During category learning, top-down and bottom-up processes battle for control of the eyes

Caitlyn M. McColeman¹ (caitlyn_mccoleman@sfu.ca)  
Mark R. Blair² (mark_blair@sfu.ca)  
Department of Psychology¹,² & Cognitive Science Program², Simon Fraser University  
8888 University Drive, Burnaby, BC V5A 1S6

Abstract

Information in the visual environment is largely accessed through a series of fixations punctuated by saccades. Changes in fixation patterns in response to learning are well documented in studies of categorization, but the properties of the saccades that precede them and the role of visual salience in effecting eye movements remains poorly understood. This eye tracking study examines oculomotor changes in a categorization task with salient distractors. The design examines high-level, goal-directed attention that serves the purpose of learning, and making decisions based on that learned knowledge in the presence of salient distractors. We find that salient distractors draw fixation durations and saccade velocities that display similar properties to eye movements directed to task relevant items, challenging existing accounts that salience draws rapid saccades.

Keywords: Saccades; eye movements; categorization; visual attention; salience; learning.

Vision is critical for many life forms. Input processed by the visual system can be used to interact with an organism’s surroundings, informing what dangers might be present or, depending on the animal in question, whether there is a coffee shop ahead. The relative contribution of exogenous salience and observers’ goals in driving the eye movements that gate input to the visual system remains contested. Eye movements are closely linked to attention, which is thought to be captured by salient items (Theeuwes, 2004; 2010). However, the goals of the observer can override capture (Leber & Egeth, 2006). Understanding the impact of salience is further complicated by an apparently disparate influence of stimulus properties (e.g. colour vs. shading) reported to influence visual search performance, implicating the physical manifestation of salient items as a factor in explorations of salience and attention (Wolfe & Horowitz, 2004). The present study explores the influence of salient distractors on attention during learning. Understanding how visual information is obtained and subsequently processed is important to understanding cognition more broadly.

When viewing a static scene, eyes are usually either resting on a stimulus (a fixation) or moving to a new location (a saccade). Little visual input is processed during saccades (Higgins & Rayner, 2014; Matin, 1974) so fixations play a particularly important role in visual processing.

The locations of fixations are determined in part by low level properties of the objects in the scene. Areas of rich color, high brightness, or high contrast are salient, and thus get fixated more often (Parkhurst, Law & Nieber, 2002) These findings have been incorporated into algorithms (e.g. Itti, Koch & Nieber, 1998) which calculate salience across a scene based on properties such as colour, intensity and orientation, and have been shown to predict fixations (Judd, Ehinger, Durand & Torralba, 2009; Ouerhani, von Wartburg, Hügli & Müri, 2004). Although salience is an important factor in predicting where fixations occur, it is not the whole story (Tatler, Hayhoe, Land & Ballard, 2011).

Fixations can also be influenced by the task at hand. Yarbus (1967) showed that participants freely viewing a painting displayed a fixation pattern that differed from fixation patterns to the same images with instructions to look for particular information in the environment by instructions such as “estimate the material circumstance of the family [in the painting]”. Another study showed that when participants were asked to find a subset of a target image, much like placing a piece in a mostly completed puzzle, they used the information from the subset to guide their search (Pomplun, 2006). The flexibility of the visual system to seek information in support of task objectives shows that goal-directed attention is also an important contributor in propelling oculomotor activity.

For example, when eye movements are recorded during a naturalistic task like making tea, fixations tend to precede larger motor movements such as reaching (Land, Mennie & Rusted, 1999) and during driving, better drivers display different eye movement patterns than novices (Underwood, Chapman, Brocklehurst, Underwood & Crundall, 2014). These findings link the global task property of expertise and task knowledge to the low level execution of eye movements. Since fixations are required to extract and subsequently process richer information, tasks can be designed to purposely enact demands on the visual system such that researchers may infer what parts of the environment are prioritized by the participant.

The changing patterns of fixations during learning are informative about what the participant knows, and the stage of knowledge they are in as the efficiency of information access increases. For instance, in a category learning task where spatially separated cues have different diagnostic value, fixations uninformative to making a category decision decline in duration and frequency relative to informative cues as learning progresses (Blair, Watson & Meier, 2009). Category learning has been an effective
source of uncovering when, and inferring why fixation patterns change (Blair, Watson, Walshe & Maj, 2009; McColeman, Barnes, Chen, Meier, Walshe & Blair, 2014; Rehder & Hoffman, 2005), but fixations are only one element of oculomotor activity.

