Eye Movements Disclose Decisions in Set

Joost Broekens [a] Walter A. Kosters [b] Timo de Vries [b]

[a] *Section Man-Machine Interaction, Delft University of Technology, The Netherlands*
joost.broekens@gmail.com
[b] *Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands*
kosters@liacs.nl, vriestimo@gmail.com

Abstract

During the past several decades, much research has been done in the field of visual attention and cognitive processing, i.e., how does the human brain process visual stimuli. We show that the card game SET is a good candidate for investigating the relation between top-down versus bottom-up visual information processing. Further, we show that Machine Learning techniques can be effectively used to study this relation. We use these techniques to predict if (*a*) the player thinks he/she found a set (a particular three-card pattern), and (*b*) he/she correctly identifies a set, both based on eye movements of the player.

Our results indicate the following. First, pop-out plays a role in SET-playing performance. Second, the more we move towards predicting correctness (the player found a correct set), the less important eye movements are and the more important player experience is. Third, the more we move towards predicting the claiming of a set (the behavior), the more important eye movements are and the less important player experience is. This indicates that eye movements disclose whether a player thinks he/she found a set or not, but not whether that player correctly identified a set.

1 Introduction

During the past several decades, much research has been done in the field of visual attention and cognitive processing, i.e., how does the human brain process visual stimuli. A continuing debate is how the interplay between top-down (i.e., voluntary attention control) and bottom-up (i.e., stimulus driven attention control) is organized [3, 4]. Here we show that the game of SET[1] [8] is particularly interesting for studying this tradeoff.

When a person plays this card game, he or she gathers visual information that is used to make high-level cognitive decisions about the (non-)existence of particular three-card patterns (called a *set*) in a collection of cards laid out on the table before the player. This means that to play the game successfully, the player has to employ different visual strategies including ones that rely on pop-out [1] (bottom-up) and guided search [3] (top-down). In this paper we investigate two questions. First, does pop-out (which can be described as the situation in which stimuli that are defined by a unique perceptual feature automatically and selectively guide attention, see [1]) play a role in SET-playing performance. Second, is it possible to use AI [6] and data mining [10] techniques to predict if (*a*) the player thinks he/she found a set, and (*b*) he/she correctly identifies a set, both based on eye movements of the player.

The structure of this paper is as follows. In Section 2 we explain the game of SET. The background and motivation for using SET in this research are mentioned in Section 3. The experimental setup is discussed in Section 4, and Section 5 explains how the data was analyzed. Section 6 has results on the experiments, and Section 7 contains conclusions and suggestions for further research.

2 The Game of SET

The game of SET is played with cards, each having a unique *number* of objects, of a particular *shape*, *shading* and *color*. Each of these 4 features has 3 possibilities, hence a total of $3^4 = 81$ different cards.

[1] SET is a trademark of Set Enterprises, Inc, www.setgame.com.

There exist many variations on SET, all involving the concept of a “set”. A *set* consists of 3 different cards, satisfying *all* of the following conditions [8]:

- they all have the same number of objects, or they all differ in number;
- they all have the same shape, or they have three different shapes;
- they all have the same shading, or they have three different shadings;
- they all have the same color, or they have three different colors.

Figure 1, left, shows an example of a set.

