Automatic selection of eye tracking variables in visual categorization for adults and infants

Samuel Rivera [1], **Catherine A. Best**[2], **Hyungwook Yim**[2], **Aleix M. Martinez**[1], **Vladimir M. Sloutsky**[2], **Dirk B. Walther**[2]

[1]Department of Electrical and Computer Engineering, 205 Dreese Lab, 2015 Neil Avenue, Columbus, OH 43210

[2]Department of Psychology, 225 Psychology Building, 1835 Neil Avenue, Columbus, OH 43210

Abstract

Visual categorization and learning of visual categories exhibit early onset, however the underlying mechanisms of early categorization are not well understood. The main limiting factor for examining these mechanisms is the limited duration of infant cooperation (10-15 minutes), which leaves little room for multiple test trials. With its tight link to visual attention, eye tracking is a promising method for getting access to the mechanisms of category learning. But how should researchers decide which aspects of the rich eye tracking data to focus on? To date, eye tracking variables are generally handpicked, often resulting in biases of the eye tracking data. Here, we propose an automated method for selecting eye tracking variables based on their usefulness to discriminate learners from non-learners of visual categories. We presented infants and adults with a category learning task and tracked their eye movements. We then extracted an over-complete set of eye tracking variables encompassing durations, probabilities, latencies, and the order of fixations and saccadic eye movements. We applied the statistical technique of ANOVA ranking to identifying those variables among this large set that covaried with the learner/non-learner label. We were able to identify learners and non-learners above chance level using linear SVM and the top eye tracking variables.

Keywords: eye tracking; category learning.

Introduction

Categorization is the process of forming an equivalence class, such that discriminable entities elicit a common representation and/or a common response. This task has two components (a) category learning (i.e., forming a common representation for a set of items and/or learning a common response) and (b) categorization or classification (i.e., partitioning a stimulus set using this common representation). While category learning exhibits an early onset (Quinn, Eimas, & Rosenkrantz, 1993), relatively little is known about the underlying mechanism and the development of early categorization. One of the reasons is the limited duration of typical experimental paradigms with infants, yielding only a small number of data points per participant. These limitations have restricted researchers ability to answer fundamental questions about categorization in infants: How do infants learn a category? And what changes in the course of development?

Given that eye movements are tightly linked to visual attention (see Rayner, 1998, for a review), eye movements can provide critical information of how attention allocation changes during category learning. However, eye tracking yields a large amount of data, and it is usually not clear from the outset how to make sense of these data. These data could be converted into numerous variables, and there is no principled way of deciding which of these variables should be selected and why. The *key idea* of our approach is to develop algo-
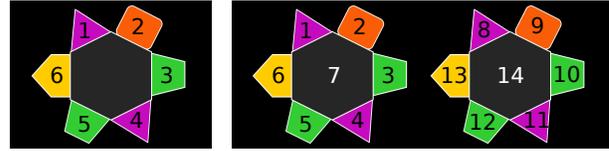
Figure 1: Image (a) is an example category object used in the eye tracking study, with the Areas of interest (AOI)s enumerated. Image (b) defines the AOIs for a pair of objects. The numbers were not shown to participants.

rithms that *automatically* select the eye tracking variables that are systematically linked to category learning.

Our approach was as follows. We extracted a large set of possible variables from the adult or infant gaze sequence during a categorization task (e.g. fixations, saccades, gaze sequences, etc). Some of the variables have been used in analyzing categorization experiments, whereas others are new. We used ANOVA to identify the eye tracking variables that best predict category learning in adults and subsequently in infants, and validated the variables using linear SVM. The significant contribution of this work is that it provides a methodology for identifying eye-tracking variables that are linked to category learning, thus allowing researchers to better understand categorization.

Method

Category Object

The category members were flower-like objects with six petals. An example category object is shown in Fig. 1(a), with the petals enumerated for clarity. The four different categories used were defined by a single petal having a distinguishing color and shape. Specifically, the category defining features were A (pink triangle at AOI 4), B (blue semi-circle at AOI 4), C (orange square at AOI 6), and D (yellow pentagon at AOI 6).