Saccades, too, are influenced by a number of features of the task environment. For example, participants in an anti-saccade task are asked to fixate away from the location of a visual cue, the reverse of a pro-saccade task where participants fixate toward the location of a visual cue. Thus, a cue might appear in the mirror location of where the participant is asked to fixate. Researchers have found that saccades directed to the cue in the pro-saccade task are faster than saccades directed to the mirror location opposite the cue in the anti-saccade task (Fischer & Weber, 1992) and erroneous eye movements - moving toward the cue in the anti-saccade task, for example - are executed more quickly (Tatler & Hutton, 2007). In a similar vein, stimulus-driven saccades are faster than goal-directed ones (van Zoest & Donk, 2007). Still other research suggests that saccades to salient distracting items are faster (as measured by shorter latency) than saccades directed to a target in a search array (van Zoest, Donk & Theeuwes, 2004). These studies look at eye-movements in simple tasks wherein the goal of the participant remains constant and so the responsiveness of these effects to changing task knowledge remains unknown.

The present study aims to explore the role of salient versus non-salient features when they serve as distractors in a learning task, effectually pitting bottom-up attention and top-down attention against each other. The experimental task is category learning: participants learn about the relevance of different features for making a category response through trial and error. The experiment design allows for insight as to how the role of bottom-up attention changes over time, and points during learning wherein top-down attention takes precedence over stimulus-driven attentional capture. As with earlier work in categorization, eye movements are used as an index of attention patterns. The fixations and the saccades are both of interest here, in that both oculomotor measures are necessary for moving about the visual environment and accessing information. Examining both the fixation and the saccade is a step toward developing a fuller understanding of the visual system in the context of learning.

Methods

Participants were asked to sort images of sham alien cells into four possible categories. The information necessary for making the category choice is communicated by features that vary trial-to-trial (Figure 1) and are described to the participants as organelles of the alien cells during the instructions which ask them to sort the images into categories.

Participants were randomly assigned to one of three conditions: One Salient Distractor (n=77), Two Salient Distractor (n=68) and Baseline without salient distractors (n=65). All participants in the study received partial course credit from the Simon Fraser University Research Participation Pool. Those who failed to complete the 320 trials in the task are excluded from analysis (20 from Baseline, 23 from Two Salient Distractor and 22 from One Salient Distractor conditions respectively). Of primary interest is in how salient distractors affect oculomotor measures as participants learn, so we restrict our analysis to participants who achieved 16 correct trials in a row. Twenty-five participants from the Baseline Condition, 20 from the Two Salient Distractor Condition and 22 from the One Distractor Condition were excluded. The remaining participants all met a pre-specified gaze criterion of >70% of trials with >75% gaze data collected.

Stimuli and Category Structure

The features can take on one of two possible values during each trial and are pasted on a random selection of rotations of the alien cell background (Figure 1). For any one participant, features are pasted to the same location trial-by-trial, and the location of the features is randomly assigned to four possible locations between participants. Each feature subtends approximately 1.15° and is separated by 12.4°. The combination of Features 1 and 2 is sufficient to correctly respond with a category label on all trials (Table 1), and thus Features 1 and 2 are considered relevant. Participants learn through trial and error which are relevant features and the combination of feature values that correspond to the four category labels.

Procedure

Participants’ eye movements are recorded using a desk-mounted Tobii X120 eye tracker sampling with a temporal resolution of 120Hz and a spatial resolution of .5° of visual angle. Fixations are calculated using the raw gaze data and a modified version of Salvucci and Goldberg’s dispersion threshold algorithm (2000) with a temporal threshold of 75ms and a spatial threshold of 1.1° such that a fixation is a minimum of 75ms long and consists of gaze points within

Figure 1. The stimuli used in the experiments. The features (right) could take on two possible values. They were pasted on the stimulus background (left). The top four pairs are the possible relevant items and non-salient distractors, and the bottom two pairs are the possible salient distractors.
Results

At the level of overt response, there is no influence of the number of salient distractors as reflected by accuracy, reaction time or the probability of fixating irrelevant information. That is, participants are able to perform the task and are unaffected by salient distractors at this level of analysis. Accuracy data were binned into blocks, where one block is the average of 20 trials. Condition is a between subjects effect, defined by the number of salient distractors (0, 2 or 1 salient distractors). There is a main effect of Block in the accuracy (Figure 2) indicated by an ANOVA ($F_{15,1125}=132.54, p<0.0001$, Greenhouse-Geisser correction) and no noticeable effect of Condition ($F_{2,75}=3.8, p=0.068$) or the interaction between Condition and Block ($F_{30,1125}=1.07, p=0.36$). Accuracy clearly improves over learning, and is not noticeably influenced by the number of salient distractors.