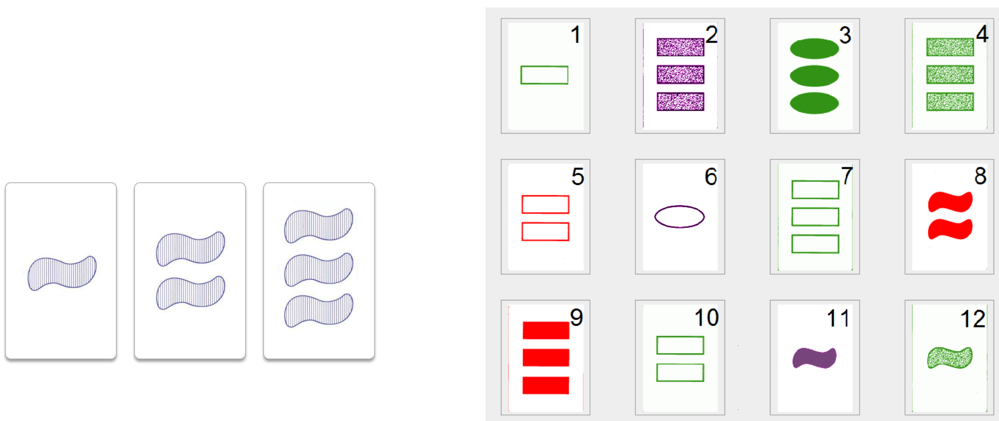


Figure 1: Left: Example of a set: each card has a different number of objects, while the shape, shading and color are the same for each card. Right: Example of a 12 card SET layout. This same layout is also used for the heat maps from Figure 2, and visible in the setup from Figure 3, right.

SET can be played with two or more players (but in our experiment, a human played against one imaginary opponent simulated by the computer). Usually, the dealer puts 12 cards on the table, the so-called *layout*. If one notices a set among those cards, the player calls “Set!” and points out the cards that form the set: he/she claims a set. Once the cards have been verified as a set, the person takes the 3 cards, after which 3 new cards are put on the table. If a player cannot stand the claim, he or she is not allowed to play for the current layout anymore. If there is apparently no set within the 12 cards on the table, the dealer will add 3 more and the game continues. When the deck is empty and there is no set left in the layout, the player with the highest amount of sets is the winner.

The layout from Figure 1, right, contains four sets. The cards 1, 10 and 7 form a set: they are all green, rectangular and have the same shading, whereas the number of objects differs between the three of them. The cards 4, 6 and 8 also form a set (this is the set found in Figure 2, right, from the next section), a fact that beginners find somewhat harder to see. The cards 3, 8 and 11 are not a set, but they come close: only the shapes of 8 and 11 are the same, but the other features do satisfy the required properties. There is always precisely one card that makes a set with two given cards; however, this card is not necessarily in the layout.

The game of SET gives rise to many interesting mathematical questions, see, e.g., [2]. However, in this paper we use the game to investigate visual attention. In the next section we explain in detail why SET is a good vehicle to study this phenomenon.

3 Background

The game of SET is easy to understand, but difficult to play at first sight. As explained above, players have to find patterns of three different cards that range from easy to quite difficult. In some of these patterns pop-out plays an important role (e.g., when all cards in a to-be-found set are red), while in others visual search plays an important role (e.g., when all four conditions are different). In visual search the player controls his/her attentional spot-light to find a particular thing (card in SET) according to a goal or hypothesis (e.g., “I found two cards with two objects on each card, let's find a third card with two objects on it and figure out if it fits the other three SET criteria").

It seems plausible to assume that eye movements of a player reveal something about his or her mental processes. In particular, when noticing or claiming a possible pattern, this could be reflected in the eye movements, as exemplified in Figure 2.

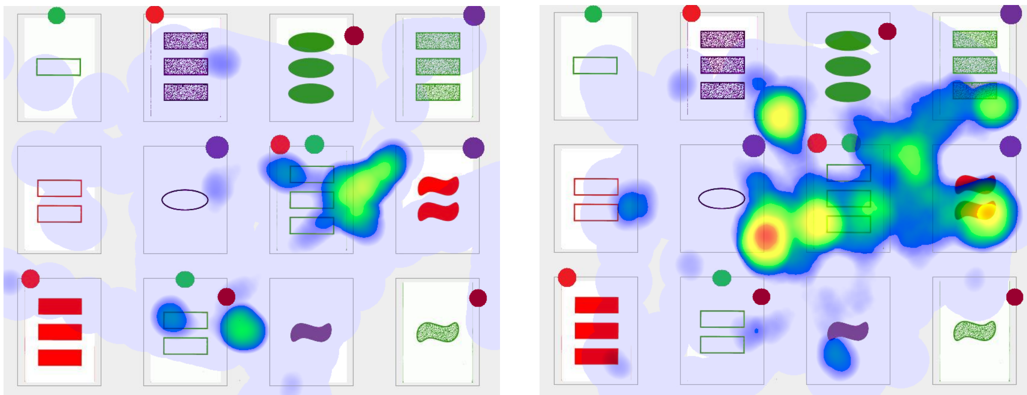


Figure 2: An example of two heat maps showing the distribution of attention. Dots (not visible for the participant) indicate the sets. The colored regions try to give the main focus of attention as obtained from eye movement data. In the left figure no set was found, in the right one a set (consisting of the cards with the purple dots) has just been detected.

