-0285(72)90021-7)

Love, B., Medin, D., & Gureckis, T. (2004). SUSTAIN:A network model of category learning. *Psychological Review*, *111*(2), 309. (DOI: https://doi.org/10.1037/0033-295X.111.2.309)

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298. (DOI: https://doi.org/10.1037/h0076778)

Maddox, W., & Ashby, F. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 795–817. (DOI: https://doi.org/10.1037/0096-1523.22.4.795)

Matsuka, T., & Corter, J. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology*, *61*(7), 1067–1097. (DOI: https://doi.org/10.1080/17470210701438194)

McKinley, S., & Nosofsky, R. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 294–317. (DOI: https://doi.org/10.1037/0096-1523.22.2.294)

Medin, D., & Shaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207. (DOI: https://doi.org/10.1037/0033-295X.85.3.207)

Medin, D., & Smith, E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(4), 241. (DOI: https://doi.org/10.1037/0278-7393.7.4.241)

Meier, K., & Blair, M. (2013). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition*, *126*(2), 319–325. (DOI: https://doi.org/10.1016/j.cognition.2012.09.014)

Murphy, G. (2002). *The big book of concepts*. Cambridge, Ma: MIT Press.

Myung, J., & Pitt, M. (2004). Model comparison methods. *Methods in Enzymology*, *383*, 351–366. (DOI: https://doi.org/10.1016/S0076-6879(08)03811-1)

Navarro, D. (2007). On the interaction between exemplar-based concepts and a response scaling process. *Journal of Mathematical Psychology*, *51*(2), 85–98. (DOI: https://doi.org/10.1016/j.jmp.2006.11.003)

Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39. (DOI: https://doi.org/10.1037/0096-3445.115.1.39)

Nosofsky, R. (1991). Tests of an exemplar model for relating classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. (DOI: https://doi.org/10.1037/0096-1523.17.1.3)

Nosofsky, R., Gluck, M., Palmeri, T., McKinley, S., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369. (DOI: https://doi.org/10.3758/BF03200862)

Palmeri, T., Love, B., & Turner, B. (2017). Model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 59–64. (DOI: https://doi.org/10.1016/j.jmp.2016.10.010)

Pearce, J., & Hall, G. (1980). A model for pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*, 532–552. (DOI: https://doi.org/10.1037/0033-295X.87.6.532)

Peterson, M., Kramer, A., Wang, R., Irwin, D., & McCarley, J. (2001). Visual search has memory. *Psychological Science*, *12*(4), 187–292. (DOI: https://doi.org/10.1111/1467-9280.00353)

Plebanek, D., & Sloutsky, V. (2017). Costs of selective attention: When children notice what adults miss. *Psychological Science*, *28*(6), 723–732. (DOI: https://doi.org/10.1177/0956797617693005)

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59. (DOI: https://doi.org/10.1037/0033-295X.85.2.59)

Rehder, B., & Hoffman, A. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1-41. (DOI: https://doi.org/10.1016/j.cogpsych.2004.11.001)

Rehder, B., & Hoffman, A. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(5), 811–829. (DOI: https://doi.org/10.1037/0278-7393.31.5.811)

Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning:variations in the effectieness of reinforcement and non-reinforcement. In A. Black & W. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Rich, A., & Gureckis, T. (2018). The limits of learning: Exploration, generalization, and the development of learning traps. *Journal of Experimental Psychology: General*, *147*(11), 1553–1570. (DOI: https://doi.org/10.1037/xge0000466)

Shepard, R., Hovland, C., & Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42. (DOI: https://doi.org/10.1037/h0093825)

Smith, L., & Kemler, D. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, *24*(2), 279–298. (DOI: https://doi.org/10.1016/0022-0965(77)90007-8)

Snodgrass, J., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*(1), 34–50. (DOI: https://doi.org/10.1037/0096-3445.117.1.34)

Storn, R., & Price, K. (1997). Differential evolution: A simple and efficient hueristic for global optimization over continuous spaces. *Journal of Global Optimization*, *11*(4), 341. (DOI: https://doi.org/10.1023/A:1008202821328)

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. (DOI: https://doi.org/10.1037/h0027242)

Turner, B., Forstmann, B., Love, B., Palmeri, T., & van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. (DOI: https://doi.org/10.1016/j.jmp.2016.01.001)

Turner, B., Forstmann, B., & Steyvers, M. (2019). *Joint models of neural and behavioral data* (A. Criss, Ed.). Springer International Publishing. (DOI: https://doi.org/10.1007/978-3-030-03688-1)

Turner, B., Palestro, J., Miletic, S., & Forstmann, B. (2019). Advances in techniques for imposing reciprocity

in brain-behavior relations. *Neuroscience & Biobehavioral Reviews*, *102*, 327–336. (DOI: https://doi.org/10.1016/j.neubiorev.2019.04.018)

Turner, B., Schley, D., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review*, *125*(3), 329. (DOI: https://doi.org/10.1037/rev0000089)

Usher, J., M McClelland. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550. (DOI: https://doi.org/10.1037/0033-295X.108.3.550)

Wan, Q., & Sloutsky, V. (2023). Driven by information: Children's exploration shapes their distributed attention in category learning. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the annual meeting of the cognitive science society* (Vol. 45).