Adult Experiment

To validate the efficacy of the approach before applying it to infants, adult participants were tested. Adult participants from an introductory psychology course at The Ohio State University underwent a series of category learning tasks while their eye gaze was tracked by a Tobii T60 eye tracker at 60Hz. Adults had normal or corrected-to-normal vision, and sat approximately 60 cm away from the display.

Adult participants were assigned to either a supervised or an unsupervised condition. In the supervised condition participant were advised to look for a single distinguishing fea-

ture prior to the start of the experiment. In the unsupervised condition, no hint was given. Previous research suggests that this hint has large consequences with respect to how quickly participants learn to classify the objects, especially when there are few overlapping features (Kloos & Sloutsky, 2008).

The experiment had 8 blocks, with each block consisting of a learning phase and a testing phase. The learning phase had 8 trials, and during each trial a category member was displayed on the center of the screen one at a time for 1.5 seconds each. The testing phase immediately followed the learning phase. There were 4 testing blocks, and on each testing block a category member of the to-be-learned category and a member of a new category were presented side by side. The images were displayed until the participant identified the learned category object by pressing a key. The position of the learned category member (left or right) was counter-balanced. In addition, in-between trials a randomly located fixation point (cross-hair) directed the participants gaze to a position on the monitor.

The to-be-learned category remained the same for the first 4 blocks. A second to-be-learned category was introduced in the final 4 blocks. If the experiment started with category A or B, the second category was either the C or D, and vice-versa. That is because while the A and B categories were defined by a relevant feature at AOI 4, categories C and D were defined by a relevant feature at AOI 6. This provided a mechanism to verify the reproducibility of the variables determined most important.

Infant Experiment

The infant experiment was similar to the adult experiment, but was adapted for infants by using a familiarization paradigm. To aid infant learning, category exemplars were shown in pairs on each trial. This was also done so that the images in the learning and testing phases had an identical layout. Furthermore, there was only a supervised condition, in which the infants were presented with a pre-trial fixation video of synchronized sound and motion (e.g., looming with whistle sound) to draw the infants attention to the single category-relevant feature. Once the infant looked at the fixation video, the learning trial commenced. Infants had to accumulate 3 seconds of looking to the category exemplar pairs. Whenever an infant looked away, an attention-grabbing fixation was presented until the infant reconnected with the images on the screen. After accumulating 3 seconds of looking to the stimulus pair, the supervisory fixation video was again presented followed by another learning image pair. This procedure was repeated with 8 learning pairs per block.

The testing phase of the infant experiment used a paired preference test with one category member paired with one non-category member. The standard assumption is that the infant can discriminate between the category and non-category object if he or she consistently looks at one object significantly more than the other object (i.e., whether the infant displays a novelty or familiarity preference). There were two test trials per block, where an exemplar from the learned category was paired with an exemplar from a novel category. Test trials were presented for a fixed duration of 6 seconds, and left/right position of familiar or novel category objects was counterbalanced.

Filtering eye tracking data

The gaze data obtained with the Tobii eye tracker contains noise, missing data, and micro-saccades, which makes identifying true fixations and saccades difficult. Therefore, raw eye tracking data exported from every experimental block was filtered using a Kalman filter (Murphy, 2004) before extracting the variables of interest. The eye gaze data from both the left and right eye were filtered separately. The average of the filtered data from left and right eyes yielded the mean eye gaze data, which were used in the current analyses.

Labeling the Data

The eye movement sequences during the *learning phase* of the experiment aid in understanding category learning, while the sequences during the *testing phase* aid our understanding of category use. Before applying our methodology to understand these processes, however, the eye tracking data from both the learning and testing phases of the experiments were labeled as *learner* (class 1), *non-learner* (class 0), or *indeterminate* (class 2). Indeterminate samples were not used.