As the layouts in the game of SET can be generated by a computer, it is easy to manipulate the type of patterns that can occur. Therefore, the game of SET allows direct investigation of the relation between bottom-up and top-down processes. In this paper we report on experiments that are an initial and exploratory investigation of this relation.

Others have already used SET in scientific research. For example, in [9], an attempt is made to model human SET-playing behavior. The authors have done experiments with human participants playing against an artificial intelligent SET-playing agent and rating the humanness of the agent. They focused on the development of a cognitively plausible model of human SET playing behavior.

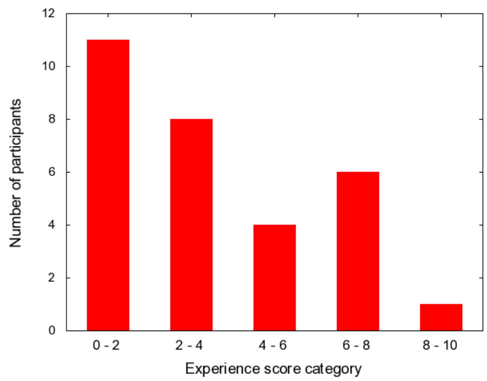
Craig [5] on the other hand focused on differences in SET-playing strategies. The goal of Craig's first experiment, a training study, was to test the influence of training on problem solving abilities. After subjects had taken a pre-test, the training group was asked to do the "daily puzzle" every day for approximately three weeks. Finally, all subjects took a post-test. The difference in solving times for both groups indicated that the training did indeed enhance performance. In a second experiment, Craig investigated the difference between expert and novice players. Craig concluded that "Novices tend to see the cards as an ordered succession of attributes while experts view the cards as composed of concurrent attributes." Also, "Novices tend to reconstruct their mental model of the board with each new deal, while experts merely update their mental model with the most current information."

The findings of Craig [5] are of direct importance to our study: they show that complex pattern detection can be learned, and that there are individual differences in SET-playing strategies. We investigate a related issue, i.e., if pop-out plays a role in finding such complex patterns. Secondly, we investigated if eye movements can be related to high-level mental processes while playing SET.

4 Experimental Setup

We devised an experiment, in which 31 individuals, mostly students, participated. The experience subjects had with playing SET was distributed as shown in Figure 3, left. All participants had to perform the same SET tasks in the same order. The experiment consisted of three SET-tasks, intertwined with dummy tasks that had nothing to do with SET, such that the participants would not immediately catch the true intent of the experiment.

For the first task, we generated 10 random layouts with each layout containing multiple sets. These same 10 layouts were shown in random order to each participant. For each layout, the goal was to find a set be-



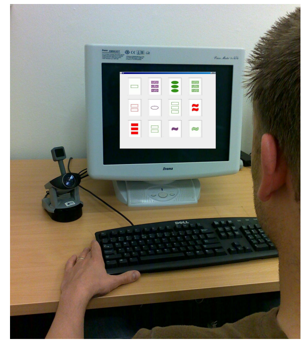


Figure 3: Left: Distribution of participants across experience categories. Right: Setup.

fore the computer opponent would. To mimic the aspect of time-pressure, the task description implicitly suggested there was a human opponent, though some participants did have questions about this. The "opponent" program adjusted the difficulty (i.e., the period after which the "opponent" finds a set) according to the response time of the participant.

For the second task, we generated a SET layout sequence that contained five layouts with each layout containing exactly one set. When a SET is found, the three cards are replaced by three new cards, just as in the normal game of SET. Each participant played the same sequence. Again an "opponent" played against the player.