Weichart, E., Darby, K., Fenton, A., Jacques, B., Kirkpatrick, R., Turner, B., & Sederberg, P. (2021). Quantifying mechanisms of cognition with an experiment and modeling ecosystem. *Behavior Research Methods*, *53*, 1833–1856. (DOI: https://doi.org/10.3758/s13428-020-01534-w)

Weichart, E., Galdo, M., Sloutsky, V., & Turner, B. (2022). As within, so without; as above, so below: Common mechanisms can support between- and within-trial category learning dynamics. *Psychological Review*, *129*(5), 1104–1143. (DOI: https://doi.org/10.1037/rev0000381)

Weichart, E., & Sederberg, P. (2021). Individual differences in attention allocation during a two-dimensional inhibitory control task. *Attention, Perception, & Psychophysics*, *83*, 676–684. (DOI: https://doi.org/10.3758/s13414-020-02160-6)

Wiecki, T., Poland, J., & Frank, M. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. *Clinical Psychological Science*, *3*(3), 378–399. (DOI: https://doi.org/10.1177/2167702614565359)

Appendix A
Key effects observed between groups

Strategy groups were initially identified using variants of standard GCM fits to behavior alone, before involving measures of gaze (Table 2). This approach provided an important baseline test for the appropriateness of the exemplar-similarity framework for capturing data in the present context. If the standard model not fit when given the flexibility of freely-estimated attention parameters for each test phase, there would be no point in imposing additional constraints by way of eye-tracking data. As discussed below, however, the exemplar-similarity framework provided an excellent account of the data and key effects (Figure 2).

Mean and 95% confidence intervals of best-fitting α_D parameters for recognition and categorization are shown in Figure A1A. Predictions from GCM using best-fitting parameters provide good fits to data as determined by qualitative assessment. In the sections to follow, we verify that our individual-level, model-based approach was effective for identifying strategy groups that replicate Deng & Sloutsky.

To analyze data from the recognition test phase, we first calculated each participant's sensitivity to new features that occurred in the D and P positions via d' (d-prime). We applied the formula $d' = Z(HitRate) - Z(FalseAlarmRate)$ where Hit Rate refers to the proportion of correct "old" responses to high-match items, and False Alarm Rate refers to the proportion of incorrect "old" responses to new-D and one-new-P items. To address the issue of extreme values, we adjusted the hit rates and false alarm rates using methods described by Snodgrass & Corwin (1988), ensuring that no accuracy values were equal to 0 or 1. We submitted d' values to a 2 (feature type: D, P) by 4 (group) mixed ANOVA with feature type as a within-subjects factor and group as a between-subjects factor. This analysis identified a significant interaction ($F(3,106) = 35.71$, $MSE = 27.90$, $p < 0.001$, $\eta^2 = 0.50$).

We then performed post hoc tests to assess the differences in sensitivity to new D and P features within each group. Sidak's correction was applied to control for multiple comparisons, resulting in adjusted p-values for each test ($\alpha = 0.05$). For Group $Sel_R - Sel_C$, a paired samples t-test revealed higher d' for D ($\mu_D = 4.00$, $\sigma_D = 1.15$) than P features during recognition ($\mu_P = 0.95$, $\sigma_P = 1.16$, $t(42) = 14.49$, $p < 0.001$, $d = 2.62$). A similar effect was found for Group $Sel_R - Dist_C$ ($\mu_D = 3.85$, $\sigma_D = 1.05$, $\mu_P = 1.20$, $\sigma_P = 1.27$, $t(29) = 9.88$, $p < 0.001$, $d = 2.23$). For Groups $Dist_R - Sel_C$ and $Dist_R - Dist_C$, however, participants were equally likely to identify novel D and P features ($Group\ Dist_R - Sel_C$: $\mu_D = 2.69$, $\sigma_D = 1.53$, $\mu_P = 2.33$, $\sigma_P = 1.28$, $t(19) = 1.94$, $p = 0.07$, $d = 0.25$; $Group\ Dist_R - Dist_C$: $\mu_D = 1.94$, $\sigma_D = 1.30$, $\mu_P = 1.71$, $\sigma_P = 1.03$, $t(16) = 1.20$, $p = 0.12$, $d = 0.19$).