Adult Labels: Intuitively, labels for adult data are readily identified based on the accuracy of the responses during the testing phase. An uninterrupted string of correct responses during the testing phase suggests that the participant has learned the category. Each adult experimental block yielded 12 eye movement sequences. These correspond to eye movements during the presentation of 8 exemplar images during the learning phase and 4 test images during the testing phase. Adult participants had 4 blocks of learning and discriminating the same category before switching to a new category. This amounted to 32 samples of the learning phase, and 16 samples of the testing phase for each category per participant. The 16 samples from the testing phase were associated with a 16 digit binary string, called the *response string*. This data structure shows performance over the first and last 4 blocks of the experiment. A one identifies a correct response, while a zero denotes an incorrect response on the associated test trial.

We expect a learner's response string to contain a series of ones beginning within the string and terminating at the end of the response string. This pattern indicates that at some point the participant learned the category and correctly discriminated the category from that point on. A participant who has not learned the category (non-learner) would select one of the two stimuli by chance in each trial. A non-learner could get lucky and achieve a series of correct guesses. In order to determine if a participant is a learner or a non-learner we need to establish a criterion that allows us to reject chance as the cause for a series of ones. The question that we need to answer is how many ones we should expect for a learner. We address this problem by assessing how likely it is that we see a sequence of M consecutive ones in a binary response string

of length $R = 16$. Under the null hypothesis, the participant does not know the category label and selects one of the stimuli by chance, giving her a 50% chance of correctly guessing the category member. Each sequence is equally likely given this assumption, so the probability of guessing at least M right in a row is the total number of sequences having M ones in a row ($(R - M + 1) \times 2^{(R-M)}$) divided by the total number of binary sequences of length R (2^R). This yields the probability $p = (R - M + 1)/(2^M)$. For $R = 16$, $M = 10$ is the minimum number that achieves a significance level of $p < 0.01$ ($p = 0.0068$). Therefore, we rejected the null hypothesis that a participant was guessing randomly when we identified a consecutive string of 10 correct responses.

We call the position of the first correct response in this string of correct responses the point of learning (POL). The *test phase* and *learning phase* samples before the POL were labeled as non-learner, while the samples after the POL were labeled as learner. The learning phase samples from the block associated with the POL were labeled as indeterminate, because it was unclear at exactly which trial during the block the category was learned.

If the learning criterion was not achieved, we then identified the remaining non-learning and indeterminate samples. We first labeled correct responses at the end of the respond string as indeterminate. Those samples did not meet the learning criterion, but might be attributed to learning late in the experiment. The remaining samples were labeled as non-learner. Approximately 8% of the adult eye track samples were labeled indeterminate.

Infant Labels: Obviously, infants are not able to respond by keyboard to identify a category object. Instead, we used a variant of the preferential looking paradigm to determine if an infant could discriminate between novel exemplars of a familiar category object and a novel category object. Recall that the preferential looking paradigm assumes that infants who consistently look more to one class of stimuli when shown two classes of stimuli are able to discriminate between the two classes. This means that if the infant consistently looks longer at the learned category object (or novel category object), then he or she is assumed to be discriminating between the familiar and novel categories.

Given this paradigm, we labeled each infant's gaze data by blocks. Each block consisted of two test phase samples. We determined novelty preference as the ratio of total looking time to the novel category object compared to the total looking time to the novel category plus the familiar category object. We sorted the mean of the novelty preference for each block according to the absolute difference from 0.5. A third of the blocks with mean novelty preference closest to 0.5 were labeled as non-learner. The third of the blocks with novelty preference furthest from 0.5 in absolute value were labeled learner. Otherwise, the samples were labeled indeterminate. Approximately 33% of the infant eye track samples were labeled indeterminate.

Variables List

We compiled an over-complete list of eye tracking variables. We began with the fundamental variables, fixations and saccades. Fixations occur when eye gaze is maintained at a single position for at least 100ms. They were identified using the dispersion threshold algorithm of (Salvucci & Goldberg, 2000). Saccades are rapid eye movements that move the eye gaze between points of fixation. To be considered a saccade, the eye movement needed to exceed smooth pursuit velocity of 30° per second or 0.5° per sample at 60Hz (Stampe, 1993). The fixations and saccades were determined with respect to a specific Area of Interest (AOI) within an object. AOIs are regions of an object image or scene that can be grouped in some meaningful way, such as color uniformity, and are relevant or non-relevant based on their role in determining the object category.

