

Temporal characteristics of overt attentional behavior during category learning

Lihan Chen · Kimberly M. Meier · Mark R. Blair · Marcus R. Watson · Michael J. Wood

Published online: 14 November 2012
© Psychonomic Society, Inc. 2012

Abstract Many theories of category learning incorporate mechanisms for selective attention, typically implemented as attention weights that change on a trial-by-trial basis. This is because there is relatively little data on within-trial changes in attention. We used eye tracking and mouse tracking as fine-grained measures of attention in three complex visual categorization tasks to investigate temporal patterns in overt attentional behavior within individual categorization decisions. In Experiments 1 and 2, we recorded participants' eye movements while they performed three different categorization tasks. We extended previous research by demonstrating that not only are participants less likely to fixate irrelevant features, but also, when they do, these fixations are shorter than fixations to relevant features. We also found that participants' fixation patterns show increasingly consistent temporal patterns. Participants were faster, although no more accurate, when their fixation sequences followed a consistent temporal structure. In Experiment 3, we replicated these findings in a task where participants used mouse movements to uncover features. Overall, we showed that there are important temporal regularities in information sampling during category learning

that cannot be accounted for by existing models. These can be used to supplement extant models for richer predictions of how information is attended to during the buildup to a categorization decision.

Keywords Attention · Eye tracking · Categorization · Eye movements · Optimization · Learning · Error · Modeling · Temporal regularity · Visual cognition

To interact with the world, people must learn how to sort through the vast amount of information it presents. One of the difficulties in doing so is that much of this information is irrelevant to the task at hand. Consequently, it is essential to allocate limited cognitive resources to important aspects of the environment through selective attention. The notion of selective attention is recognized within a broad range of research areas, from working memory (Awh, Anillo-Vento, & Hillyard, 2000) and priming (Tipper & Cranston, 1985) to cognitive development over the human life span (Plude, Enns, & Brodeur, 1994). Category learning, a paradigm that elicits voluntary control of attention, provides a unique opportunity for observing and formally analyzing how attention shifts over time in response to the relevance of available information. The present study uses eye tracking and mouse tracking to document temporal characteristics of overt top-down attentional behavior during categorization tasks and how this changes over the course of learning.

One of the most influential models in the category-learning literature, the generalized context model (GCM; Nosofsky, 1986), models the role of attention in learning as a weighting mechanism for incoming stimulus information. This type of model hypothesizes that relevant information gains more attention weight as a consequence of learning. For example, after learning the task of distinguishing limes from lemons, the GCM assigns a higher attention weight to color than to shape. While exact implementations

L. Chen · M. R. Blair
Cognitive Science Program, Simon Fraser University,
Burnaby, Canada

M. R. Blair (✉) · M. J. Wood
Department of Psychology, Simon Fraser University,
8888 University Drive,
Burnaby, BC V5A 1S6, Canada
e-mail: mblair@sfu.ca

K. M. Meier · M. R. Watson
Department of Psychology, University of British Columbia,
Vancouver, British Columbia, Canada

Present Address:
M. J. Wood
School of Psychology, University of Kent,
Canterbury, UK

differ, nearly all models of categorization containing an attention component, such as ALCOVE (Kruschke, 1992) and SUSTAIN (Love, Medin, & Gureckis, 2004), incorporate it as weights applied simultaneously to incoming stimulus dimensions. In these models, attention weights are adjusted on the basis of trial-by-trial feedback to produce more accurate classification decisions. Originally, these weighting mechanisms were supported only by indirect measures, such as participants' performance on transfer stimuli. Recent studies on the relationship between eye-tracking measures and the attention weights of category-learning models (e.g., Rehder & Hoffman, 2005a) have shown that gaze behavior can also be used as a directly measurable proxy for modeling attention weights (Rehder & Hoffman, 2005b).

One important limitation of implementing attention as weights is that it does not capture participants' use of attention to obtain stimulus information during individual decisions. While attention weights do a fine job of predicting the amount of time spent looking at visual features of varying usefulness (Blair, Watson, & Meier, 2009; Rehder & Hoffman, 2005a), they say nothing about how the allocation of attention might progress over the course of a trial. As such, they cannot address temporal regularities that emerge as these decisions are being made. What information is attended to (and when) might be invariant from trial to trial, or attention could be highly dynamic and sensitive to within-trial context. Weighting models make no predictions either way.

Although categorization in the real world is not a passive process in which one is given a set of necessary and sufficient information for making decisions, much of the literature effectively treats it this way. Participants in experiments generally classify stimuli that require little processing and few overt behaviors to obtain information. The primary dependent variables, responses and response times, conflate all the processes leading up to a decision, including choices about what information to attend to and when. It is natural, therefore, that categorization models do not incorporate these processes. Because experiments are usually designed to test theoretical predictions (i.e., model predictions), processes that are not directly addressed by prominent theories are often overlooked. Thus, traditional categorization models, experimental paradigms, and methods of analysis all have mutually reinforcing tendencies to ignore within-decision attentional shifts. As a result, relatively little work has been done that even elicits, much less describes or predicts, attentional behavior in categorization.

Attentional behaviors are only now beginning to be addressed. Mouse-driven displays that require participants to actively select information for viewing have been used in several studies (e.g., Matsuka & Corter, 2008; Nelson, McKenzie, Cottrell, & Sejnowski, 2010). Another rich source of information is studies of eye movements, which provide a steady stream of information about how people

obtain information from the environment to learn and make decisions (Blair, Watson, Walshe, et al., 2009; Kruschke, Kappenman, & Hetrick, 2005; Rehder & Hoffman, 2005a, b). Overt top-down attentional behaviors are interesting in their own right, but there are also indications that accounting for these processes may improve categorization response predictions. For example, the amount of time people spend looking at stimulus features during feedback on incorrect trials is predictive of learning speed (Watson & Blair, 2008). Colner and Rehder's (2009) exemplar fragment model has predicted the key findings of a number of published studies using eye tracking to obtain estimates of how effectively exemplars are encoded in memory, which in turn are used to predict responses.

Previous work has shown that after categories are learned, participants' total time spent fixating on stimulus features during a single decision (i.e., a single trial in a categorization experiment) is less for irrelevant features than for relevant ones (Blair, Watson, & Meier, 2009), but this measure conflates the duration of fixations and the number of fixations. The latter measure does change with learning: Irrelevant features are fixated less often (Rehder & Hoffman, 2005a), but it is unknown whether the durations of individual fixations are sensitive in the same way. Since these measures are independent, understanding how each of them operates is necessary to accurately describe overt attentional behavior. The fact that people make shorter fixations to task-irrelevant words when reading (Kaakinen, Hyona, & Keenan, 2002; van der Schoot, Vasbinder, Horsley, & van Lieshout, 2008) suggests that attentional learning impacts not only the number of fixations, but also individual fixation durations, at least with sufficient training.