To analyze data from the categorization test, we focused on probabilities of D-consistent responses during high-match and conflict items. Data were analyzed with a 2 (item-type: high-match, conflict) by 4 (group) mixed ANOVA with item type as a within-subjects factor and group as a between-subjects factor. After identifying a significant interaction ($F(3,106) = 19.48$, $MSE = 0.22$, $p < 0.001, \eta^2 = 0.36$), we used post hoc paired-samples t-tests to explore effects within each group. As before, we applied Sidak's correction to control for multiple comparisons ($\alpha = 0.05$).

Group $Sel_R - Dist_C$ made significantly more D-

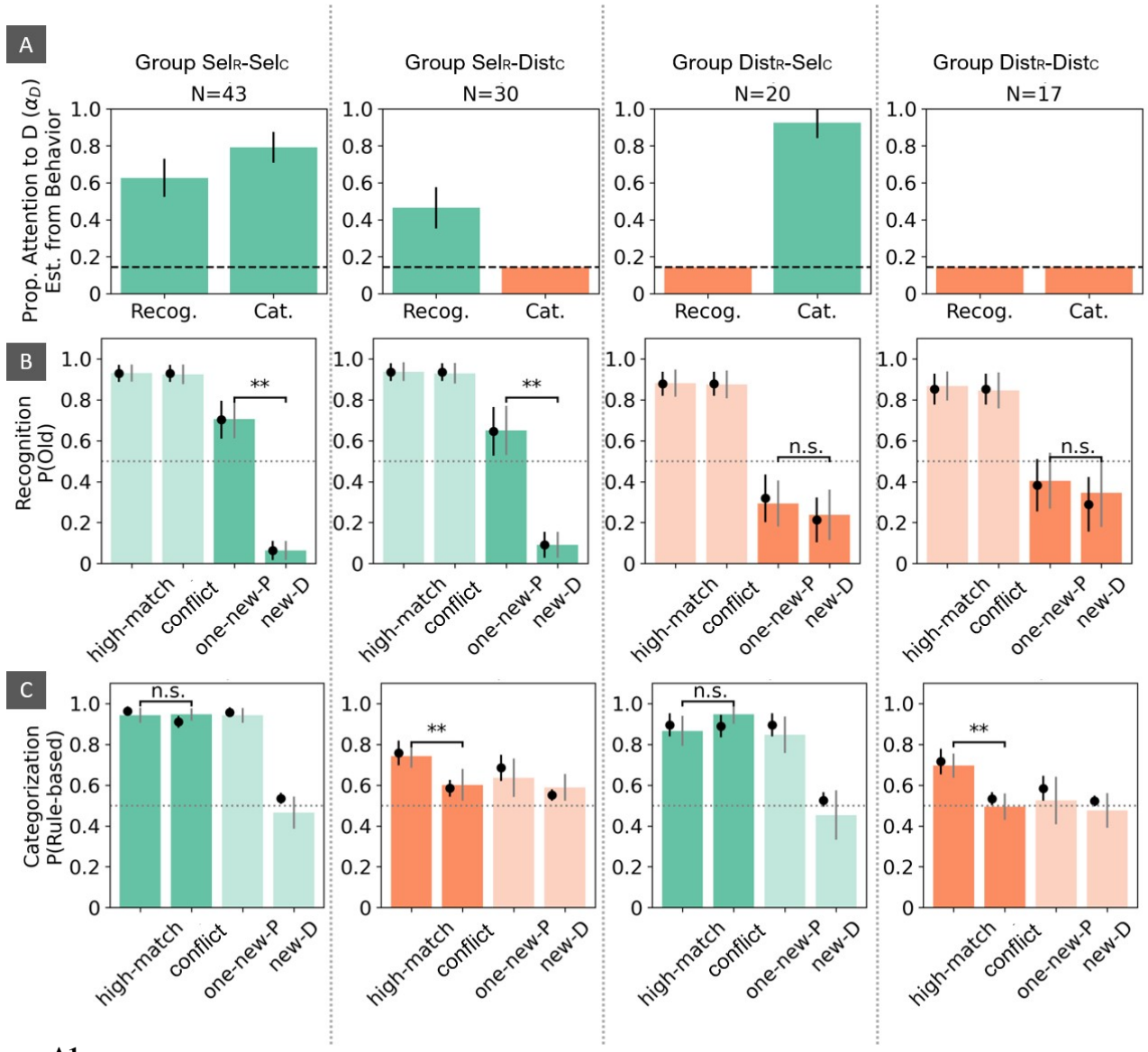

Figure A1

Behavioral correlates of selective and distributed attention; model predictions with freely-estimated α. *D=deterministic, P=probabilistic, P(X)=proportion of X, N=number of subjects, α=attention parameter, Recog.=recognition, Cat.=categorization, **=p<0.001, n.s.=not significant. Green bars represent patterns of behavior consistent with a selective strategy of attention, and orange bars correspond to distributed attention. Bold bars and significance markers denote key effects in the observed behavior. (A) We identified four groups of participants via comparison of Generalized Context Model (GCM) variants with contrasting specifications of attention. Bars show mean and 95% confidence intervals of best-fitting estimates of α_D for each test phase. (B) Bars show mean probabilities of making an "old" response to each item type during the recognition test phase. Points show aggregate simulations using best-fitting parameters. (C) Bars for high-match, conflict, and one-new-P items reflect mean probabilities of responding consistently with the D feature. Bars for new-D reflect probabilities of responding consistently with the majority of P features.*

