

Cognitive Science

Set as an Instance of a Real-world Visual-cognitive Task

--Manuscript Draft--

Manuscript Number:	11-1183R2
Full Title:	Set as an Instance of a Real-world Visual-cognitive Task
Article Type:	Regular Article (12,000 words)
Keywords:	visual perception; competitive parallelism; SET; ACT-R; Cognitive architecture; problem solving.
Corresponding Author:	Enkhbold Nyamsuren, Ph.D. candidate University of Groningen Groningen, Groningen NETHERLANDS
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Groningen
Corresponding Author's Secondary Institution:	
First Author:	Enkhbold Nyamsuren, Ph.D. candidate
First Author Secondary Information:	
Order of Authors:	Enkhbold Nyamsuren, Ph.D. candidate Niels A. Taatgen, Professor
Order of Authors Secondary Information:	
Abstract:	Complex problem solving is often an integral between perceptual processing and deliberate planning. But what balances these two processes, and how do novices differ from experts? We investigate the interplay between these two in the game of SET. The paper investigates how people combine bottom-up visual processes and top-down planning to succeed in this game. Using combinatorial and mixed-effect regression analysis of eye-movement protocols and a cognitive model of a human player, we show that SET players deploy both bottom-up and top-down processes in parallel to accomplish the same task. Combination of competition and cooperation of both types of processes is a major factor for a success in the game. Finally, we explore strategies players use during the game. Our findings suggest that within-trial strategy shifts can occur without need of explicit meta-cognitive control, but rather implicitly as a result of evolving memory activations.

Set as an Instance of a Real-world Visual-cognitive Task

Enkhbold Nyamsuren and Niels A. Taatgen

University of Groningen

(Last revised on March 07 2012)

Author Note

Enkhbold Nyamsuren, Department of Artificial Intelligence, University of Groningen; Niels A. Taatgen, Department of Artificial Intelligence, University of Groningen.

Correspondence concerning this article should be addressed to Enkhbold Nyamsuren or Niels A. Taatgen, Department of Artificial Intelligence, University of Groningen, Nijenborgh 9, 9747 AG Groningen, Netherlands.

E-mail: e.nyamsuren@rug.nl; n.a.taatgen@rug.nl

Abstract

Complex problem solving is often an integration of perceptual processing and deliberate planning. But what balances these two processes, and how do novices differ from experts? We investigate the interplay between these two in the game of SET. The paper investigates how people combine bottom-up visual processes and top-down planning to succeed in this game. Using combinatorial and mixed-effect regression analysis of eye-movement protocols and a cognitive model of a human player, we show that SET players deploy both bottom-up and top-down processes in parallel to accomplish the same task. The combination of competition and cooperation of both types of processes is a major factor of success in the game. Finally, we explore strategies players use during the game. Our findings suggest that within-trial strategy shifts can occur without the need of explicit meta-cognitive control, but rather implicitly as a result of evolving memory activations.

Keywords: visual perception, competitive parallelism, SET, ACT-R, cognitive architecture,
problem solving

Human performance in complex tasks is often a combination of internal planning and responding appropriately to the environment. Nevertheless, cognitive models of complex tasks typically focus on the mental planning aspects, and fail to take into account that the external world can heavily influence the control of behavior.

The role of the environment was first recognized in robotics (Brooks, 1991) but was later extended to human cognition to form *embodied cognition* (e.g., Clark, 1997). However, in more complex tasks it is clear that the control of behavior is not entirely in the environment. The challenge is therefore to understand how control is shared between goal-driven planning and processes that are driven by perceptual input. Moreover, the balance between goal and perceptually driven control is likely to change with expertise (Kirsh & Maglio, 1994). The approach we take in this paper follows the *threaded cognition* theory of multitasking (Salvucci & Taatgen, 2008). We will assume two parallel processes: a bottom-up visual process that scans the visual field on the basis of saliency and similarity, and a top-down planning process that tries to achieve the goal, but also biases the bottom-up process. The interaction between the two processes follows the central idea in threaded cognition that there is no overall executive process that balances parallel goals. Instead, the two processes alternate in using the cognitive resources (e.g., vision, declarative memory, procedural memory, etc.). Changes in the balance between the two occur if one process benefits more from learning than the other and therefore makes more efficient use of the resources available to it.