To see if the number of salient distractors affected how quickly people decided to make their category response, an ANOVA was conducted on reaction time, using Condition and Block as above. There is no main effect of Condition ($F_{2,97}=0.27, p=0.97$) or interaction between Block and Condition ($F_{30,1455}=0.527, p=0.988$), but there is a main effect of Block ($F_{15,1455}=47.578, p<0.0001$) indicating that response times are faster as the experiment progresses.

The probability of fixating an irrelevant feature measure acts as a coarse descriptor of the oculomotor efficiency during a trial by capturing whether a participant is accessing irrelevant information (Figure 2). As learning progresses and the participant is aware of which of the features in the stimulus are important to making a category decision, it is unnecessary to fixate irrelevant items to extract information from them. In this experiment’s particular design, with two irrelevant features (Table 1), it is possible to look at half of the irrelevant features during a trial. For this reason, it is possible to obtain a score of 0.5 on this measure on a trial by looking at only one of the two irrelevant items.

Again, an ANOVA is conducted with Condition and Block as factors to investigate their influence on the probability of fixating irrelevant features and there is a main effect of Block ($F_{15,1125}=109.96, p<0.0001$, Greenhouse-Geisser correction) finding no noticeable effect of Condition ($F_{2,75}=0.4, p=0.91$) or interaction between Condition and Block ($F_{30,1125}=0.75, p=0.84$). That is, in none of the three conditions were participants more likely to look at the irrelevant features during learning. Overall, there is no distinguishable performance difference between the three groups. Accuracy, reaction times, and the probability of fixating irrelevant information all show a similar change over learning where there is no detectable effect of condition. This result is consistent with work that suggests that task goals can eliminate salience effects (Tatler, Hayhoe, Land & Ballard, 2011). Still, salience may yet influence more basic ocular motor measures.

To investigate this, the overall dwell time on the four features is reported as the fixation duration (Figure 3) and is meant to uncover how much fixation time is allocated to the four features of interest. That is, if there are four fixations to feature 3 in a trial, then this measure reports the sum of the durations of those four fixations. To uncover the relative role of repeated exposures to stimuli on a trial-by-trial basis, the fixation duration is estimated using a linear mixed effects model built with trial, feature relevance and their interactions as possible predictors. Estimates of the predictiveness of feature relevance use Feature 1 as a reference category.

In the baseline condition (Figure 3A), with no particularly salient items, there were main effects of Feature 3 and 4 ($p < 0.28$, ps = 0.0002). There is also an interaction between Trial and Feature 3 ($p = 0.251$, ps = 0.00024) while the remaining factors play no discernible role in motivating the fixation duration. Generally, then, there were decreased fixation durations to irrelevant items relative to Features 1 and 2.

In the condition with two salient distractors (Figure 3B),
there were main effects of Features 2, 3, and 4 ($\beta$s $<$ -31, $p$ $<$ .00001) Trial ($\beta$ = .0659, $p$ = .0047), and interactions between Trial and Feature 2 ($\beta$ = -.18426, $p$ = .00024) and Trial and Feature 3 ($\beta$ = -1.743, $p$ = .00017) with the remaining predictors having minor significance in affecting fixation duration as predicted by the model. Here again, irrelevant items were fixated less overall.

In the condition with a single salient distractor, Feature 4 is coded as the salient feature. There is a main effect of Feature 3 ($\beta$ = 14.778, $p$ = .0132), Trial ($\beta$ = -.1367, $p$ < .0001), and an interaction between Trial and Feature 2 ($\beta$ = 0.716, $p$ = .0190) while the remaining predictors do not play a significant role in predicting fixation duration. In all three conditions, less time is spent fixating irrelevant information.

A second series of models was constructed to explore the role of feature salience among other factors of interest in predicting saccade velocities. The data subject to analysis are eye movements directed to four features, and so fixations that terminate elsewhere are excluded from the dataset prior to analysis. That is, a saccade that begins in the centre of the screen and ends with a fixation to Feature 1 is viable, while a fixation beginning on Feature 1 and moving to the centre of the screen is excluded. All saccades included in this analysis end with fixations that are described above, such that the overall number of saccades decreased over the course of the experiment.