consistent responses to high-match compared to conflict items ($\mu_{HM} = 0.75$, $\sigma_{HM} = 0.12$, $\mu_C = 0.58$, $\sigma_C = 0.18$; $t(29) = 4.82$, $p < 0.001$, $d = 1.08$). Group $Dist_R - Dist_C$ performed similarly, with more D-consistent responses during high-match ($\mu_{HM} = 0.72$, $\sigma_{HM} = 0.09$) compared to conflict items ($\mu_C = 0.49$, $\sigma_C = 0.12$; $t(16) = 5.34$, $p < 0.001$, $d = 2.14$). Group $Sel_R - Sel_C$ did not show a difference in proportions of D-consistent responses between the relevant item types ($\mu_{HM} = 0.95$, $\sigma_{HM} = 0.09$, $\mu_C = 0.94$, $\sigma_C = 0.09$; $t(42) = 0.48$, $p = 0.64$, $d = 0.10$), nor did Group $Dist_R - Sel_C$ ($\mu_{HM} = 0.88$, $\sigma_{HM} = 0.13$, $\mu_C = 0.95$, $\sigma_C = 0.08$; $t(19) = -3.04$, $p = 0.99$, $d = 0.67$). The results of additional post hoc tests to evaluate the pairwise differences in effects between groups are presented in Tables 3 and 4.

These analyses confirmed a connection between GCM parameterizations of selective and distributed attention, and the patterns of behavior that Deng & Sloutsky (2016) hypothesized to be indicative of each. The reader is invited to note that the observed effects presented in Figure A1 directly correspond to the predictions shown in Figure 2, but are reconfigured to highlight within-group effects.

Appendix B
Modeling: Conceptual overview

In the set of analyses described in Appendix A, attention was freely-estimated in each phase to delineate participants according to behavior. Group $Sel_R - Sel_C$, for example, used a strategy of selective attention to the D feature during both test phases of the experiment. While this may be considered to be an effective strategy during categorization, selective attention during recognition resulted in an extremely high proportion of false alarm "old" responses when stimuli contained a novel feature in one of the P dimensions ($\mu_{FA} = 68\%$). This could have happened if 1) selective sampling of D features during training resulted in insufficient memory precision to correctly reject one-new-P items at test; or 2) participants failed to sample sufficient information from the test stimuli themselves, and therefore were not equipped to appropriately weight the novel P features during their decisions.

Simulations presented in Figure B1 illustrate the proposed dissociation between memory precision (represented as η) and decision weight (represented as ζ) components of attention. We present this specification as an alternative to the unitary view in which these components are indistinguishable (represented as α). Panels A and B depict the impacts of attention on the probability of correctly rejecting a one-new-P item during the recognition test, as described above. In order to correctly reject a one-new-P item as "new," the unitary view posits that the observer must distribute attention across dimensions (Panel A). The proposed specification allows for additional nuance: even if the observer aptly distributes decision weights across all dimensions when presented with a one-new-P item during the recognition test ($\zeta_D \approx \frac{1}{7}$), they will not incur an accuracy advantage unless they *also* stored features in all dimensions with sufficient precision during training ($\eta_D \approx \frac{1}{7}$; Panel B).

Panels C and D show an analogous set of simulations for critical conflict items during the categorization test. Because conflict items contain a D feature drawn from one category prototype and the majority of P features from the other, modulating the proportion of attention allocated to the D feature directly impacts how the observer will respond.

The depictions of attention in Panels A and C of Figure B1 reflect the original presentation of the exemplar-similarity framework (Context Model; Medin & Shaffer, 1978), in which encoding strength of exemplar features and the weighting of information from the test probe were described interchangeably as the impetus for observed variability in responses. Contrast this with the similarly unitary description of attention provided by GCM in which exemplar features are perfectly encoded (e.g. $\eta_k = 1$), and decision weights at test are what determine response variability (Nosofsky, 1986). By freely estimating attention within either the Context Model or GCM, these two accounts make identical predictions under certain conditions, despite using incongruous language to describe attention's theoretical actions. The current work presents a novel eye-tracking approach to disentangle these forces, such that fixations during training directly correspond to encoding strength for exemplar features, and fixations during test directly correspond to the weighting of features for making decisions.