Another important aspect of overt attentional behavior is the order in which information is viewed. During free viewing and recognition tasks, individuals develop idiosyncratic *scanpaths*—sequences of fixations—for particular images (cf. Humphrey & Underwood, 2010; Noton & Stark, 1971). This order of viewing information has recently been investigated in the context of categorization learning. As people learn a category structure, changes occur in the temporal order in which they view stimulus features during individual decisions. If a category structure is designed such that there is a most efficient order to attend to features, participants begin to learn this order over time and develop stereotypical sequences of eye movements to particular features, even if the optimal sequence changes from trial to trial (Blair, Watson, Walshe, et al., 2009). Does this develop solely in response to task demands, or do participants develop such sequences even when the task does not favor one pattern over another? Rehder and Hoffman's (2005a) study of the classic Shepard, Hovland, and Jenkins (1961) type II category structures suggests that sequential patterns of attention tend to develop even when they are not inherently

encouraged by the structure itself, although there was no quantitative measure of this finding and no estimate of its statistical reliability was reported. In the present study, we investigate the formation of temporal patterns of fixations, using two structures that encourage their formation (Experiments 1 and 3) and two that do not (Experiment 2). We develop a novel measure of the degree to which individual participant's overt attentional behavior is temporally patterned—that is, to what extent they view stimulus features in a stereotypical order over repeated trials. We investigate whether the degree of temporal patterning is predictive of accuracy or response time. More ordered scanpaths are associated with shorter response times on visual search tasks (Myers & Gray, 2010), but not with differences in accuracy on recognition tasks (Humphrey & Underwood, 2010). However, numerous differences between these tasks and categorization learning make predictions in this domain difficult.

The present article reports three experiments on temporal qualities of attention in category learning. Our primary measures are mean duration of single fixations, with a focus on fixation duration differences between relevant and irrelevant stimulus features, and temporal patterning of those fixations, as indicated by a quantitative measure of temporal ordering effects we call *strength of temporal pattern* (STP). STP indexes the degree to which information sampling is temporally structured or randomly ordered by assessing how uniform the probability of fixating different stimulus features is across trial time.

The first experiment used a hierarchical rule-based category structure for which the irrelevant feature differed trial by trial and depended on the superordinate category of the stimulus. We replicated the temporal ordering shown in prior studies with this category structure (e.g., Blair, Watson, Walshe, et al., 2009) and investigated the temporal measures that were the focus of the present work. Two questions that followed from the results of Experiment 1 were addressed in Experiments 2 and 3. First, do participants develop stereotypical temporal patterns of fixations to stimulus features even when there is no advantage to doing so? Experiment 2 used two different category structures for which the two relevant features were equally informative. Since it did not matter which order participants viewed these two features in, any temporal patterns that developed were not inherent to the category structure. Second, are the temporal aspects of attention we found in Experiments 1 and 2 restricted to cases involving the oculomotor system, or do they extend to other overt attentional mechanisms? Experiment 3 revisited the hierarchical structure from Experiment 1, using a mouse-driven interface to investigate whether the temporal effects discovered in the first two experiments were properties exclusive to eye movements or, instead, were properties of overt attentional behaviors more generally.

Experiment 1

Method

Participants

Two hundred twenty students from Simon Fraser University with normal or corrected-to-normal vision participated in this experiment for course credit.

Stimuli and categories

Stimuli were images of fictitious microorganisms represented by organelles occupying the microorganisms' three arms (see Fig. 1). Features subtended approximately 3° of visual angle, and their centers were located 10.6° apart. Features could take one of two possible forms, comprising the categories A1, A2, B1, and B2. On any given trial, only two features were necessary for making a correct classification (see Fig. 1): A1 and A2 required knowledge of features at locations 1 and 2, while B1 and B2 required knowledge of features at locations 1 and 3. Feature images and locations remained constant for each participant but were counter-balanced across participants.

Procedure

Participants learned to classify the microorganisms on the basis of corrective feedback. Participants began each trial by clicking on a central fixation cross to reveal a microorganism. After inspecting the stimulus, participants clicked to advance to a response screen and indicated their classification decision by clicking on a box corresponding to the category label. The stimulus, which disappeared from the

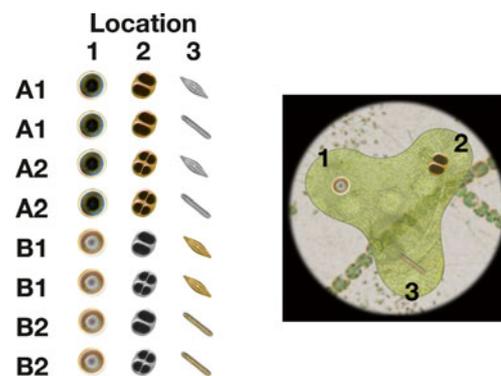


Fig. 1 Left: Category structure of Experiment 1. The stimulus feature at location 1 was relevant for both A and B categories. For stimuli in the A categories, the feature at location 2 is relevant. For stimuli in the B categories, the second relevant feature is at location 3. Features that are not relevant for a category are shown in gray. Right: Example stimulus. The location and relevance of features are counterbalanced across participants and remain constant for the entire experiment

screen during the response phase, was re-presented concurrently with performance feedback. If participants were correct, the category label they selected turned green; if they were incorrect, their selection turned red, and the correct response turned green. Feedback remained on the screen until participants made a mouse-click to proceed to the next trial. The number of trials participants completed was adapted to their performance: If participants reached the learning criterion defined by 25 correct classifications in a row before 200 trials were completed, feedback was removed, and participants completed 72 trials wherein the button they clicked turned purple. Stimuli were randomized such that participants saw a new random order of all stimuli every 8 trials. The task took approximately 30 min to complete.

A Tobii X120 eyetracker sampling at 60 Hz with a spatial resolution of 0.5° recorded participants' eye gaze. Participants were seated approximately 70 cm from the computer screen. Fixations were generated from raw gaze data using a modified dispersion threshold algorithm (Salvucci & Goldberg, 2000) with temporal and spatial thresholds of 75 ms and 1.9° , respectively. Participants were prompted to correct for position drift with a display showing their eye positions if the previous trial had fewer than 80 % valid gaze points collected. Invalid gaze points are defined as samples for which the eyetracker cannot register eye position coordinates and, instead, outputs an invalid marker. Further drift correction was performed post hoc using fixations to the central cross as reference points.

Results and discussion

Thirteen participants were excluded from the analysis for failing to reach our minimum gaze collection criterion of 75 % valid gaze points. A further 61 participants were excluded for failing to reach learning criterion. This left 146 participants for analysis. Their mean overall accuracy was .81 ($SD = .07$), with a mean of 83.27 ($SD = 39.04$) trials to reach learning criterion.

Duration of fixations to features across learning

The mean duration of single fixations was examined in three blocks: (1) all trials before criterion was reached, (2) the first 48 postcriterion trials, and (3) the last 48 postcriterion trials. Mean fixation durations decreased over the experiment for both the relevant and irrelevant features, and the decrease was larger for the irrelevant feature (for the relevant feature, $M = 298, 297,$ and 283 ms, $SD = 71, 81,$ and 77 , for the three blocks, respectively; for the irrelevant feature, $M = 298, 243,$ and 233 ms, $SD = 81, 58,$ and 77 ; see Fig. 2). Fixation durations were analyzed using a 3 (block: 1–3) \times 2 (relevance: relevant for the trial, irrelevant for the trial) within-subjects ANOVA. There was a significant main effect of

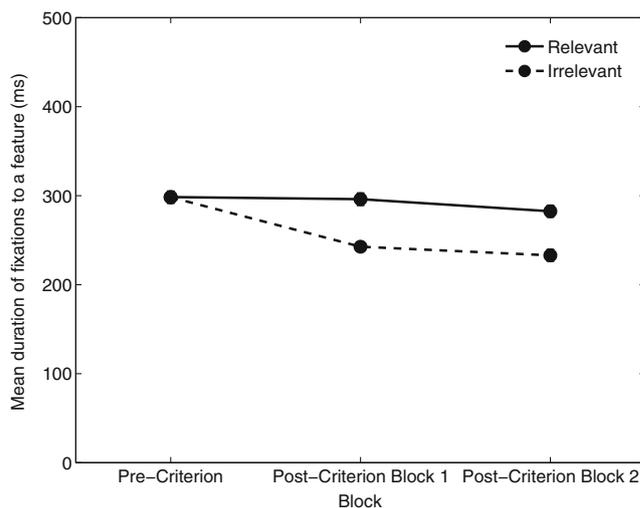


Fig. 2 Mean duration of fixations to relevant and irrelevant features during stages of Experiment 1, collapsing across category. See Fig. 1 for feature relevance. Error bars reflecting the standard error of the mean are too small to be seen in the figure

block, $F(2, 290) = 58.51, p < .001$, and a significant main effect of relevance, $F(1, 145) = 84.47, p < .001$. This was qualified by a block \times relevance interaction, $F(2, 290) = 40.00, p < .001$. Feature differences emerged only after criterion was reached. There were simple main effects of relevance for the second and third blocks, $F(1, 290) = 465.03$ and 396.61 , respectively; both $ps < .001$, but not for the first, $F(1, 290) = 0.006, p = .95$. Durations decreased to both types of relevance: There was a significant effect of block for both relevant, $F(2, 290) = 22.15, p < .001$, and irrelevant, $F(2, 290) = 392.53, p < .001$, features.