These fundamental eye tracking variables were combined in various ways to derive a larger set of variables. Our variables list is defined as follows:

1. *AOI fixation percentage* describes the percentage of time fixated at the different AOIs during a trial. All non-AOI fixations were discarded in this and all of the variables defined. The fixation percentages were normalized so that they sum to 1, unless there were no fixations at AOIs. In that case, all percentages were set to 0.

2. *Relevant AOI fixation density* describes the percentage of time fixated at the relevant AOI(s).

3. *AOI fixation sequence* describes the sequence of AOI fixations during one trial. We limited this sequence to a fixed number of fixations, starting with trial onset (not counting fixations to the fixation mark). The number of fixations to consider as well as the start position were determined using cross validation (CV). In addition, the fixation sequence was represented as a sequence of relevant and non-relevant AOI fixations. The analysis showed that the latter representation was more informative in some cases.

4. *Duration of fixations in sequence* describes the duration of each fixation in the sequence described by variable 3.

5. *Total distance traveled by eye* is a scalar describing the total distance traveled by the eye gaze during a trial.

6. *Histogram of fixation distances to relevant AOI* describes how much time is spent fixated near or far from the relevant AOI(s). The number of bins was determined using CV.

7. *Number of unique AOIs visited* is a scalar describing the total number of unique AOIs fixated during a trial.

8. *Saccade sequence* is similar to variable 3, but describes the sequence of AOI saccades during one trial. All saccades whose targets were not to AOIs were discarded in this and all of the variables defined. The sequence was limited to a fixed number of saccades, starting at the first saccade. The

number of saccades to consider as well as the start saccade were determined using CV. In addition, the saccade sequence was represented as a sequence of saccades to relevant and non-relevant AOIs.

9. *Relative number of saccades to an AOI* is the saccade analogue of variable 1, and describes the relative number of saccades to the AOIs during eye movement.

10. *Fixation latency to relevant AOI* describes the delay before fixating at a relevant AOI during an eye movement. It is a scalar between 0 and 1, where 0 corresponds to fixating to a relevant AOI immediately and 1 describes a sequence where a relevant AOI is never fixated.

11. *Saccade latency to relevant AOI* is a scalar between 0 and 1 defining the delay before saccading to a relevant AOI.

Thus, eye movement was represented by a *feature vector* $\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$ whose d entries correspond to the variables described. For clarity, *features* denote the entries of the feature vector which encodes the eye tracking variables, while *variables* correspond to the measures of the eye tracking enumerated above. Therefore, d is much larger than 11, because encoding certain variables requires multiple feature values. In addition, the information encoded by several of these features overlaps. This over-complete representation allows us to find the encoding that is best suited to describe the categorization task. To this end we performed variable selection on this over-complete set.

Variable Selection

Our goal is to identify the subset of variables from the set defined above that can best separate the classes: category learners and non-learners. This was achieved using ANOVA feature selection by ranking. ANOVA feature selection relies on a standard hypothesis test on each feature of $\mathbf{x}$. Specifically, let x_i denote the ith feature of $\mathbf{x}$. Using a dataset of eye tracking feature vectors and the associated class labels, we performed a two tailed t-test of the null hypothesis which states that samples of x_i coming from classes 1 and 0 are independent random samples from normal distributions with equal means, μ_{i1} and μ_{i0}, respectively. The alternative says that the class means are different. We calculated the test statistic and the corresponding p-value. A low p-value means the null hypothesis is rejected with confidence. Since the goal is to find the variables which best separate the classes, the feature with lowest p-value is ranked as best. The p-values were calculated for all features $x_i, i = 1 \ldots d$, and they were ranked from best to worst according to increasing p-values. If we vectorize the indices of the t top ranked features as $\mathbf{k} = (k_1, k_2, \ldots, k_t)^T$, then after feature selection $\mathbf{x} = (x_{k_1}, x_{k_2} \ldots, x_{k_t})^T$.