Appendix C
Model fitting procedures

We used a binomial likelihood to fit all gaze-based model variants to recognition and categorization test response data from each subject independently. We identified best-fitting parameter values for each model and subject using a three-step procedure. First, we implemented Differential Evolution (DE) using the Python package RunDEMC (`https://github.com/compmem/RunDEMC`) with 50 particles for $100q$ iterations, where q was the number of free parameters in the relevant model. We did this to effectively sample the parameter space and identify reasonable initial values for each subject (Storn & Price, 1997; Brest et al., 2006). Second, we used the DE output values as input to the Nelder-Mead function optimization algorithm implemented in SciPy to identify stable estimates of best-fitting parameters. Third, in the event of failure to meet the base convergence criterion after 10000 iterations, DE sampling recommenced for sets of 100 iterations until convergence was achieved. All parameter values were exponentially transformed to achieve support $(0, \infty)$.

Model fits were assessed using AIC, which accounts for goodness-of-fit as well as model parsimony. Within each comparison, models were selected on the basis

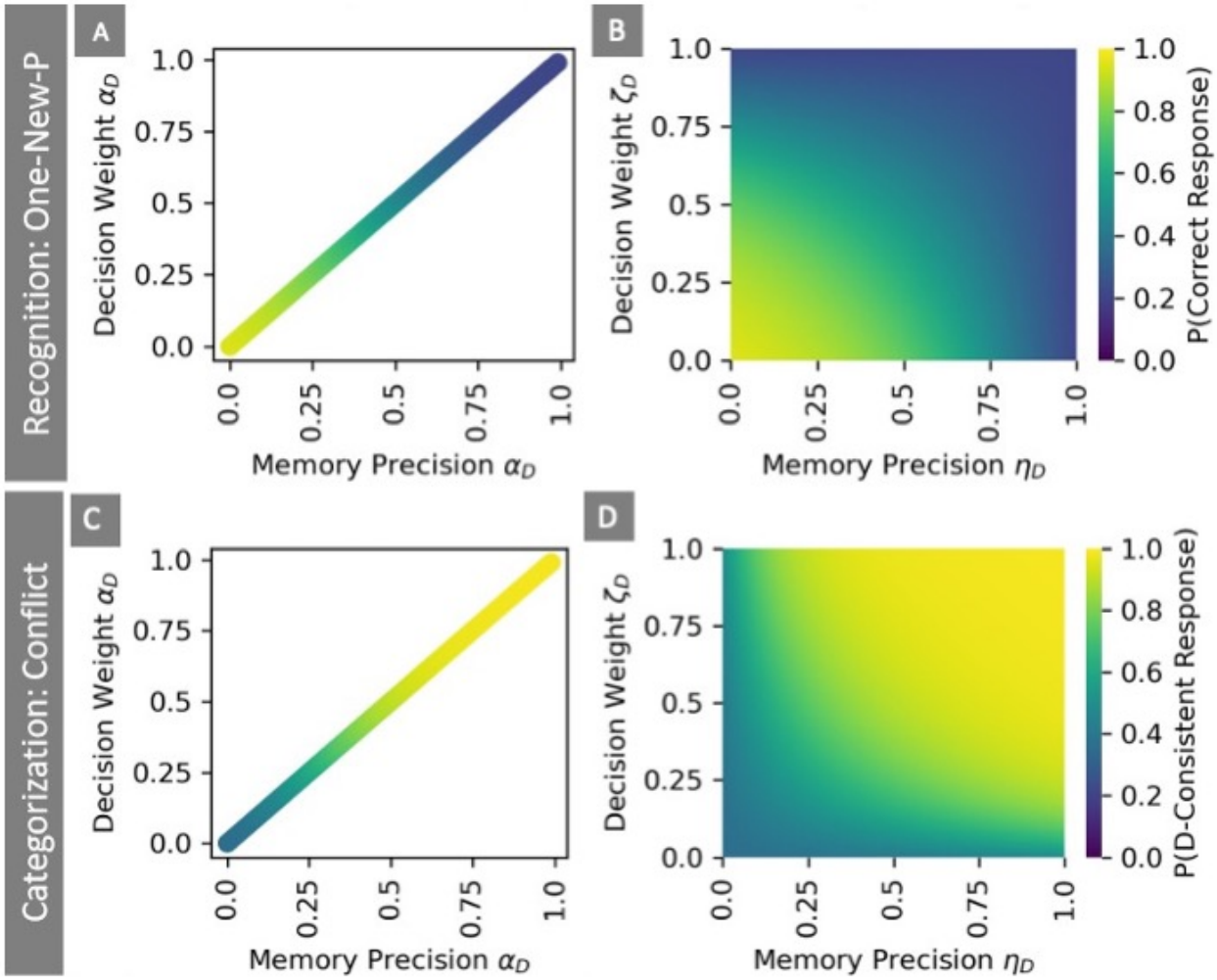

Figure B1

Relating attention to choice probability during critical items. *Panels depict simulated response probabilities. Most parameter values were selected arbitrarily and fixed across simulations; only parameter values representing attention were varied. X and Y values of each panel show the proportion of attention allocated to the deterministic dimension. The proportion of attention allocated to the probabilistic dimensions was specified as* $\sum \alpha_P = 1 - \alpha_D$. *(A) Z values (colors) indicate the probability of correctly rejecting a one-new-P item as "new" during the recognition test. Attention was specified as a single vector where* $\sum \alpha = 1$. *(B) Z values indicate the probability of correctly rejecting a one-new-P item as "new" during the recognition test. Attention was specified as the product of two vectors where* $\sum \eta = 1$ *and* $\sum \zeta = 1$. *(C) Z values indicate the probability of making a categorization response consistent with the deterministic feature of given conflict item. Attention was specified as a single vector where* $\sum \alpha = 1$. *(D) Z values indicate the probability of making a categorization response consistent with the deterministic feature of given conflict item. Attention was specified as the product of two vectors where* $\sum \eta = 1$ *and* $\sum \zeta = 1$.

of lowest mean AIC across subjects.

After identifying best-fitting parameters for each model and subject, we simulated responses using the relevant participant's gaze data as input. We then aggregated model-simulated responses within participant group, test phase, and item type. This allowed us to evaluate each model by its ability to re-produce the key effects. If gaze is indeed an effective index of latent attention, we determined *a priori* that a direct input approach should predict significant differences in responses between *selective* and *distributed* attention strategy groups.