The third task consisted of 80 randomly generated 6-card layouts, with approxmately half of the layouts containing exactly one set. Each layout was presented in random order with a short pause in between layouts. Further, the first series of 10 layouts were presented for 2 seconds, the second series for 1.5 seconds, and so on until the last series of 10 layouts were presented for 0.4 seconds. The goal of this third task was to evaluate the performance of participants based on pop-out, hence the short presentation times and the smaller layout (6 cards). In the first two tasks, eye tracking data was collected (see Figure 3, right, for an impression). In the last task it was not. In the first two tasks, if a participant thought to have found a set in a layout, he/she pressed the space bar, and indicated the set with the mouse. In the third task the particant only indicated using the spacebar if he/she thought a set was found. Further, participants filled in a questionnaire that included items about their SET playing experience (how often per month, last time one played, subjective expertise). These items were aggregated into one *Subjective Experience Score* (also used as basis for Figure 3, left).

5 Data Analysis

During all three tasks, we gathered the time needed to press the space bar (time to press), whether or not the bar was pressed / not pressed, and —if pressed— whether the indicated set was correct or incorrect. During the first two tasks, the participant's gaze position was captured every 20 ms and the coordinates of the focal position were written to a log file, along with a timestamp. Unfortunately, many participants' gaze records seemed compressed and shifted in some direction, some more than others. This must be due to calibration errors. For this reason, we have chosen not to rely on absolute coordinates for feature extraction, but on relative positions. As a result, we could not match eye tracking data to cards or screen location, leaving an even greater challenge of finding a relationship between eye movement and whether a player thinks he/she found a set, and whether a player correctly identifies a set. In Figure 2 two heat maps are presented that give an impression of data gathered during one task.

Preprocessing of these raw eye-tracking data was necessary in order to extract meaningful features that could be fed into the prediction models used to predict space-bar pressing and correctness. During preprocessing we discriminated between *fixations* and *saccades* (fast eye movements), see [7]. Fixations maintain the visual gaze on a single location, whereas saccades move it to a different location. We have used a fixed distance threshold to distinguish saccades from drifts/tremors.

One participant was excluded from the analysis, as this participant, having a neurobehavioral developmental disorder (ADHD), generated eye-tracking data that indeed indicate extreme eye movement (in the order of 4 times the average screen distance crossed during a trial).

For each trial in task one and two (10×31 and 5×31 trials, respectively), we extracted the following fixation/saccade based features: (*a*) duration of fixation (in ms), (*b*) saccade length (in pixels), (*c*) saccade angle (in degrees), (*d*) delta saccade angle (the difference in degrees between the angles of two, successive saccades), (*e*) velocity (in pixels per second), (*f*) delta velocity, i.e., acceleration (in pixels per second^2). In addition, the following two features were extracted based on the heat-maps: (*g*) attention spread (in pixels), (*h*) attention-intensity maxima (based on the time spent gazing at the 10 most frequently visited locations). For each of these features (with the exception of (*g*) as that is a single number) a binned frequency distribution was calculated (10 bins), as well as the average and standard deviation. The values of the bins were used as independent variables (features) in the prediction models. This resulted in 7x(10+2)+1=85 features per trial. Further, to eliminate the possibility that the time a participant needs to decide upon a set is implicitly used as predictor for finding a set, we also created these 80 features for different segments of each trial, i.e., the last 1, 2, 4, 6 and 10 seconds just before the trials ends (opponent or participant found a set).

For each subject, an experience score was calculated based on the Subjective Experience Score from Section 4 and an *Objective Experience Score* based on their performance during the experiment, see Figure 3, left. These two numbers have also been used as features in the predictive models.

Our goal was not to optimize an individual machine learning model, nor to figure out which model would work best for this particular data set. Instead, we want to show that there is a relation between eye movement and set detection (a cognitive decision). Therefore, we used six different, and often-used prediction models as available in WEKA [12]: (*a*) Naïve Bayes, (*b*) Locally Weighted Learning, (*c*) Bagging, (*d*) Dagging, (*e*) Bayesian Network, and (*f*) MultiLayer Perceptron. The accuracy of the prediction models have been averaged. For this paper we trained each prediction model using 10-fold cross-validation, in order to prevent over fitting on the training data as much as possible. Each type of model was trained to predict four different binary classifications based on three possible behaviors of the participants in the second task (i.e., one set per layout, with set-based card-replacement). Either a participant did not claim a set (38 % of the trials), referred to as "non-push"; called "Set!" correctly (47 %), referred to as "correct"; or called "Set!" but failed to show a set (16 %), referred to as "incorrect". Together, "correct" and "incorrect" constitute "push". The binary classification tasks were:

- $\mathcal{C}_1$ "non-push" vs. "correct" or "incorrect", to predict the decision to push the space bar if a set seems to be found,
- $\mathcal{C}_2$ "non-push" or "incorrect" vs. "correct", to predict correctly identifying a set,
- $\mathcal{C}_3$ "correct" vs. "incorrect", to predict if after the decision to push a correct set is indeed indicated, and
- $\mathcal{C}_4$ "non-push" vs. "incorrect", to predict the type of error made by the participant (not found or incorrectly indicated).

The models were trained on each segment size (1, 2, 4, 6, and 10 seconds) separately and then averaged. Each training was done 5 times to compensate for models that do not have a deterministic output, so each final accuracy result is based on $5 \times 5 \times 6$ predictions.

Finally, we used standard statistics to detect relations between performance on the pop-out task on the one hand, and eye movement behavior and performance in the first task, as well as the Subjective Experience Score.

6 Results

In this paper we investigate two questions, as mentioned in Section 1 (for more details regarding this study see [11]). First, does pop-out play a role in SET-playing performance. Second, is it possible to use Machine Learning techniques to predict if (*a*) the player thinks he/she found a set (i.e., calling "Set!"), and (*b*) he/she correctly identifies a set, both based on eye movements of a player.

Indeed, it seems that pop-out plays a significant role in playing SET. First, we obtain a significant positive correlation (0.53, $n = 30$, $p < 0.05$) between the Subjective Experience Score as indicated in the

	gaze data only				scores only				all data			
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$
NaiveBayes	74.6	68.1	48.9	73.2	59.2	70.7	84.6	68.5	74.3	69.9	52.3	73.5
LWL	72.5	69.8	68.8	73.6	56.7	62.1	79.2	66.4	72.7	69.9	77.6	74.3
Bagging	77.9	73.2	73.8	77.8	56.5	69.7	80.7	69.1	78.1	75.2	80.1	78.3
Dagging	73.1	66.0	74.2	73.9	62.4	65.6	75.4	70.9	71.6	69.7	75.0	74.3
BayesNet	76.6	72.6	74.8	78.1	62.4	65.4	76.1	70.9	75.7	72.9	75.4	78.5
MLP	74.9	69.2	75.1	76.5	62.4	54.5	75.4	70.9	74.3	74.2	74.9	77.9
average	**75.0**	**69.8**	**69.3**	**75.5**	**60.0**	**64.7**	**78.6**	**69.4**	**74.5**	**72.0**	**72.5**	**76.0**

Table 1: Classification accuracies for the classifications tasks $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$, when using gaze data only, with performance/experience scores only, and for all features, with Naïve Bayes, Locally Weighted Learning, Bagging, Dagging, Bayesian Network and MultiLayer Perceptron.

questionnaire and the amount of correctly identified sets in the pop-out task (pop-out score). This indicates that more experienced players are better at quickly identifying sets. This is an obvious result, and in line with [5]. Second, however, we found a significant negative correlation (-0.63, $n = 30$, $p < 0.05$) between the total gaze path length in the first task and the pop-out score, as well as a positive correlation between scores on the first task and pop-out score (0.58, $n = 30$, $p < 0.05$). We interpret these findings as follows: experienced SET players perform better at quickly identifying sets, because they detect patterns more efficiently. As a result they move less with their eyes. As the pop-out task was constructed in such a way that visual search was practically impossible, the best explanation for the performance in this task is that the set pattern is detected due to pop-out. This would mean that, probably as a result of training, complex, rule-based patterns and not only simple perceptual-feature based patterns can be detected via pop-out.