Linear Classification

Once the important variables were identified, we used them to classify the gaze data as having originated from a learner
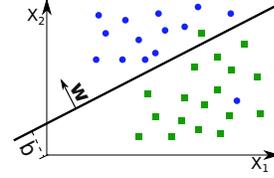

Figure 2: Illustration of a linear classifier. $\mathbf{w}$ is the normal vector of the hyperplane which separates the feature space into two decision regions, and b is the distance from the origin to the hyperplane. The blue circles represent samples from class 1, while the green squares represent samples from class 0. All but one of the blue circles exists on the positive side of hyperplane, and are classified correctly.

or non-learner. This required that we train a classifier to distinguish between two classes of data. Recall that each eye movement results in a feature vector, or *sample* $\mathbf{x}$. A classifier defines a decision rule for predicting whether a sample is from class 0 or 1. A linear classifier was used because of its ease of interpretation (Martinez & Zhu, 2005) – the absolute model weights give the relative importance of the eye tracking variables. We illustrate in Fig. 2 with a 2-dimensional linear classifier model specified by $\mathbf{w}$ and b. $\mathbf{w}$ is the normal vector of the hyperplane which separates the feature space into two decision regions, and b is the distance from the origin to the hyperplane (i.e., the offset).

All samples $\mathbf{x}$ above the hyperplane are assigned to class 1 while the samples below are assigned to class 0. Data samples $\mathbf{x}$ existing on the boundary satisfy $\mathbf{w}^T \mathbf{x} - b = 0$. Therefore, samples are classified according to the sign of $\mathbf{w}^T \mathbf{x} - b$. In this example $\mathbf{w} = (-.55, .83)^T$, so the second dimension, x_2, is more informative for classification because it has a larger absolute value. Note that in our case the feature space has not two but up to 334 dimensions, depending on the cut-off for variable selection.

In this work we used the Support Vector Machine (SVM) classifier. SVM is a linear classifier which maximizes the margin between two classes of data (Burges, 1998). In the case that the training samples are perfectly separable by a hyperplane, we can find $\mathbf{w}$ and b such that the data satisfies the following constraints,

$$\mathbf{x}_i^T \mathbf{w} - b \geq 1 \text{ for } y_i = 1, \tag{1}$$
$$\mathbf{x}_i^T \mathbf{w} - b \leq -1 \text{ for } y_i = 0. \tag{2}$$

Essentially, these constraints specify that the samples from the different classes reside on opposite sides of the decision boundary. The margin between the classes, defined by $\frac{2}{\|\mathbf{w}\|_2}$ where $\|\cdot\|_2$ defines the L2-norm, is then maximized subject to the above constraints. The dual formulation of the constrained optimization problem results in a quadratic program for $\mathbf{w}$ and b. In the case that samples from each class are not linearly separable, a penalty is introduced to penalize the amount that a sample is on the wrong side of the hyperplane. Again, the dual formulation results in a quadratic program for $\mathbf{w}$ and b. We used the implementation of (Chang & Lin, 2001).

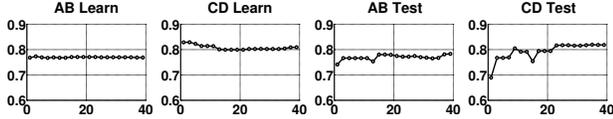The classification accuracy used for adults was the leave-

Figure 3: Leave-one-subject-out cross-validation accuracy for adult participants as a function of the number of top ranked variables used for classification. Classification was done with SVM while ANOVA was used for feature selection. AB and CD correspond to category object A or B and C or D respectively. In almost all cases, the classification accuracy was near the maximum after including very few features.

one-subject-out cross-validation (LOSO-CV) accuracy. In LOSO-CV, the samples belonging to one participant are sequestered, and the remaining samples are used to train the classifier. The sequestered samples are then classified with the learned classifier, and the procedure is repeated for every participant in the database. The total number of correctly classified samples divided by the total number of samples is the LOSO-CV accuracy. The classification accuracy used for infants was the leave-one-experiment-block-out cross-validation (LOBO-CV) accuracy. This alternative accuracy measure makes more effective use of the eye movement data when the sample size is very small. In LOBO-CV, the samples belonging to one experiment block are sequestered, and the remaining samples are used to train the classifier. The sequestered samples are then classified with the learned classifier, and the procedure is repeated for every block in the database. The total number of correctly classified samples over the total number of samples is the LOBO-CV accuracy.