Appendix D

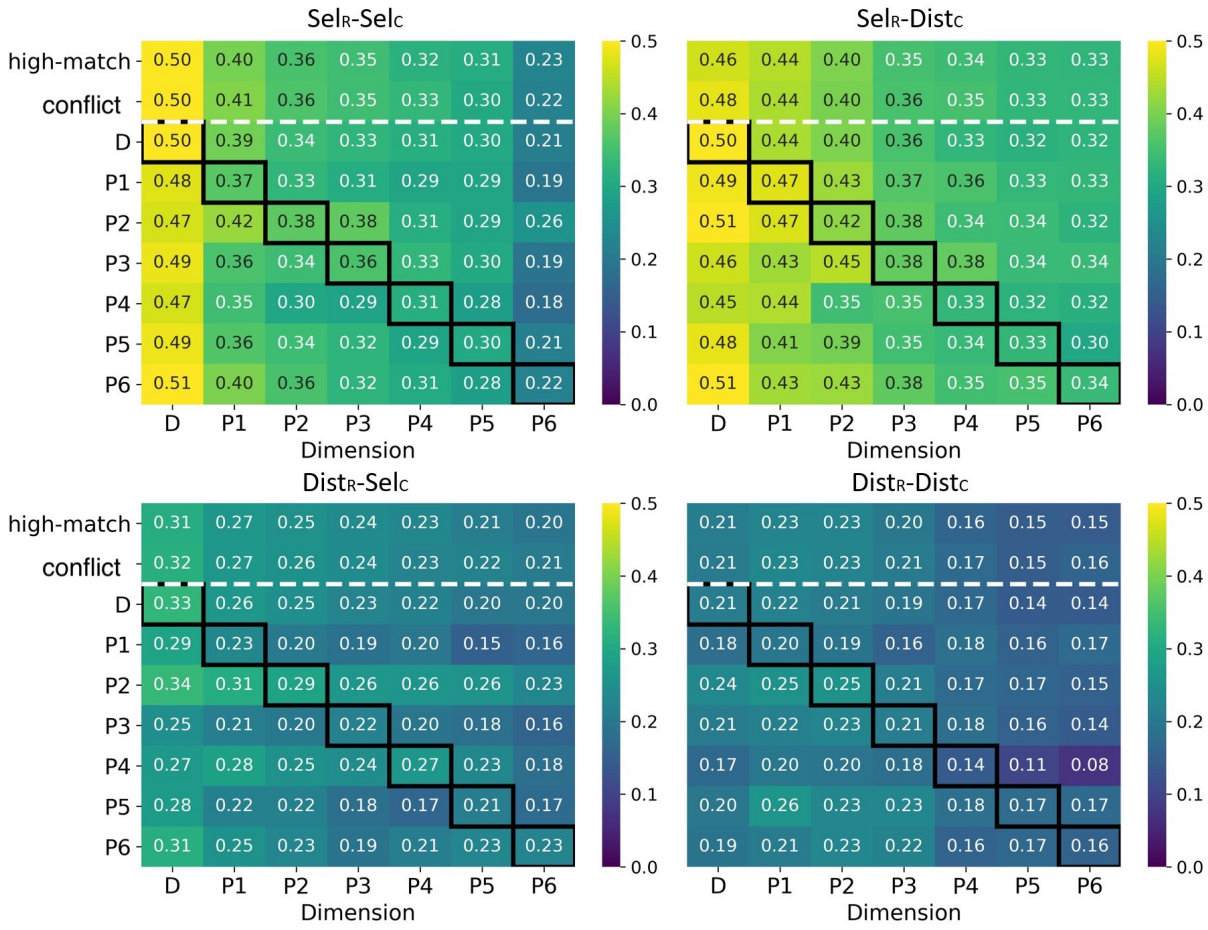

Figure D1

Recognition: Combined gaze-based memory precision and decision weights. *D=deterministic, P=probabilistic. Heatmaps show aggregate feature discriminability across subjects. X-ticks indicate stimulus dimensions, where P dimensions were rank-ordered within-subject according to gaze preference. Y-ticks indicate the dimension location of a novel feature within the relevant subset of trials. Subject-wise discriminability maps were calculated by subjecting raw dwell time data to best-fitting model-based transformations.*

Extended eye-tracking results

A one-way ANOVA revealed significant differences between groups in proportions of gaze allocated to the D dimension during the recognition test ($F(3,106) = 6.28), p < 0.001^*$). Post hoc comparisons of means using Tukey's honestly significant difference (HSD) revealed that Groups $Sel_R - Sel_C$ and $Sel_R - Dist_C$ tended to look more at the D feature more than Groups $Dist_R - Sel_C$ and $Dist_R - Dist_C$ ($Sel_R - Sel_C$ vs. $Dist_R - Sel_C$: $p < 0.05^*$, $Sel_R - Sel_C$ vs. $Dist_R - Dist_C$: $p < 0.01^*$, 2 vs. 3: $p = 0.08^\cdot$, $Sel_R - Dist_C$ vs. $Dist_R - Dist_C$: $p < 0.05^*$). Groups $Sel_R - Sel_C$ and $Sel_R - Dist_C$ ($p = 0.98$) and Groups $Dist_R - Sel_C$ ($p = 0.86$) did not differ from one another.

Figure D1 shows aggregate model-predicted feature discriminability during the recognition test, which was calculated using Equation 1 with $d_j = 1$ and best-fitting parameters from Model $C - B$. Values therefore combine transformed gaze data from both training and test to visualize attention as a product of *eta* and *zeta*. The heatmaps display features of the new-D and one-new-P items as a matrix, where Y-ticks indicate the position of a novel feature. P dimensions on the X-axis were rank ordered within-subject by gaze preference. Gaze-informed estimates of discriminability show that Groups $Sel_R - Sel_C$ and $Sel_R - Dist_C$ favor D features more than P features when making decisions, whereas Groups $Dist_R - Sel_C$ and $Dist_R - Dist_C$ do not appear to show any discriminability bias toward a particular dimension.