Please note, however, that an important characteristic of pop-out is that the speed of detection is more or less constant with respect to the size of the to-be-searched space and number of objects therein. This means that technically speaking, we can not claim the responsible mechanism is pop-out, as this would imply that in a very large layout (e.g., containing 100 cards) a person would still be able to detect sets immediately. Obviously that is not possible, unless by chance that set would be consisting of, e.g., one colour while all other cards in the layout have a different colour. This, however, can be explained by normal pop-out based on simple visual features such as color and has nothing to do with rule-based patterns. Probably a better term for what we found is fast, automatic and rule-dependent pattern recogntion.

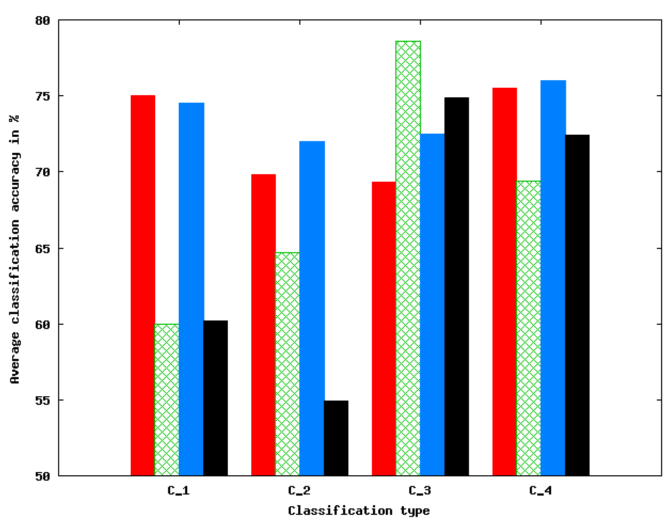


Figure 4: Average classification accuracy of the six machine learning methods with eye-tracking data features only (red solid), performance/experience scores only (green pattern) and using all features (blue solid), for each classification task $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$. The black columns indicate the performance of the naïve majority class predictor and act as a baseline.

With regards to predicting decisions based on eye tracking data we found the following. For interpretation of the accuracy we use the performance of the naïve majority class predictor. This predictor acts as a baseline. Detailed outcomes of the experiments can be found in Table 1 and Figure 4. When the eye tracking data features were used (red bars in Figure 4), the average accuracy for the prediction of the decision to call "Set!" ($\mathcal{C}_1$, $75\,\%$) was well above the baseline. This means that even when features abstract away from absolute eye-gaze location and time needed for finding a set (two potentially strong predictors), there is information about a participant's intention to call "Set!". Eye movement therefore tells us something about the decision to call "Set!".

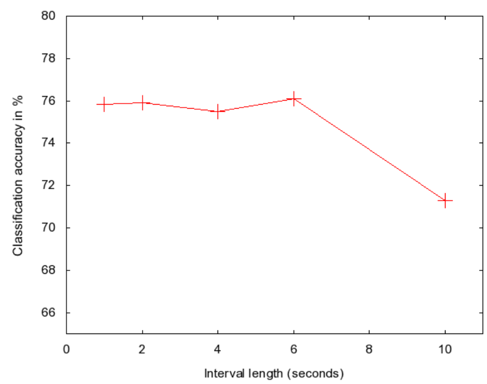


Figure 5: Average classification accuracy of the six machine learning methods with eye-tracking data features only, detailed for each segment of eye tracking data, for classification task $\mathcal{C}_1$.

Initially it seems that the models trained on eye movement data also predict the correctness of the found set ($\mathcal{C}_2$). However, when the model is trained to predict the correctness, given that it knows that the participant has called "Set!" ($\mathcal{C}_3$, correct vs. incorrect), the accuracy is poor. The good performance of the models when predicting correctness ($\mathcal{C}_2$) is thus a result of the good performance of predicting to call "Set!". This means that, based on the currently selected features, there is no difference between thinking that a set is found and finding an actual set. This makes sense, given that the eye movements of participants just before deciding to call "Set!" probably reflect their *impression* that they have found a set. The participant actually believes he/she has found a set. On the other hand, if the experience score of a subject is used (two features) to predict correct vs. incorrect set identification ($\mathcal{C}_3$), this does seem to outperform the baseline predictor (although marginally). This is plausible, as a participants' experience is probably a good indication that someone has found a correct set instead of an incorrect one. The interpretation that the Machine Learning models indeed predict the action of calling "Set!" is also shown in Figure 5. The features from the eye-tracking data in the segments from 1 to 6 seconds make a fairly stable predictor of a person calling "Set!", while when the last 10 seconds are used, the accuracy drops significantly.