Results

Adult Experiment

A total of 24 adults were tested in the supervised experiment while 46 adults were tested in the unsupervised adult experiment. This resulted in 728 learning class samples and 1256 non-learning class samples for the learning phase, and 473 learning class samples and 601 non-learning class samples for the testing phase in the A or B category learning condition. There were 496 learning class samples and 1568 non-learning class samples for the learning phase, and 323 learning class samples and 717 non-learning class samples for the testing phase in the C or D category learning condition. The indeterminate samples were not used in any of the experiments. After labeling the data, the eye tracking variables were extracted from each gaze sequence. Each labeled data sample resulted in a 182-dimensional feature vector for the learning phase samples, and a 334-dimensional feature vector for the testing phase samples.

SVM was applied to determine the LOSO-CV error as a function of the number of top features selected by ANOVA feature selection. The results are summarized in Fig. 3, and show that a very small set of variables yields a high classification rate. Adding more variables does not improve the accuracy.

The stable performance beyond just a few variables suggests that a small number of variables is sufficient for dis-

Table 1: The following variables were determined most relevant during the adult category learning and testing phases. The results suggest that looking at the relevant AOI is important during the learning phase. In addition, the first few fixations are very important during testing. The bold face entries show variables that were consistently determined most relevant across category conditions. We used shorthand notation for a few words: fixation (fix), saccade (sac), relevant (rel).

A or B Learning Condition	C or D Learning Condition
1. Latency to relevant AOI fix	**1. Latency to relevant AOI fix**
2. Density of fix at AOI 4	**2. Density of fix at AOI 6**
3. Distance to AOI 4, hist bin 2 of 30	3. Distance to AOI 6, hist bin 5 of 30
4. Look at AOI 4 on fix 2	**4. Look at AOI 6 on fix 1**
5. Look at AOI 4 on fix 1	5. Look at non-rel AOI on fix 1

A or B Testing Condition	C or D Testing Condition
1. Look at non-relevant AOI on fix 3	1. Look at non-relevant AOI on fix 4
2. Look at non-relevant AOI on fix 2	**2. Look at non-relevant AOI on fix 3**
3. Look at non-relevant AOI on sac 2	**3. Look at non-relevant AOI on sac 2**
4. Look at non-relevant AOI on fix 1	**4. number of unique AOIs fixated**
5. number of unique AOIs fixated	5. Look at non-relevant AOI on sac 3

criminating learners and non-learners. The top five variables are listed in Table 1. The bolded entries are present in the top five variables across both the A or B and C or D conditions. Note that AOI 4 for the A or B condition is equivalent to AOI 6 in the C or D test condition.

Infant Experiment

A total of 16 infants ranging from 6 to 8 months of age were tested in the supervised infant experiment. One participant's data were discarded because the infant would not cooperate. This resulted in 135 learning class samples and 137 non-learning class samples for the learning phase, and 40 learning class samples and 40 non-learning class samples for the testing phase in the A or B category learning condition. For the C or D category learning condition this resulted in 139 learning class samples and 127 non-learning class samples for the learning phase, and 40 learning class samples and 40 non-learning class samples for the testing phase. The indeterminate samples were not used in any of the experiments. After labeling the data, the eye tracking variables were extracted from each gaze sequence. Each labeled data sample resulted in a 334-dimensional feature vector for the learning and testing phase samples.