This finding demonstrates that the relevance of a stimulus feature for categorization has an impact on the durations of individual fixations. This extends previous findings that showed that the probability of fixating relevant stimulus features was greater than that of fixating irrelevant ones (Rehder & Hoffman, 2005a). It also clarifies some ambiguity in studies using mean total fixation duration as a measure (e.g., Blair, Watson, & Meier, 2009) by showing that decreases in fixation time to irrelevant information are due not only to a reduction in number of fixations, but also to a reduction in the durations of these fixations.

Strength of temporal pattern: A measure of within-trial consistency

The next phenomenon of interest is the degree to which participants view stimulus features in a stereotypical order within individual trials (i.e., the degree to which they develop a consistent scanpath). Figure 3 shows the likelihood of fixating each of the three stimulus features as a function of trial time for 1 participant. Each trial of the experiment for

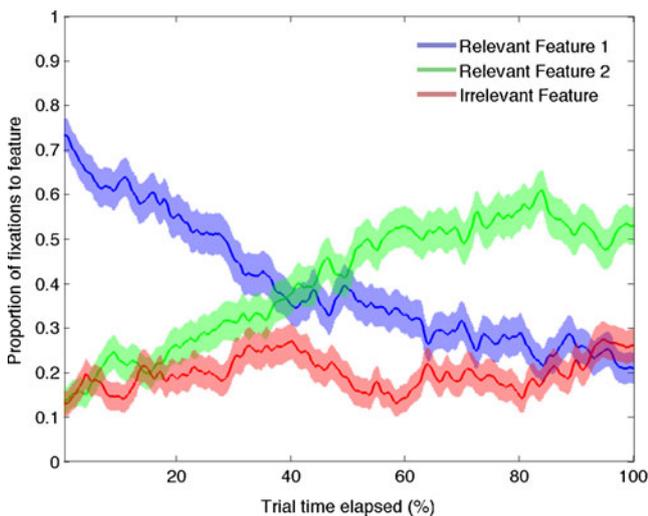


Fig. 3 Proportion of fixations to features for a single participant as a function of trial time. Trial time ranges from the start of the first fixation to a stimulus feature (0 %) to the end of the last fixation to a feature (100 %). “Relevant Feature 1” represents the feature at location 1 in Fig. 1. “Relevant Feature 2” represents the feature at location 2 (for A trials) or 3 (for B trials), and “Irrelevant Feature” represents location 3 (for A trials) or 2 (for B trials). Shaded areas represent the standard error of the mean

this participant was separated into 500 time bins, and the features fixated during each bin were averaged across all trials. Intuitively, the participant shows a stronger temporal pattern for obtaining relevant information than for obtaining irrelevant information. At the beginning of the trial, around 70 % of the fixations were to relevant feature 1 (in blue), which was not fixated at all toward the end of the trial. Fixations to relevant feature 2 (in green) follow almost exactly the opposite pattern. In contrast, fixations to the irrelevant feature (in red) scatter throughout the trial, never wandering very far from the mean value.

To quantify the strength of a participant’s use of a consistent temporal pattern for attending to each feature within a trial, we calculated their STP. The STP is the mean absolute deviation from the mean fixation proportions on that feature, divided by 0.5. STP indexes the extent to which fixation proportions are uniformly distributed across trial time. Since the upper limit of the mean absolute deviation is half the upper limit of the fixation proportion—that is to say, .5—the measure ranges theoretically from 0 (randomly ordered with no structure) to 1 (perfectly consistent in temporal structure). Figure 4a shows the calculation of STP on relevant feature 1 for the same participant as that shown in Fig. 3. As is shown in the figure, a moderate STP (.48) was obtained for this feature. Conversely, since the participant had no preference for when to look at the irrelevant feature during these trials (see Fig. 3), a very low STP (.12) was obtained. Since the temporal pattern on a given feature would be meaningless for any trial that contained

only one fixation or contained no fixations to that feature, it was considered undefined in those cases. Therefore, a trial required at least two fixations, at least one of which was on a particular feature, to be included in the calculation of STP for that feature; otherwise, data for that trial were considered missing.

Within-trial temporal consistency across learning

As with fixation duration, we investigated participants’ temporal patterns by calculating STP in three blocks over the experiment. Figure 5 shows the mean STP, averaged across participants and broken into three blocks: one containing all trials before the learning criterion was reached and two bins of 48 postcriterion trials. As can be seen in the figure, mean STP increased over the experiment for both the relevant and irrelevant features, and the increase was larger for the irrelevant feature (for the relevant features, $M = .19, .31, .38$, $SD = .07, .13, .16$, respectively, for the three blocks; for the irrelevant feature, $M = .14, .19, .23$, $SD = .04, .06, .09$). Again, a 3 (block) \times 2 (relevance) within-subjects ANOVA was performed. One participant was excluded from this analysis for having no fixations to the irrelevant feature during block 3. There were significant main effects of both block, $F(2, 288) = 176.77, p < .001$, and relevance, $F(1, 144) = 226.50, p < .001$, qualified by a significant block \times relevance interaction, $F(2, 288) = 63.96, p < .001$. There were significant simple main effects of relevance at every level of block, $F(1, 288) = 119.18, 787.51, \text{ and } 1,158.30$, respectively; all $ps < .001$. STP increases for both relevance types: There was a significant effect of block for both relevant, $F(2, 288) = 961.39, p < .001$, and irrelevant, $F(2, 288) = 200.84, p < .001$, features.

Predicting response time and accuracy from attentional consistency

Participants’ STPs were associated with faster responses but were not predictive of accuracy. This conclusion stems from two multiple regression analyses predicting response time and accuracy. In both, we used postcriterion STP for relevant and irrelevant features as predictors. The former model was moderately predictive of response time (adjusted $R^2 = .15$), $F(2, 143) = 14.26, p < .001$. The coefficients for both the relevant and irrelevant STPs were significantly lower than zero ($\beta = -.30$ and $-.19$), $t(144) = -3.5$ and $-2.2, p < .001$ and $.03$, respectively, indicating that less-patterned participants tended to take longer to respond. The latter model was not predictive of accuracy, $F(2, 143) = 2.19, p = .12$, nor were coefficients for relevant and irrelevant STPs in this model significantly different from zero, $t(144) = 1.77$ and $0.31, p = .08$ and $.75$, respectively.