In these models, saccade velocity is predicted by trial number, feature relevance, and the distance of the saccade subjected to a logarithm transform to meet assumptions of the model. Trial number captures the change in saccade velocities over time, feature relevance indicates the role of the individual features in predicting saccade velocity and also captures the influence of salience, since salient distractors are always irrelevant when they are present in this experiment design. The distance of the saccade in degrees of visual angle is reported as well, since the distance of saccades is known to play an important role in the characteristics of saccades (Balo, Konrad, Sills & Honrubia, 1975).

In the baseline condition where there are no particularly salient distractors (Figure 4A), Distance was the best predictor of saccade velocity ($\beta$ = 0.8038, $p$ < .0001) and Trial is also a strong predictor ($\beta_{\text{Trial}}$ = -.0003, $p$ < .0001). Feature 3 shows a main effect ($\beta$ = -.05689, $p$ = .0092) and the interaction between Trial and Feature 3 is significant ($\beta$ = -.0009, $p$ = .0002). The interaction between Trial and Distance is notable ($\beta$ = .0001, $p$ = .00002) as is the interaction between Feature 2 and Distance ($\beta$ = -.0300, $p$ = .00089). The three-way interactions between each of Features 3 and 4 with Distance and Trial ($\beta$s $>$ -.0003, $p$s $<$ .0068) are also significant predictors. The model reflects a complicated, interacting relationship between the progress in the experiment, feature relevance and distance. Saccades spanning a larger distance are faster, as are saccades to irrelevant features.

In the two salient distractors condition, features 3 and 4 are both irrelevant and salient (Figure 4B). There is an effect of Distance ($\beta$ = .8522, $p$ < .0001) and Trial ($\beta$ = -.0002, $p$ = .0048). The interaction between Distance and Trial is predictive ($\beta$ = .00015, $p$ = .00001). The interaction between Feature 2 and Distance is a minor but notable contributor ($\beta$ = -.0213, $p$ = .0161). This condition reflects little by way different effects for the four features, in that the saccade velocities are similar regardless of feature relevance - and, indirectly, feature salience.

Saccade velocities from participants in the single salient distractor condition (Figure 4C) are analyzed using the same model format as above. In this condition, the single salient distractor is called Feature 4, but is as equally irrelevant to completing the task as Feature 3 (Table 1). Here again, distance is best predicted by distance ($\beta$ = .8550, $p$ < .0001) where longer saccades correspond with a higher velocity. There is also a main effect of trial number ($\beta$ = -.0002, $p$ = .0106), and Feature 3 ($\beta$ = -.0815, $p$ = .0001). The interaction between Trial and Feature 3 ($\beta$ = .0004, $p$ = .0027), and the interaction between Feature 2 and distance ($\beta$ = -.030, $p$ = .0014) are both significant, while the remaining predictors having minor significance in affecting saccade velocity. Failure to find a main effect of- or interactions involving Feature 4 (the salient item) here is attributed to high variability in the saccade velocity data for eye movements directed to it.

Analyses reflect a clear difference in saccade velocities between relevant and irrelevant items in the baseline condition where there are no salient distractors. This is not so in the presence of salient distractors, reflecting a
mediating influence of salience in the execution of task-relevant saccades.

**Discussion**

This study pits high-level, goal directed attention against bottom-up attention by manipulating the salience of irrelevant features. We find that task relevance takes precedence insofar as information seeking and response-level efficiency is concerned: participants are no more likely to look at irrelevant information in the presence of salient distractors and regardless of the number of salient distractors, overall accuracy and reaction times improve at the same rate over the course of the experiment. Salience does, however, affect the lower level properties of eye movements by way of saccade velocity and, to a lesser extent, fixation duration.

By beginning to uncover properties of both the fixation and the saccade together, we are moving toward a fuller understanding of oculomotor changes in response to learning. The fixation, thought to be more closely related to cognitive processing, is analyzed through the probability of fixating irrelevant information (Figure 2) and the fixation durations (Figure 3). Participants made fewer fixations to the irrelevant features, regardless of whether those features were salient. Relevant items are typically fixated longer, although to a lesser extent when the irrelevant items were salient.