Finding an appropriate task to study the cognitive aspects of human behavior in real-life situations is not easy. However, games provide environments that often require the same type of complex processes that are usually involved in real-world situations (Green & Bavelier, 2004). This has the advantage that the behavior of a player can be studied in a controlled environment.

These qualities make games on a computer an ideal tool for studying complex cognitive processes. One such game is the card game SET¹.

The SET card deck consists of 81 cards. Each card differs from other cards by a unique combination of four attributes: color, number, shape and shading. Each attribute can have one of three distinct values: red, green and blue for the color; open, solid and textured for the shading; one, two and three for the number; oval, rectangle and wiggle for the shape. The gameplay for SET is relatively simple. At any moment in the game, 12 cards are dealt face up, as is shown in Figure 1. From those 12 cards, players should find any combination of three cards, further referred to as a set, satisfying a rule stating that in the three cards the values for each particular attribute should be all the same or all different. We will further refer to the number of attributes for which the three cards in the set have different values as the set *level*. A level 1 set has only one attribute with three different values, but three attributes with identical values. Correspondingly, there can be sets of level 2, 3 or 4. Figure 1 shows an example of a level 1 (different shape) and a level 4 set (all attributes are different). In a similar manner, we can quantify *perceptual similarity* of two cards as number of attributes that are shared between two. For example, the cards in a level 1 set have a perceptual similarity of 3 among each other since they have three attributes with identical values. Cards in a level 4 set have perceptual similarity of 0 because all attribute values are different.

In the regular game, when a set is found, the corresponding set cards are picked up and replaced with new cards from a deck. After the deck runs out, the player with the most cards wins. Even though a regular game of set consists of multiple rounds, we will refer to a “game of set” in what is normally a single round: finding a set in 12 displayed cards.

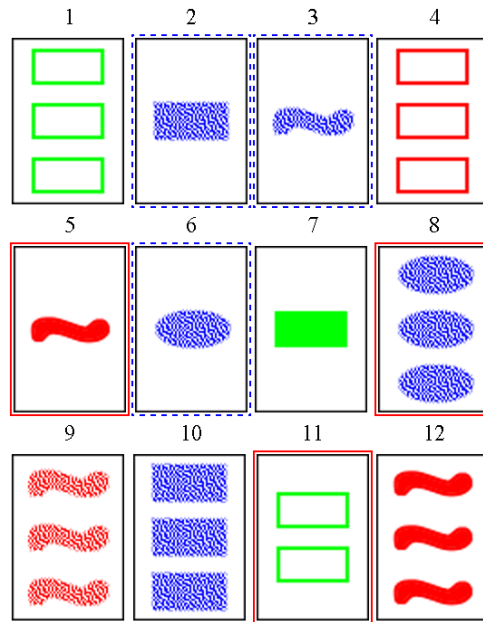


Figure 1. An example array of 12 cards. The cards with the solid highlight form a level 4 set (all attributes are different), and cards with dashed highlight form a level 1 set (Shape is different, and all other attributes are the same).

There are several advantages of choosing SET as a target game of study. Firstly, SET has an appealing simplicity of the game dynamics. The game has very simple rules to follow and a relatively static game environment. Despite the simplicity, SET requires complex cognitive processes including pattern recognition, visual processing and decision making. Previous studies on SET have established that both cognitive and perceptual processes are important (Jacob & Hochstein, 2008; Taatgen, Oploo, Braaksma and Niemantsverdriet, 2003). Without consideration of both of them in combination, important information in understanding of how players play the game will be inevitably lost. As such, the game of SET provides an excellent opportunity to study the dynamics of such processes in a relatively simple environment.

Next, the game is quite unpredictable in its structure, and players are not likely to replay the exact same sequence again. There are $7 \cdot 10^{13}$ possible combinations of 12 cards which makes it highly unlikely that player will play through the same 12 cards again. There are also 1080 different sets. This means that even experienced players will periodically have to find a set they have never encountered before.