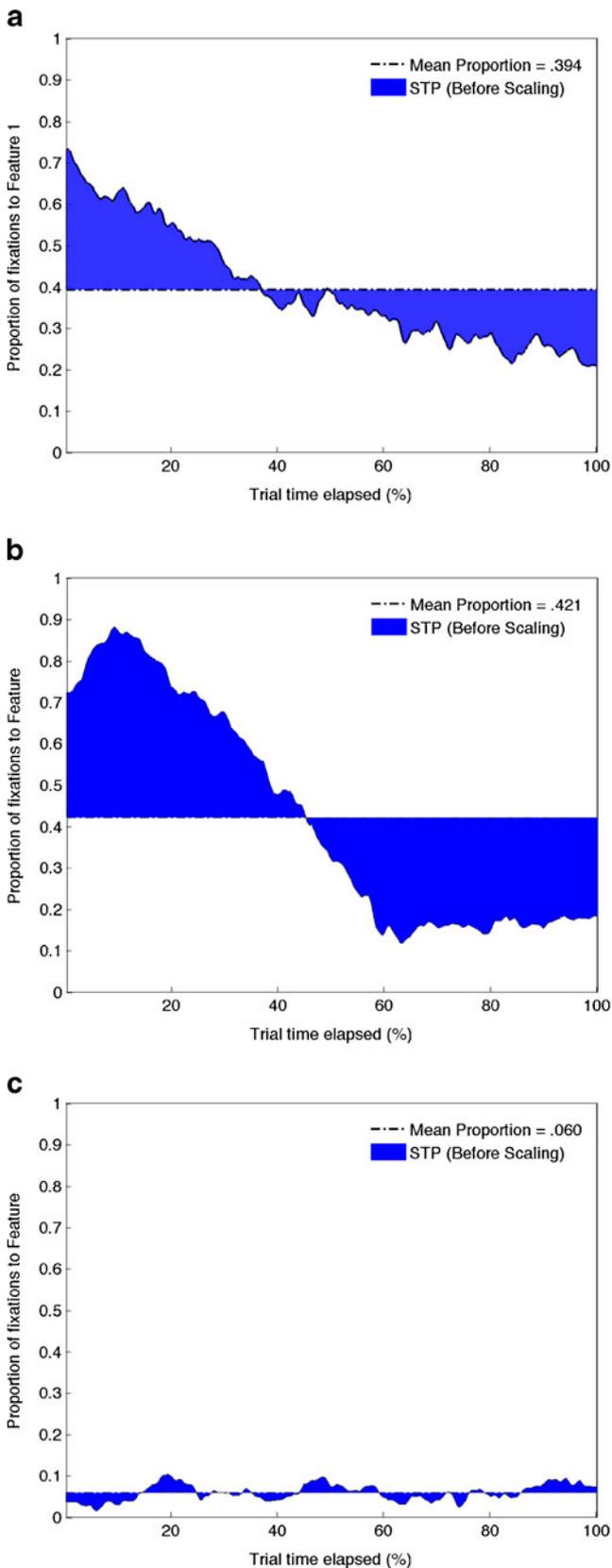


Fig. 4 The mean absolute deviation from the mean—namely, the area of the shaded region—is used to quantify the strength of the participant’s temporal pattern of attention shifts. This is scaled by dividing by .5 so that the theoretical range of the measure is [0, 1]. **a** Visual representation of the strength of temporal pattern (STP) associated with relevant feature 1 of the participant shown in Fig. 3 (STP = .48). **b** For reference, the feature with the highest STP value obtained in the experiment (STP = .998). This is for relevant feature 1. **c** The feature with the lowest STP value obtained in the experiment (STP = .07). This is for the irrelevant feature

A possible reason for the temporal regularities we found is that the category structure used in this experiment may have encouraged participants to fixate features in a particular order. It is true that participants can achieve perfect accuracy by fixating all three features, and there is no particular order that is necessary if participants choose to do this. Nevertheless, if participants look at location 1 first, they can target their next fixation on the basis of its value and eliminate the need for a third fixation. Participants have been shown to be capable of this when learning category structures of this kind (Blair, Watson, Walshe, et al., 2009). Experiment 2 used two different category structures that eliminated any particular incentive to fixate features in a specific order, allowing us to confirm whether temporal patterns develop even without structural incentives.

Experiment 2

Participants in Experiment 2 learned one of two category structures that approximately corresponded to the four-category information-integration and rule-based structures used by Maddox, Filoteo, Hejl, and Ing (2004). These structures are difficult to learn perfectly and may encourage a variety of learning strategies but equate the number of category dimensions that are required for classification. These

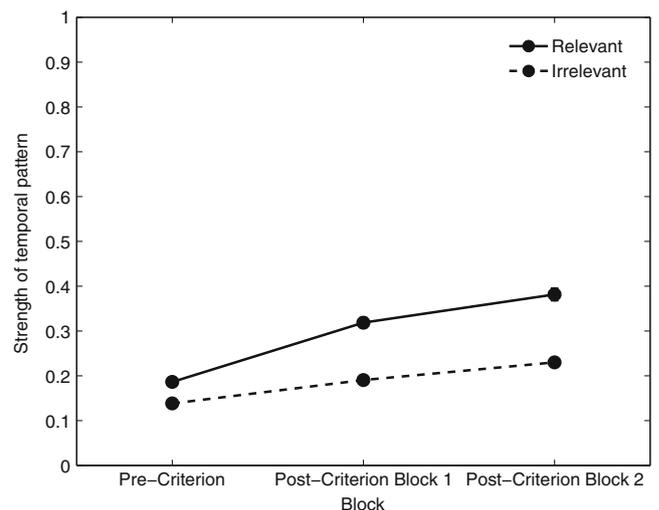


Fig. 5 Strength of the temporal pattern of relevant and irrelevant features during stages of Experiment 1. Error bars reflecting the standard error of the mean are too small to be seen in the figure

categories have not been tested using eye tracking before and can provide an excellent source of data about how eye movements differ, or not, across complex learning tasks. They also provide less incentive for developing a temporal pattern than does the structure used in Experiment 1.

Method

Participants

Seventy-eight undergraduates with normal or corrected-to-normal vision participated in this experiment.

Stimuli and categories

Participants classified microorganisms similar to those used in Experiment 1, but with continuous-valued, rather than binary, features. Participants were randomly assigned to one of two category structures. Each was defined by two decision bounds: structure 1 by an “x” (see Fig. 6, left) and structure 2 by a “+” (Fig. 6, right). Each feature varied along one of three dimensions (rotation, size, or curvature; see Fig. 7). Images for each feature were created to vary between two extremes across 90° of change. Distinct stimuli were generated by combining these features according to the bivariate distribution parameters in Table 1. The slopes of the ideal decision boundaries separated the four categories, labeled A–D, such that two of the three features were relevant for classification. Feature location and relevance were counterbalanced across participants. Note that many feature values close to the category boundaries are very hard to perceptually discriminate, and thus learners are likely to have lower asymptotic accuracy.

Procedure

During the instruction phase, values 0, 45, and 90 (see Fig. 7) of each feature were shown to familiarize participants with

each feature’s range of variation. The procedure for the category-learning task was identical to that in Experiment 1 in both conditions, with the exception of number of trials completed. Because we expected the continuous feature values used in this experiment to lead to more gradual learning than in Experiment 1 (Zeithamova & Maddox, 2007), no learning criterion was implemented. This task contained 360 trials total and took approximately 40 min to complete. As with Experiment 1, participants’ eye movements were recorded.