Generally, the eyes travel to relevant, informative features more slowly than they do to irrelevant distractors as is indicated by the saccade velocity. This is evident as learning progresses: the eyes travel to irrelevant items more quickly, reflecting a responsiveness of low-level oculomotor activity to shifts in high level cognitive changes, since eye movements to the irrelevant features are comparatively fast as learning progresses. The increase in saccade speed to irrelevant features is especially true in the baseline condition (Figure 4A), where the distractors are of similar salience to the relevant features. When there are two salient distractors (Figure 4B), the velocities directed to irrelevant items differ less - it’s not that the velocities to relevant features are slower in this case relative to saccade speeds in the baseline condition as much as it’s due to decreased saccade velocities to the irrelevant items in the condition where those irrelevant items are salient. The condition with a single distractor is particularly noteworthy in that one irrelevant feature is salient and one irrelevant feature is not, and the saccade velocities to each differ as a function of the irrelevant items’ salience (Figure 4C).

The rapidity of saccades is clearly affected by both salience and learned task relevance, and surprisingly the direction of the influence of salience is opposite what existing reports suggest: salient items draw slower saccades than less salient items of equal task importance. Other work investigating the role of salience in saccade velocities suggests that goal-directed eye movements are typically slower than stimulus-driven (or salience responsive) eye movements (van Zoest, Donk & Theeuwes, 2004; Xu-Wilson, Zee & Shadmehr, 2009). When the saccade itself is rewarded, an interesting divergence in contributing factors can be noted. Slow, long latency saccades are attributed to rewarded saccades and faster, short latency saccades are linked to salience (Schütz, Trommershäuser, Gegenfurtner, Wilson, Zee & Shadmehr, 2009). When the saccade itself is rewarded, an interesting divergence in contributing factors can be noted. Slow, long latency saccades are attributed to rewarded saccades and faster, short latency saccades are linked to salience (Schütz, Trommershäuser, Gegenfurtner, Wilson, Zee & Shadmehr, 2009). When the saccade itself is rewarded, an interesting divergence in contributing factors can be noted. Slow, long latency saccades are attributed to rewarded saccades and faster, short latency saccades are linked to salience (Schütz, Trommershäuser, Gegenfurtner, Wilson, Zee & Shadmehr, 2009).

These results may seem incongruent with previous work. A possible factor in mediating eye movements such that purposeful saccades travel more slowly is volition. Our data can align well with earlier reports. Considering work in visual search environments, where search is informed by visual properties of prior interest to the observer (Wolfe, 1989), the three conditions can be interpreted as three different types of search environments. In the baseline condition, there is not a clear exclusion property in the same way that there is when there are two salient distractors; in the latter condition the participant may use luminosity to exclude two of the features from contention as a target for the next saccade. That is, reflexive, loosely controlled saccades may be less likely when there is a filter or conjunction search influencing viable regions of interest for subsequent fixations. If this is the case, then saccades made to these salient distractors are more likely to be volitional. Alternatively, the salient distractors may simply be interesting in their own right due to their conspicuity and not fitting in with the rest of environment, which also increases the likelihood of volitional saccades relative to reflexive saccades directed to salient targets.

Exploring the relative influence of goal-directed and stimulus-driven attention during learning is uniquely informative in that goal-directedness develops over repeated exposure to stimuli. Through changes over time it’s possible to see the dynamic relationship between oculomotor activity and higher level performance such as response accuracy. In
this task, the influence of salient items is not clear until some degree of category mastery is achieved in that the three conditions show similar saccade velocities to all four features until the 7th Block (approximately 140 trials). This suggests that achieving the goal of the task - learning the category structure - is imperative for participants and generally, that high level goals can drastically affect low level motor execution.

The main contributions of this work are threefold. We gather additional insight into the role of salience in high-level, cognitive task performance, adding to the knowledge base of bottom-up versus top-down attention with data that challenge basic assumptions of the primacy of bottom-up attention in driving saccades. Methodologically, exploring purposeful manipulations of the environment’s properties over learning provides a foundation for insight into how developing top-down knowledge dynamically responds to task-irrelevant properties of the domain. Finally, exploring the profile of a saccade in a high level task of this nature is also novel, since most work - at least in category learning - focuses on the fixations to probe attentional changes in the task. Moving forward, designs of this nature are valuable for informing how the cognitive system uses information in the environment, and cases in which the properties of the environment might help or hinder an agent’s ability to most effectively use the information available in its surroundings.

Acknowledgements
Thank you to the members of the Cognitive Science Lab who provided support and suggestions during this project. This work was funded by the National Science and Engineering Research Council.

References