Finally, game difficulty can differ significantly based on a player's strategy. Given an array of 12 cards with a single set in it, a player may choose to compare every possible combination of three cards. There are 220 possible combinations, and a probability of finding a set with random choice is $1/220$. However, player may also consider combinations of two cards since a pair uniquely defines a third card. In that strategy, a player would pick two cards, would then determine what the third card should be to complete a set, and would then see whether the predicted third card is actually among the remaining 10 cards. There are 66 possible pairs, and the same set is defined by three different pairs. Therefore, a probability of finding a set with a random choice of a pair is $1/22$. However, with an optimal search strategy player still has to consider a maximum of 54 pairs before finding a set. This is a significant decrease in complexity compared to a strategy where player has to compare every combination of three cards.

The above two strategies are both top-down in the sense that they do not take into account what the properties of the particular array of 12 cards are. However, players are likely to be using perceptual processes and clues, such as visual grouping and visual similarity, to decrease complexity or speed up the search. As an example, suppose that there are eight red cards and two cards each for blue and green. Furthermore, let's assume that player is using similarity in color to find a set. Blue and green cards can't have a set since there are only two cards in each group. There are 56 combinations of three cards among red cards and 32

combinations of three cards with different colors. It is already significant decrease in complexity from 220 to 88 possible combinations and 2.5 time increase in a chance probability of finding a set. Chance probability of finding a set among red cards is even higher $56/88$ or about $2/3$. This leverage in a chance probability only comes from a larger group size for red cards. For example, if there is an even split of four cards for each color then the chance probability of a set being among cards of the same color is only $4/76$. As will be discussed next, players actually exploit the advantage of a larger group size.

There are two studies directly relevant to the work in this paper. Jacob and Hochstein (2008) did several experiments with human subjects playing SET on a computer without any opponent. Each experiment was designed to test a particular aspect of the game including a strategy of playing the game, dependency of the performance on the set level, attribute preference and the learning. Taatgen et al. (2003) also did similar experiment aimed for studying the strategy of playing the game and developed a computer model of a human player.

Jacob and Hochstein (2008) demonstrated that SET players prefer to look at perceptually similar cards, and, for the comparison of the cards, mainly rely on the perceptual processes such as similarity-detecting process. According to the authors, bias to the perceptual similarity and corresponding bottom-up processes can explain why players need less time to find lower-level sets than higher-level sets. Taatgen et al. (2003) also reached the conclusion that the perceptual elements play a greater role in finding lower-level sets. They suggested a strategy where a player looks at an arbitrary first card then at a second card that shares an attribute value. Next, the player predicts the third card and determines whether that card is one of the remaining ten cards. Taatgen et al. (2003) also hypothesized that the choice of the first card might not be arbitrary in some cases. They proposed that players try to find the set among the cards that have an attribute

value occurring in more than half of 12 cards. For example, if there are many red cards, it is attractive to search for a set among those cards. Taatgen et al. (2003) implemented this strategy in an ACT-R model. However, the data they collected did not have enough detail to determine whether subjects used such strategy.

Jacob and Hochstein (2008) proposed a generalization of the above strategy based on the notions of the most abundant value and the most abundant value group. The former refers to an attribute value that occurs most, and the latter refers to the group of cards that have the most abundant value. They found that the sets belonging to the most abundant value group are preferred to the sets outside of that group. In addition, the time required to find the set in the most abundant value group decreased as the size of the group increased. Most abundant value group was preferred to any other value group independently of the attribute type. Jacob and Hochstein (2008) suggested a *dimension reduction strategy* where players try to reduce the four-dimensional search space into three by choosing to look at cards that have one or more attribute values in common. It was assumed that dimension reduction strategy is primarily used with the most abundant value.

Research Objectives

Cognitive and Perceptual Processes

The dimension-reduction strategy is an example of a strategy that combines perceptual processing and goal-directed planning. Dimension reduction's gain in efficiency is due to the fact that the perceptual system is good at detecting similarity, but goal-directed planning is needed to decide what attribute value to focus on and for how long. Even though the earlier studies have established that dimension reduction is used, their methodology did not allow studying the dynamics within a trial. Moreover, not all sets can be found with that strategy. In particular, level

4 sets have no attributes in common, making them impossible to find with dimension reduction. To gather real-time behavioral data that can provide more insight into previously hidden aspects of user behavior, we decided to use eye tracking. Since many studies have shown that the eye-movement protocols directly or, at least, indirectly reflect both the cognitive processes and the amount of cognitive load (Kong, Schunn, & Wallstrom, 2010; Rayner, 1995; Salvucci, 1999), we considered the eye tracking a viable choice for studying human behavior.