Results and discussion

Three participants were excluded for failing to complete the experiment. A further 10 participants were excluded due to eyetracker failure, and 9 participants for failing to meet our minimum gaze collection criterion of 75 % valid gaze points. Three participants solving structure 2 were excluded for performing at chance. Analyses were conducted on 25 participants solving structure 1 and 28 participants solving structure 2. Accuracies in the two conditions were similar throughout the experiment. Mean overall accuracy for participants learning structure 1 was .74 ($SD = .10$), and for structure 2 was .70 ($SD = .16$), and there was no significant difference between conditions overall, $t(51) = 1.01$, $p = .32$, or in the last block of 24 trials, $t(51) = 0.49$, $p = .63$. Participants’ performance is comparable to that in a similar study by Maddox et al. (2004). Due to the lack of differences between these groups in the initial analyses, we collapsed across the two conditions. Since participants did not reach a clear criterion point in this task, the experiment was divided into four bins of 90 trials each for the purposes of analysis.

Duration of fixations to features across learning

As in Experiment 1, the mean duration of individual fixations decreased over the experiment for both the relevant

Fig. 6 Example category structures from each condition. Category membership is represented in shades of gray. The left panel shows structure 1; the right shows structure 2. Category features vary along three dimensions: orientation, size, and curvature (see Fig. 7 for some examples). The x and y axes represent variation along two of these features. An irrelevant third feature was generated for both structures. Parameters used to generate the structures are described in Table 1

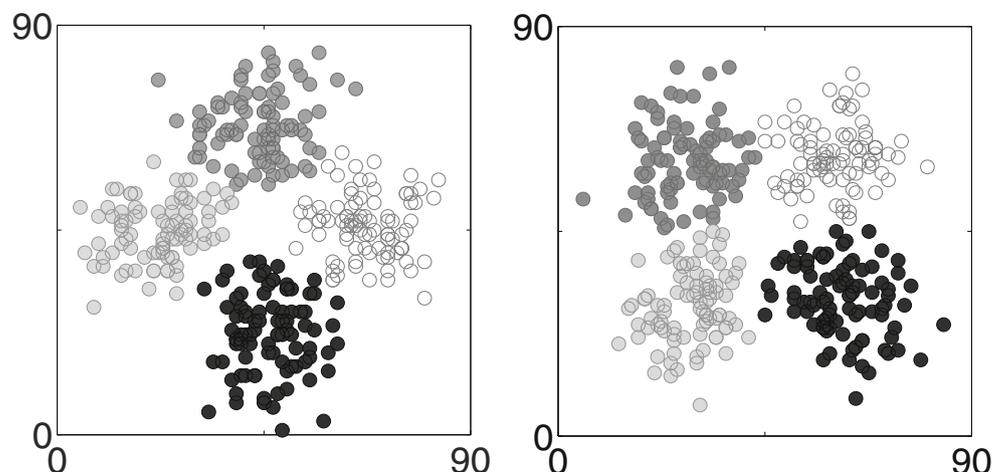




Fig. 7 Sample stimulus for Experiment 2 (above) and examples of the moderate and extreme values of each of the three continuous-valued features: orientation, size, and curvature (below)

and irrelevant features, and the decrease was larger for the irrelevant feature (for the relevant feature, $M = 294, 281, 263,$ and 250 ms, $SD = 62, 74, 69,$ and $67,$ for the four blocks, respectively; for the irrelevant feature, $M = 283, 243, 223,$ and 204 ms, $SD = 56, 91, 66,$ and $66;$ see Fig. 8). As with Experiment 1, a 4×2 ANOVA was performed using within-subjects factors of block (1–4) and relevance (relevant for the task, irrelevant for the task). There were significant main effects of both block, $F(3, 117) = 32.39, p < .001,$ and relevance, $F(1, 39) = 12.63, p = .001,$ qualified by a

Table 1 Category distribution parameters

	μ_1	μ_2	μ_3	μ_4	σ^2	cov_{xy}
x	45.0	22.5	67.5	45.0	65	0
y	22.5	45.0	45.0	67.5	65	0

This was used to create stimuli for structure 1; structure 2 was created by rotating the resulting coordinates by $45^\circ.$ $\mu =$ mean of the coordinate; $\sigma^2 =$ variance of the bivariate distribution; $cov_{xy} =$ covariance between the two coordinates. Each category was restricted within one of the quadrants defined by the equations $y = x$ and $y = -x + 90.$ Values for irrelevant features were generated from a normal distribution ($\mu = 45$ and $\sigma = 18$)

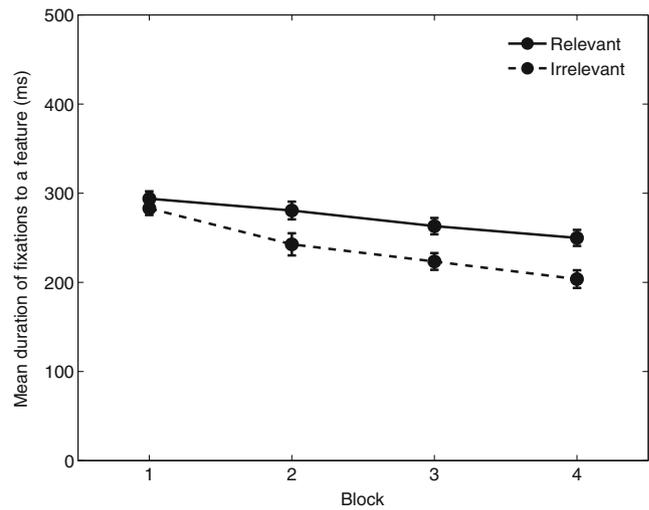


Fig. 8 Mean fixation durations to relevant and irrelevant features during stages of Experiment 2. Error bars represent the standard error of the mean

significant block \times relevance interaction, $F(3, 117) = 5.43, p = .002.$ A follow-up analysis showed that relevance had no significant simple main effect on the first block, $F(1, 117) = 0.70, p = .002,$ whereas the simple effects on all other levels of block were significant, $F(1, 117) = 144.77, 236.86,$ and $110.376,$ all $ps < .001.$ Durations decreased for both levels of relevance: There was a significant simple main effect of block for relevant, $F(3, 117) = 12.3, p < .001,$ and irrelevant, $F(3, 117) = 37.14, p < .001,$ features.

It is clear from these data that the influence of relevance on fixation durations is not limited to the category structure used in Experiment 1. The fixation durations are of similar magnitudes across experiments (~ 200 – 300 ms), and the differences between relevant and irrelevant fixations are likewise similar (~ 50 ms).

Within-trial temporal consistency across learning

As is shown in Fig. 9, mean STP increased for both relevant and irrelevant features over the blocks of the experiment (for relevant features, $M = .17, .24, .28,$ and $32, SD = .07, .13, .16,$ and $.16;$ for irrelevant features, $M = .15, .24, .28,$ and $37, SD = .07, .15, .15,$ and $.19).$ A 4×2 ANOVA using within-subjects factors of block and relevance was performed. Unfortunately, 15 (8 solving structure 1, and 6 solving structure 2) out of the 53 participants had to be excluded from the analysis for having undefined STP due to single-fixation trials. It appears that these participants were relying on a suboptimal single-dimensional strategy. There was a significant main effect of block, $F(3, 111) = 40.85, p < .001,$ but no significant effect of relevance, $F(1, 37) = 0.18, p = .67,$ and no significant block \times relevance interaction, $F(3, 111) = 2.25, p = .09.$ In this task, participants’ STP appears to increase over time equally for relevant and irrelevant features.

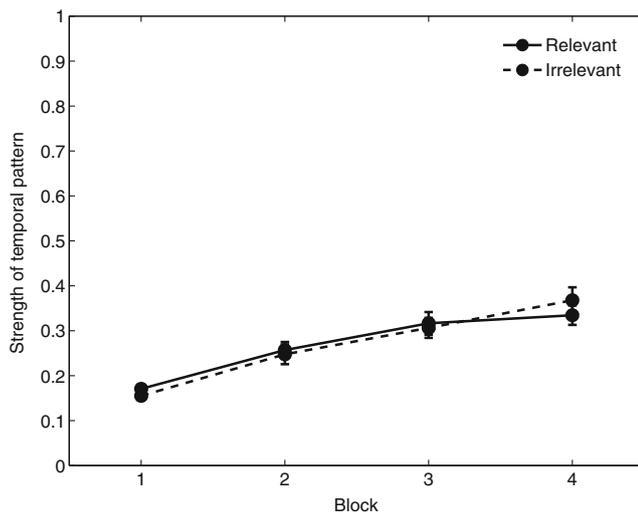


Fig. 9 Strength of the temporal pattern of relevant and irrelevant features during stages of Experiment 2. Error bars represent the standard error of the mean

Predicting response time and accuracy from attentional consistency

As in Experiment 1, participants with stronger temporal patterns made faster, but not more accurate, decisions. The STPs on the final block for relevant and irrelevant features were used as predictors in two multiple regression analyses predicting response time and accuracy. The former model was moderately predictive of response time (adjusted $R^2 = .32$), $F(2, 38) = 10.2$, $p < .001$, and coefficients for both relevant and irrelevant STPs were significantly lower than zero ($\beta = -.36$ and $-.35$), $t(39) = -2.5$ and -2.4 , respectively; $p = .02$ in both cases. The latter model was not predictive of accuracy, $F(2, 38) = 0.73$, $p = .49$, nor were relevant and irrelevant STPs significantly nonzero, $t(39) = -0.58$ and 1.2 , $p = .57$ and $.24$, respectively.

The fact that many participants resorted to the single-dimensional strategy indicates that this task was not a particularly efficient way of obtaining temporal pattern effects. Despite this shortcoming, we observed the emergence of clear temporal patterns in attentional behavior and replicated the finding that stronger temporal patterns were related to faster responses, but not to more accurate responses. The replication of most aspects of the findings from Experiment 1 demonstrates an impressive generality of results. Thus far, the key findings have been obtained with three different category structures of different complexities and sizes, with binary and continuous-valued stimulus features, with tasks that are readily learned perfectly, and with tasks in which perfect performance is difficult.

Unlike Experiment 1, Experiment 2 did not replicate differential STP increases for relevant and irrelevant features. One possible account is that temporal patterns develop

automatically as participants' experience with the task increases but that relevance differences develop only as participants gain mastery of the task. Indeed, accuracy was much lower in Experiment 2, suggesting that the latter condition was not met—whether due to task difficulty or perceptual difficulty. Because participants in Experiment 2 completed many trials, temporal patterns emerged. However, they did not completely master the categories; hence, the temporal patterns did not differ across relevant and irrelevant features.

Experiment 3

In the two experiments so far, we have examined gaze-based attentional behaviors. We hypothesize that those effects found in Experiment 1 and replicated in Experiment 2 reflect top-down attentional control directed by underlying decision mechanisms. However, it is possible that these effects might be peculiar to the oculomotor systems used to obtain information about the stimuli in these experiments. For instance, there is significant evidence that the nature of visual processing is time dependent (van Zoest, Hunt, & Kingstone, 2010) and, thus, regularities in fixation duration and STP might not replicate in a task where information is obtained in a slower and more deliberate manner. The interactions between bottom-up salience and top-down attentional control may change when different systems are responsible for obtaining information. Thus, in Experiment 3, we made stimulus features available through hand movements, rather than through eye movements alone, slowing the process of attentional allocation, to see whether the two findings generalize beyond simple oculomotor control. Given the unexpected prevalence of unidimensional strategies among participants in Experiment 2, we chose to use the category structure from Experiment 1.

Method

Participants

Seventy undergraduates participated in this experiment.

Stimuli and procedure

Stimuli, category structures, and procedure were the same as those used in Experiment 1, except that eye movements were not recorded. Each feature was completely obscured by an overlay at the start of each stimulus presentation. Participants used a mouse cursor to reveal features. The overlay was removed for as long as the participant kept the cursor over the feature. The total time for which the overlay was removed was recorded as a “fixation.” Overlays were not present during feedback.

Results and discussion

Four participants did not finish the experiment. As in Experiment 1, approximately 30 % of participants did not learn the categories: 21 participants were excluded from the analysis for failing to reach criterion. This left 45 participants for analysis. Their mean overall accuracy was .81 ($SD = .07$), with a mean of 76.2 ($SD = 39.6$) trials to reach learning criterion.

Duration of fixations to features across learning

Trials were divided into the same three blocks as those used in Experiment 1: (1) trials before reaching criterion, (2) the first 48 trials from criterion, and (3) the last 48 trials from criterion. Mean fixation duration (see Fig. 10) decreased over the experiment for the irrelevant feature, but not for the relevant features (for the relevant features, $M = 1,015, 1,059, \text{ and } 1,029 \text{ ms}$, $SD = 515, 435, \text{ and } 422$, for the three blocks, respectively; for the irrelevant feature, $M = 1,024, 677, \text{ and } 522 \text{ ms}$, $SD = 467, 317, \text{ and } 314$). A 3×2 ANOVA was performed using within-subjects factors of block (1–3) and relevance (relevant for the task, irrelevant for the task). There were significant main effects of block, $F(2,86) = 15.2, p < .001$, and relevance, $F(1, 43) = 31.85, p < .001$, qualified by a significant block \times relevance interaction, $F(2, 86) = 48.33, p < .001$. Once again, feature differences emerged only after criterion was reached: There were simple main effects of relevance for the second and third blocks, $F(1, 86) = 86.68 \text{ and } 152.16$, respectively; both $ps < .001$, but not for the first, $F(1, 86) = 0.052, p = .82$. While block had no significant effect on relevant features, $F(2, 86) = 0.68, p = .51$, time spent on the irrelevant feature decreased significantly, $F(2, 86) = 88.00, p < .001$.

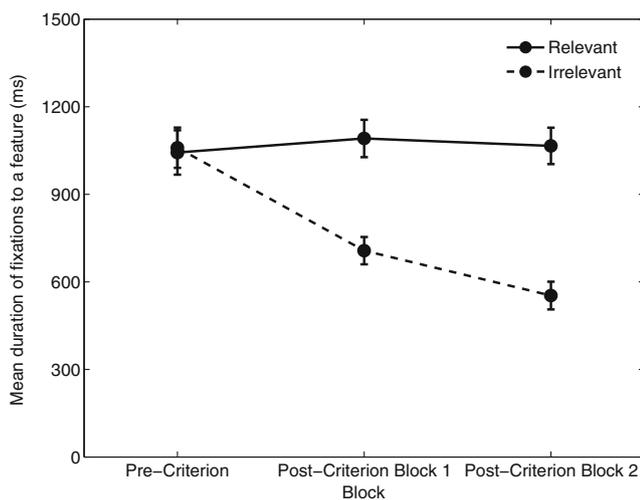


Fig. 10 Mean fixation durations of relevant and irrelevant features during stages of Experiment 3. Error bars reflect the standard error of the mean

These data replicate the finding that fixations are longer to relevant than to irrelevant features, now with a mouse-driven interface rather than gaze-based measures. Although fixations are substantially longer in the mouse-based experiment, the finding still holds and, indeed, appears to be stronger. In Experiments 1 and 2, fixation durations to relevant features were roughly 20 % longer than those to irrelevant features by the end of the experiment. In the present experiment, relevant fixations were nearly twice as long as irrelevant ones.