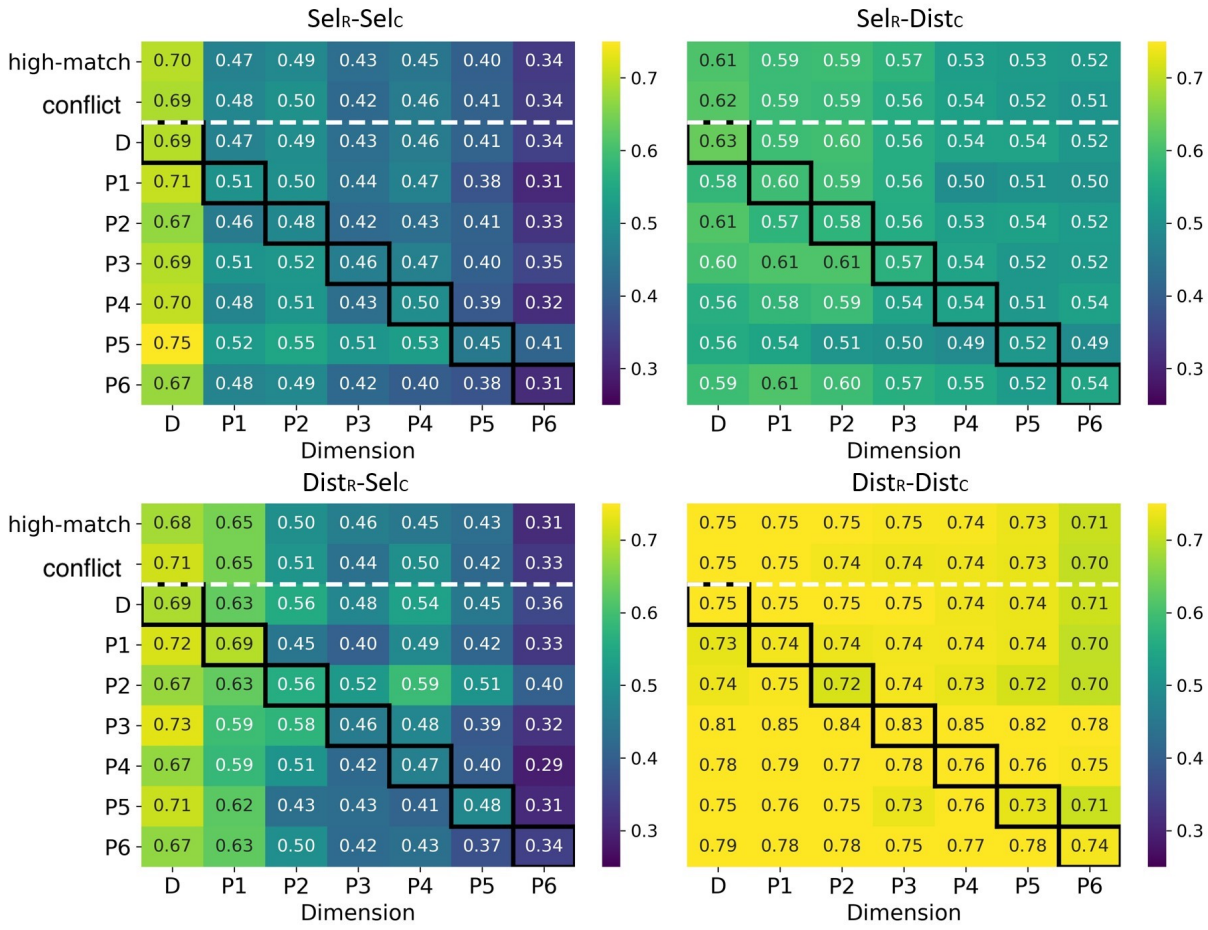

Figure D2

Categorization: Combined gaze-based memory precision and decision weights. *D=deterministic, P=probabilistic. Heatmaps show aggregate feature discriminability across subjects. X-ticks indicate stimulus dimensions, where P dimensions were rank-ordered within-subject according to gaze preference. Y-ticks indicate the dimension location of a novel feature within the relevant subset of trials. Subject-wise discriminability maps were calculated by subjecting raw dwell time data to best-fitting model-based transformations.*

A one-way ANOVA revealed significant differences between groups in raw proportions of gaze allocated to the D dimension during the categorization test as well ($F(3,106) = 9.69, p < 0.001^*$). Post hoc comparisons of means using Tukey's HSD revealed that Groups $Sel_R - Sel_C$ and $Dist_R - Sel_C$ tended to look more at the D feature more than Groups $Sel_R - Dist_C$ and $Dist_R - Dist_C$ ($Sel_R - Sel_C$ vs. $Sel_R - Dist_C$: $p < 0.01^*$, $Sel_R - Sel_C$ vs. $Dist_R - Dist_C$: $p < 0.001^*$, $Sel_R - Dist_C$ vs. $Dist_R - Sel_C$: $p = 0.57$ n.s., $Dist_R - Sel_C$ vs. $Dist_R - Dist_C$: $p < 0.05^*$). Groups $Sel_R - Sel_C$ and $Dist_R - Sel_C$ ($p = 0.23$) and Groups $Sel_R - Dist_C$ and $Dist_R - Dist_C$ ($p = 0.33$) did not differ from one another.

Figure D2 shows aggregate model-predicted feature discriminability during the categorization test. The heatmaps show that Groups $Sel_R - Sel_C$ and $Dist_R - Sel_C$ show higher attention to D features higher than P features when making categorization decisions. As in Figure D1, values combine transformed gaze data from both training and test to visualize attention as a product of *eta* and *zeta*. Although Groups $Sel_R - Dist_C$ and $Dist_R - Dist_C$ differ in overall feature discriminability, neither group appears to show a discriminability bias in favor of a particular dimension. These visualizations of transformed gaze measures are consistent with the observed behavioral effects within each group: participants whose responses were characterized by a selective attention strategy looked more at D; those whose responses were characterized by a distributed attention strategy sampled feature information more evenly.

To summarize the results, Group $Sel_R - Sel_C$ tended to fixate to the D dimension during training whereas Group $Dist_R - Dist_C$ tended to sample a broader range of features (Figure 5). These sampling behaviors were directly reflected in the distribution of decision strategies used at test, with Group $Sel_R - Sel_C$ prioritizing the D dimension during both phases and Group $Dist_R - Dist_C$ distributing attention across dimensions (Figures D2 and D1). We note, however, that our proposed framework could have predicted consistently D-selective or distributed decision strategies across phases using the *same* profile of fixation biases by simply selecting linking function parameters that magnified or diffused determinism as needed. It is therefore important to highlight that our framework was also effective for predicting choices among participants who *shifted strategies* from recognition to categorization, given that this could only occur if gaze patterns during test shifted as well (Groups $Sel_R - Dist_C$ and $Dist_R - Sel_C$). Although these groups appeared to show similar patterns of sampling and storage of features in aggregate (Figure 5), combining the influences of gaze-informed memory precision and decision weights produced the expected patterns of behaviors during both test phases (Figures D2 and D1).