SVM was applied to determine the LOBO-CV error as a function of the number of top features selected by ANOVA feature selection. The results are shown in Fig. 4, where we see that the accuracy varies when considering different features. Thus, while classification of learners and non-learners is still possible with infants, it is not as clear-cut as with adults. This is to be expected because of the amount of random movements typical of babies. The top five infant variables are shown in Table 2. There are no bold entries because no variables were consistently selected across the A or B and C or D conditions.
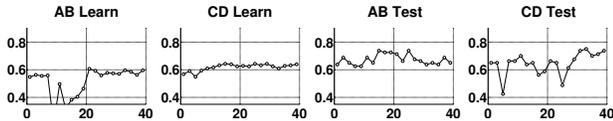
Figure 4: Leave-one-block-out cross-validation accuracy for infant participants as a function of the number of top ranked variables used for classification. Classification was done with SVM while ANOVA was used for feature selection. AB and CD correspond to category object A or B and C or D respectively. Chance level was at 0.5.

Table 2: The following variables were determined most relevant during the infant category learning and testing phases. There were no consistently relevant features across category conditions. We used shorthand notation for a few words: fixation (fix), saccade (sac), relevant (rel).

A or B Learning Condition	C or D Learning Condition
1. Den of fix at AOI 10	1. Distance to AOI 13, hist bin 5 of 35
2. look at AOI 10 on fix 3	2. Distance to AOI 6, hist bin 21 of 35
3. Distance to AOI 11, hist bin 5 of 35	3. Density of fix at AOI 13
4. Density of saccade to AOI 10	4. Distance to AOI 13, hist bin 2 of 35
5. Distance to AOI 4, hist bin 20 of 35	5. Look at non-rel AOI on sac 3

A or B Testing Condition	C or D Testing Condition
1. Look at non-relevant AOI on fix 6	1. Density of fix at AOI 7
2. Distance to AOI 4, hist bin 20 of 35	2. Look at AOI 3 on sac 3
3. Distance to AOI 4, hist bin 8 of 35	3. look at AOI 7 on fix 1
4. Look at AOI 7 on fix 4	4. Look at AOI 10 on fix 4
5. Density of sac to AOI 10	5. look at AOI 7 on fix 3

Comparing Infants to Adults

The above results raise a new question. How similar are the attention models of adults and infants? Specifically, since the infant data is so noisy, can we use the adult model to improve on the infant one? To test this, we used the adult classifier model trained with the top five variables to predict if infants were learners or non-learners. This was done only for the testing phase, because the testing phase images for adults and infants are similar so that the extracted variables correspond. Infants were classified with 49% accuracy in the A or B condition, and with 50% accuracy in the C or D condition with chance level at 50%. These findings suggest that infant category learners do not direct their attention like adult learners.

Discussion and Conclusion

We have developed a methodology for automatically determining eye tracking variables that are relevant to understanding category learning and discrimination processes. Previous research has relied on ad-hoc techniques to determine which variables should be analyzed. Instead, we used statistical methods to find the important variables in an over-complete set of variables.

The efficacy of the approach was verified with an adult categorization study. The variables determined most relevant for adults emphasize looking at the relevant AOI(s) longer, and earlier during the categorization tasks. This result is satisfying for two reasons: 1) It is expected that category learners quickly focus their efforts on the relevant AOI(s), and 2) These variables coincide with the variables *proportion fixation time* and *relative priority* of previous eye-tracking category learning studies such as (Rehder & Hoffman, 2005). Finally, we demonstrated that the adult model does not predict infant categorization. This is evidence of different attention processes for infants and adults during categorization.

Note that the important variables were verified by the *task* and *stimuli* described. Altering these parameters may result in different important variables. By comparing the important variables among different tasks and stimuli, we can further dissociate which eye tracking variables are linked to specific processes during categorization.

References

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*, 121–167.

Chang, C.-C., & Lin, C.-J. (2001). LIBSVM: a library for support vector machines [Computer software manual]. (Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm)

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*(1), 52-72.

Martinez, A. M., & Zhu, M. (2005). Where are linear feature extraction methods applicable? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *27*(12), 1934–1944.

Murphy, K. (2004). *Kalman filter toolbox for matlab*. Available from http://www.cs.ubc.ca/~murphyk/Software/Kalman/kalman.html

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463–475.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1-41.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Etra '00: Proceedings of the 2000 symposium on eye tracking research & applications* (pp. 71–78). New York, NY, USA.

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavioral Research Methods, Instruments, & Computers*, *25*(2), 137–142.