Within-trial temporal consistency across learning

Mean STP increased over the experiment for both the relevant and irrelevant features, and the increase was larger for the relevant features (for the relevant features, $M = .27, .58, \text{ and } .65$, $SD = .11, .17, \text{ and } .15$, for the three blocks, respectively; for the irrelevant feature, $M = .19, .35, \text{ and } .37$, $SD = .07, .16, \text{ and } .15$; see Fig. 11). A 3×2 ANOVA was performed using within-subjects factors of block and relevance. One participant was excluded from STP analyses due to having one block with undefined STP. There were significant main effects of block, $F(2, 86) = 119.91, p < .001$, and relevance, $F(1, 43) = 156.41, p < .001$. There was a significant block \times relevance interaction, $F(2, 86) = 25.58, p < .001$. As in Experiment 1, relevance had a significant simple main effect on each level of block, $F(1, 86) = 26.43, 214.22, \text{ and } 314.09$, all $ps < .001$, indicating that relevant features were accessed in a more stereotypical order than were irrelevant ones. However, STP increased over time for both types of features: Block had a significant simple main effect on both the relevant, $F(2, 86) = 202.28, p < .001$, and the irrelevant, $F(2, 86) = 50.56, p < .001$, features.

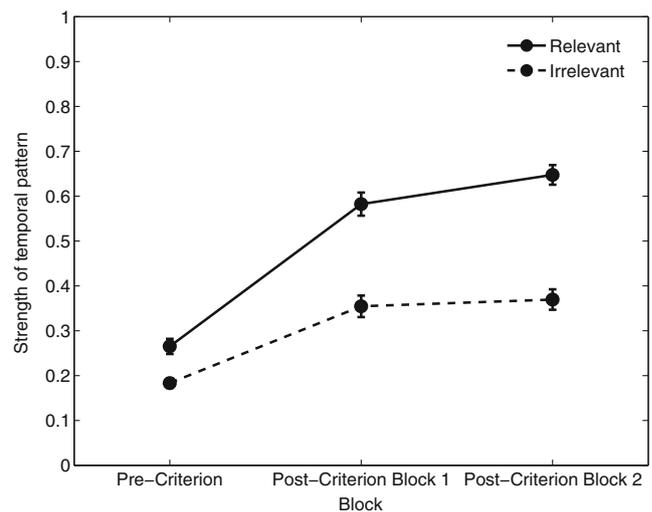


Fig. 11 Strength of the temporal pattern of the relevant and irrelevant features during stages of Experiment 3. Error bars reflect the standard error of the mean

Predicting response time and accuracy from attentional consistency

Replicating Experiments 1 and 2, the STP was associated with lower response times but was not predictive of accuracy. This conclusion stems from two multiple regression analyses predicting response time and accuracy.¹ Postcriterion STPs for relevant and irrelevant features were used as predictors in both. The former model was moderately predictive of response time (adjusted $R^2 = .23$), $F(2, 42) = 7.45$, $p = .002$. While the coefficient for relevant STP was significantly non-zero ($\beta = -.47$), $t(43) = -3.42$, $p = .001$, the coefficient for irrelevant STP was not, $t(43) = -0.76$, $p = .46$. Again, the latter model was not predictive of accuracy, $F(2, 41) = 0.88$, $p = .42$, nor were the coefficients for relevant and irrelevant STPs significantly different from 0, $t(42) = -0.44$ and 1.31 , $p = .67$ and $.20$, respectively.

Like the results for the fixation duration differences, the STP was noticeably larger in this experiment than in the gaze-based studies. The findings replicate the increase in the STP as the experiment progressed, but the relevant feature STP was far higher than in the previous experiments (.65, as compared with .38 and .32 in Experiments 1 and 2, respectively). This difference likely arises as a result of the increased cost of mouse movements: Participants may have been more deliberate in their temporal orderings when considering which features to consult.

General discussion

The present study is an investigation of overt, top-down attentional behavior in the context of category learning. The three experiments described here provide three primary novel findings. First, we found that the durations of individual gaze fixations were shorter when directed at irrelevant features than when directed at relevant ones. Second, we replicated the finding that participants readily adopt temporally structured patterns of attention (Experiment 1), with particular features being more likely to be fixated early or later in the buildup to the category decision. Experiment 2 extended this finding by showing that participants adopted temporal patterns even when the category structure did not encourage or reward them. Finally, in Experiment 3, we replicated our first experiment with a mouse-driven interface, rather than using an eyetracker. We replicated both the fixation duration differences and the temporal pattern strength, demonstrating that these effects are not exclusive to gaze measures and, instead, may reflect basic properties of overt attention.

¹ One data point was excluded specifically from this model due to its unusually high leverage (Cook's $d = 1.88$).

The first key finding of the experiments reported here is that fixations to irrelevant features were shorter in duration than those made to relevant features. Previous measures of attentional optimization have indicated that participants spend proportionately less trial time on irrelevant information (Blair, Watson, & Meier, 2009; Rehder & Hoffman, 2005a, b). Here, we show that this effect is attributable not only to overall fewer fixations to irrelevant information, but also to participants' ability to quickly orient attention away from unnecessary information once it has been fixated. This effect held not only when features were always irrelevant (Experiment 2), but also when features were irrelevant on only 50 % of the trials, whether participants were using eye (Experiment 1) or hand (Experiment 3) movements to obtain this information. Although the effect is similar to that found in reading (van der Schoot et al., 2008), we find it striking that these effects occur after minimal training in a novel task: Even by the second quarter of Experiment 2, the durations of fixations to relevant and irrelevant information began to diverge. This short training duration starkly contrasts with the thousands of hours of practice adult readers have had (e.g., Ribovich & Erickson, 1980). While our experiments did not evaluate whether these effects are due to the content of irrelevant features, rather than their spatial locations, the analogous effect in reading suggests that the former may be the case. These small but significant differences demonstrate that an understanding of feature relevance affects visual processing on short time scales and provide further empirical incentive to investigate categorization at a finer-grained level than is typical in the field.

Our second key finding, that participants develop stereotypical temporal patterns of attention shifting, complements existing studies on attentional strategies (Blair, Watson, Walshe, et al., 2009; Rehder & Hoffman, 2005b). Our STP measure provided a quantitative method of assessing the degree to which fixations are temporally structured and allowed us not only to confirm indications of temporal patterns in previous work (cf. Rehder & Hoffman, 2005a), but also to show that STP in category structures without any incentive for temporal order is comparable to that in those that do provide such incentive. Using STP also allowed us to discover that participants using more consistent oculomotor patterns for obtaining information made faster responses. Motor-learning processes may underlie these orderly overt attentional behaviors (e.g., Graybiel, 1998), and the shorter reaction times correlated with ordered fixation patterns provide an ample advantage, even if such temporal regularities do not increase accuracy. These findings are consistent with investigations of scanpaths in free-viewing, recognition, and visual search, which show that participants tend to develop stereotypical scanpaths and that these are associated with shorter search times (Myers & Gray, 2010), but not with greater recognition accuracy (Humphrey & Underwood, 2010).

Our final key result is the replication of the gaze-based findings from Experiments 1 and 2—relevance-dependent fixation duration differences and the development of temporal fixation patterns—in a study with a mouse-based interface. This confirmed that these findings are not due strictly to oculomotor processes. Although it is difficult to directly compare, it is also notable that the magnitude of these effects was larger in the mouse-based experiment than in the two gaze-based ones. It seems plausible that decisions about which features to consult in the mouse-driven interface are more deliberate and, thus, less affected by feature saliency. In essence, noise introduced by the oculomotor system is filtered by the top-down control of attention through hand movements. This allows for less stochasticity in attentional behaviors and, thus, stronger effects. These quantitative differences aside, the general qualitative results are the same as in Experiments 1 and 2, which offers confirmatory evidence that experiments using mouse-driven interfaces to study decisions in categorization (e.g., Matsuka & Corter, 2008; Nelson et al., 2010) study the same processes as gaze-based interfaces, only on a slightly different time scale. This allows researchers to combine inferences made through both avenues of research and apply them to decision processes more generally.