This analysis is supported by the fact that, based on eye tracking data features, the accuracy of predicting an incorrectly identified set versus not finding a set ($\mathcal{C}_4$) is comparable to predicting to call "Set!" in the first place. This is obvious, as it essentially is the same as predicting to call "Set!" ($\mathcal{C}_1$), with the only difference that the claim was incorrect. As the eye movements do not reveal anything about correct or incorrect set identification, one would expect to see this similarity.

Overall there seem to be two trends in prediction accuracy. First, the more we move towards predicting correctness, the less important eye movements are and the more important experience is. Second, the more we move towards prediction of calling "Set!" (the behavior), the more important eye movements are and the less important experience is.

7 Conclusions and Further Research

We claim that the game of SET is a good candidate for investigating the relation between top-down versus bottom-up visual information processing. Further, we claim that Machine Learning techniques can be used effectively to study this relation.

These two claims are supported by concrete experimental results. In this paper we have investigated two questions. First, does pop-out play a role in SET-playing performance. Second, is it possible to use Machine Learning techniques, to predict if (*a*) the player thinks he/she found a set (i.e., calling "Set!"), and (*b*) he/she correctly identifies a set, both based on eye movements of the player. Our results indicate that pop-out plays a significant role in the performance of playing SET.

With regards to predicting decision making based on eye movements, there seem to be two trends in prediction accuracy. First, the more we move towards predicting correctness (the player found a correct set), the less important eye movements are and the more important player experience is. Second, the more we move towards the prediction of calling "Set!" (the behavior), the more important eye movements are and the less important player experience is. This can be explained because participants genuinely believe they found a set, so the eye movement when a correct set is found and the eye movement when an incorrect set is found is probably identical.

As future research we mention, besides obvious issues like involving more participants and better equipment for data gathering, a deeper analysis of the low-level features connected with high-level decision making. In particular, we would like to extend the analysis with features like the absolute coordinates of the peaks of the focal position. Finally, increasing the precision of absolute coordinates, thereby making it possible to actually use these in the prediction, might enhance prediction accuracy.

References

[1] S.A. Adler and J. Orprecio. The eyes have it: Visual pop-out in infants and adults. *Developmental Science* 9(2):189–206, 2006.

[2] N.G. de Bruijn. Set! [in Dutch] *Nieuw Archief voor Wiskunde* 5–3(4): 320–325, 2002.

[3] M.M. Chun and J.M. Wolfe. Visual attention. In *Blackwell Handbook of Perception*, pages 272–310, 2001.

[4] C.E. Connor, H.E. Egeth and S. Yantis. Visual attention: Bottom-up versus top-down. *Current Biology* 14: 850–852, 2004.

[5] J.M.B. Craig. Experiments with Set, Two studies on Set game play The University of Virginia, USA, 2004. `http://people.virginia.edu/~jmc4as/docs/JMBC-710-Research.doc`

[6] S.J. Russell and P. Norvig. *Artificial Intelligence, A Modern Approach*. Second edition, Prentice Hall, 2003.

[7] D.D. Salvucci and J.H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium*, pages 71–78, 2000.

[8] Wikipedia, Set (game). `http://en.wikipedia.org/wiki/Set_(game)` [retrieved on May 6, 2009].

[9] N.A. Taatgen, M. van Oploo, J. Braaksma and J. Niemantsverdriet. How to construct a believable opponent using cognitive modeling in the game of Set. In *Proceedings of the Fifth International Conference on Cognitive Modeling*, pages 201–206, 2003.

[10] P.-N. Tan, M. Steinbach and V. Kumar. *Introduction to Data Mining*. Pearson / Addison Wesley, 2006.

[11] T. de Vries. *Investigating Visual Attention and Reasoning through Cognitive Modeling*. Research Thesis, Leiden University, The Netherlands, 2008.

[12] I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Second edition, Morgan Kaufmann, 2005.