Performance

Performance in SET is defined by how fast a player can find a set. Hence, speed is a major factor in the game. There can be different factors defining a player's speed. One of them is a strategy. This is the aspect of the game we are interested to explore. Taatgen et al. (2003) found that most players differ little in reaction times when it comes to finding lower-level sets. However, reaction times differ significantly in finding higher-level sets. One explanation for this effect might be that all players are likely to rely on general perceptual processes to find lower-level sets (Jacob & Hochstein, 2008). On the other hand, finding the higher-level sets may require strategies. As a consequence, we expect that slow players' eye-movements will be more guided by similarity between cards than faster players, because faster players' strategies will overrule the default similarity-based search.

We will not address how previous experience can affect the performance, or how subjects derive the strategies. One can get better at the game through practice, by naturally having better strategic thinking skills, or by just simply being good at pattern recognition. Learning in SET is a complex process and requires separate study.

Improved ACT-R Model

The ACT-R model created by Taatgen et al. (2003) was able to closely approximate the human player's reaction times. Its main drawback is that it fully predicts the third card given the first two cards it has looked at, and then searches for that card among the remaining cards. It therefore does not fully use a dimension reduction strategy, and also does not use perceptual similarity to find sets. In other words, it uses a pure top-down strategy. Our aim is to test whether a model with greater emphasis on perceptual elements of the game can explain the human data.

Experiment

Subjects

In total, 14 subjects participated in the experiment. The age of the subjects ranged from 20 to 30 years. All subjects were either students or staff members of University of Groningen. The subjects' previous experience with SET varied greatly: from a few played games to several years of experience.

Design and Procedure

Every subject was asked to do 60 trials. The group of 60 trials was the same for all subjects, but the order was determined randomly for each subject. Each trial consisted of 12 cards shown on a computer screen and arranged in an array similar to one shown in Figure 1. Each trial had exactly one combination of three cards that formed a set. Subjects were aware of this, but were not told about the level of the set. Subjects were asked to find a set and select the relevant cards with a mouse. A time limit of 180 s was given for each trial after which next trial was shown.

All 60 trials were randomly generated. In 30 trials one of the set cards was highlighted with a red border. These trials were distributed evenly over the four levels, with 7 or 8 trials of each level for each of the two highlighting conditions. The highlighted card belonged to the set

and served as a clue for the subject to find the other two cards. Subjects were aware about the meaning of a highlighted card. The presence of the highlighted card should make the task of finding a set much easier. In particular, it decreases the number of possible combinations from 200 to 55, and the number of possible pairs from 66 to 11. Since there are two pairs that lead to a set, in the worst case player will have to consider only 10 pairs. This is a six times reduction in complexity of the problem in terms of the search space. The main purpose of highlighting a card is that it provides a reference point on which we can base our eye-movement analysis.

Prior to an experiment, subjects were asked to do four warm-up trials to let them get familiar with experiment setup and with SET itself in the case that the subject had never played it before. Results from those trials were not included in the analysis.

Eye Tracking

An EyeLink 1000 eye tracker was used for recording the eye movements. It is a desktop-mounted remote eye tracker with monocular sampling rate of 500Hz and spatial resolution of $< 0.01^\circ$ RMS. The card images were shown on 20 inch LCD monitor with screen size of 1024×768 pixels and screen resolution of 64 pixels/inch. The card images had size of 124×184 pixels, or $4.02^\circ \times 5.95^\circ$. The horizontal and vertical distances between images were 80 and 70 pixels respectively, which constitutes to 2.59° and 2.27° . Angular sizes were calculated with an approximated viewing distance of 70 centimeters since subjects were given a certain freedom for head movement. The gaze position was calculated using the eye's corneal reflection captured with an infrared camera compensated for head movements. The eye tracker's default parameters were used to convert gaze positions into fixations and saccades. The calibration of an eye tracker was performed at the start and during the experiment, if necessary. A calibration accuracy of 0.8°

was considered as an acceptable measure. Before each trial subjects were asked to do a drift correction as an additional corrective measure.