Although our primary interest is in the overt allocation of attention, we note that across our three experiments, participants developed a systematic pattern of accessing stimulus features that is consistent with a serial rule-based approach to categorization (e.g., Bradmetz & Mathy, 2008; Fific, Little, & Nosofsky, 2010; Trabasso, Rollins, & Shaughnessy, 1971). The spatially separated features of our stimuli make this difficult to interpret, however. Gathering information by foveating stimulus features is a necessarily serial activity, and the act of planning fixation sequences likely encourages rule-based strategies. Furthermore, it is not clear how tightly attentional behavior reflects the underlying decision processes. For instance, our data do not rule out a serial, rule-based system governing top-down attention combined with a nonserial, exemplar-based decision processes. There may also be alternative accounts of the attention data. For example, it may be that attention is optimized by reinforcement learning mechanisms, as has been postulated for eye movements in reading (Reichle & Laurent, 2006), leading to the efficient allocation of attention seen in the present experiments. Without a great deal more data, we are cautious in interpreting our results as support for a rule-based model of categorization or, indeed, for any other varieties. These data are intended to provide a basis for future models that include dynamic attention shifts, rather than to settle existing disputes.

Categorization is an excellent paradigm for studying overt attentional behaviors. It captures a slice of cognitive processing that incorporates perceptual, motor, and

decision-making processes in a well-controlled yet ecologically relevant task. Looking further into the future, a model of eye movements in categorization may be profitably thought of as a special case of a more broadly applicable theory of information use and decision making. Variables such as the cost of obtaining new information may be important in these broader situations. Wood, Fry, and Blair (2010), for instance, manipulated information costs in a categorization task by imposing a time delay for information revealed by a mouse click and found that the condition with a longer delay yielded more optimal patterns of attention and more rapid learning. Although the costs of eye movements are relatively small, in many circumstances—such as in medical diagnosis—obtaining information incurs significant costs in time and money. A model that captures the properties of eye movements will likely have the right structure to be extended to account for overt attentional behaviors in a broad range of real-world cognitive activities (e.g., Hayhoe & Ballard, 2005).

Timing is a critical component in cognitive processes, especially visual processing (e.g., van Zoest et al., 2010). As it currently stands, however, not one of the major models of categorization (Kruschke, 1992; Kruschke & Johansen, 1999; Love et al., 2004; Nosofsky, 1986) can accurately predict overt attentional behavior. To be fair, this is because these models were not intended to do so. We very much doubt that any of these researchers have a strong theoretical commitment to attention as a static bias on individual stimulus dimensions that is updated only after feedback has been given, but this is effectively what the models incorporate. Given how important selective attention has been to theories of categorization and how much we know about the dynamics of attentional allocation, this limitation has lingered for too long. New models of categorization should move in the direction of incorporating a more dynamic within-trial understanding of human learning. Much as the order in which a person is exposed to category members (e.g., Mathy & Feldman, 2009; Medin & Bettger, 1994) can affect performance, the manner in which a person attends to information about these members is likely to have an impact on learning.

Humans are not passive learners; by breaking down and studying the finer-grained attentional processes that occur during categorization, we hope to emphasize the active role people play in the overt attentional behaviors they use. We began by arguing that the structure of the models in some ways constrains the experimental paradigms, but we are encouraged to see an increasing body of work designed to specifically address the active processes of learners (e.g., Castro et al., 2008; Markant & Gureckis, 2010). We consider the findings presented here to be part of a growing empirical foundation that is a necessary first step for the extension of extant theories of category learning and the development of new ones.

Author note The authors would like to thank the Cognitive Science Lab team who contributed in numerous ways to this research. This research was supported by grants from the Canada Foundation for Innovation and the Natural Sciences and Engineering Research Council of Canada to Mark R. Blair.

References

- Awh, E., Anillo-Vento, L., & Hillyard, S. A. (2000). The role of spatial selective attention in working memory for locations: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *12*, 840–847.
- Blair, M. R., Watson, M. R., & Meier, K. M. (2009a). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*, 330–336.
- Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009b). Extremely selective attention: Eyetracking studies on dynamic attentional allocation to stimulus features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1196–1206.
- Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression time in Boolean concept use. *Psychological Research*, *72*, 211–234.
- Castro, R. M., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2008). Human active learning. In D. Koller, Y. Bengio, D. Schuurmans, L. Bottou, & A. Culotta (Eds.), *Advances in neural information processing systems*, *21* (pp. 241–248). Cambridge, MA: MIT Press.
- Colner, B., & Rehder, B. (2009). A new theory of classification and feature inference learning: An exemplar fragment model. In N. Taatgen, H. van Rijn, L. Schomaker, & J. Nerbonne (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 371–376). Mahwah, NJ: Erlbaum.
- Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, *117*, 309–348.
- Graybiel, A. M. (1998). The basal ganglia and chunking of action responses. *Neurobiology of Learning and Memory*, *70*, 119–136.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*, 188–194.
- Humphrey, K., & Underwood, G. (2010). The potency of people in pictures: Evidence from sequences of eye fixations. *Journal of Vision*, *10*, 1–10.
- Kaakinen, J. K., Hyona, J., & Keenan, J. M. (2002). Perspective effects on online text processing. *Discourse Processes*, *33*, 159–173.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.
- Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1083–1119.
- Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associated blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 830–845.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.
- Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category-learning systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 227–245.
- Markant, D., & Gureckis, T. M. (2010). Category learning through active sampling. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 248–253). Austin TX: Cognitive Science Society.
- Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, *16*, 1050–1057.
- Matsuka, T., & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology*, *61*, 1067–1097.
- Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic Bulletin & Review*, *1*, 250–254.
- Myers, C. W., & Gray, W. D. (2010). Visual scan adaptation during repeated visual search. *Journal of Vision*, *10*, 1–14.
- Nelson, J. D., McKenzie, C. R. M., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science*, *21*, 960–969.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*, 929–942.
- Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta Psychologica*, *86*, 227–272.
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something category learning results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 811–829.
- Reichle, E. D., & Laurent, P. (2006). Using reinforcement learning to understand the emergence of “intelligent” eye-movement behavior during reading. *Psychological Review*, *113*, 390–408.
- Ribovich, J. K., & Erickson, L. (1980). A study of lifelong reading with implications for instructional programs. *Journal of Reading*, *24*, 20–26.
- Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eyetracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*, *1*, 71–78.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, (13, Whole No. 517).
- Tipper, S. P., & Cranston, M. C. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *37*, 591–611.
- Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology*, *2*, 239–289.
- van der Schoot, M., Vasbinder, A. L., Horsley, T. M., & van Lieshout, E. C. D. M. (2008). The role of two reading strategies in text comprehension: An eye fixation study in primary school children. *Journal of Research in Reading*, *31*, 203–223.
- van Zoest, W., Hunt, A. R., & Kingstone, A. (2010). Visual representations in cognition: It's about time. *Current Directions in Psychological Science*, *19*, 116–120.
- Watson, M. R., & Blair, M. R. (2008). Attentional allocation during feedback: Eyetracking adventures on the other side of the response. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 345–350). Austin, TX: Cognitive Science Society.
- Wood, M. J., Fry, M. D., & Blair, M. R. (2010). The price is right: A high information access cost facilitates category learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 236–241). Austin TX: Cognitive Science Society.
- Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition*, *35*, 1380–1398.