Experiment Results

Reaction Times

In total, there were 29 trials where subjects failed to find the set, constituting 3% of all trials. Given this small proportion we treated them as response trials with a reaction time of 180 s. Figure 2 shows reaction times by level and highlighted condition. It shows that having a highlighted card as a clue more than halves the reaction time, and that the reaction time increases as the set level increases. This latter effect was also observed in previous studies (Taatgen et al., 2003; Jacob & Hochstein, 2008).

As it is shown in Figure 3a, subjects differed significantly by mean reaction times. As can be seen in the graph, all subjects were divided into three groups of fast, medium and slow players based on their mean reaction times. Figure 3b indicates that there is only a small difference in speed among three groups when it comes to finding a level one set. However, as level increases the differences between three groups also increase. This result is consistent with description of fast and slow players provided in Taatgen et al. (2003). Hence, we expected the groups to exhibit different behavioral effects despite the post-hoc division.

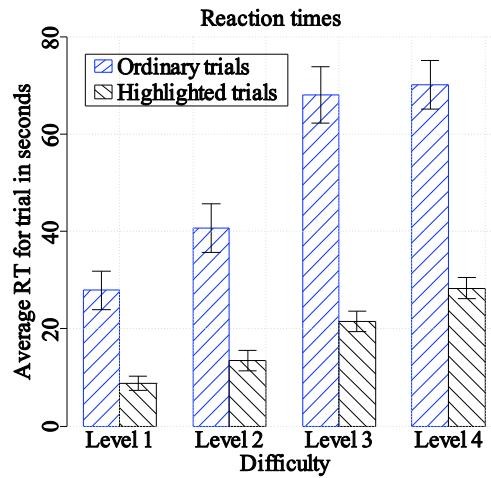


Figure 2. The mean reaction times with standard errors in ordinary and highlighted trials clustered by the levels and averaged over all subjects.

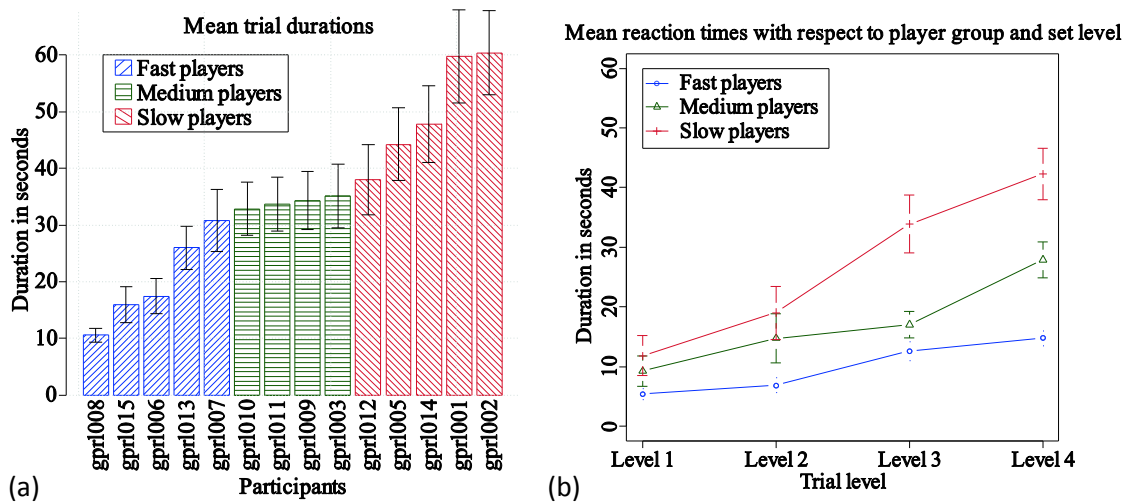


Figure 3. (a) Mean reaction times averaged over all trials for each subject. Subjects are divided into three groups of fast, medium and slow players. (b) Mean reaction times averaged over trials of the same level and player group.

Dimension Reduction

The reaction time analysis shows that subjects require less time to find sets with perceptually similar cards. This suggests that subjects apply a similarity-based strategy. Even though dimension reduction is such strategy, we want to investigate in detail to what extent this strategy is used. In this subsection we will examine evidence for the use of the dimension reduction strategy. If subjects used dimension reduction strategy then the corresponding scanpath should contain consecutive fixations on cards sharing at least one common attribute value.

To explore the existence of such pattern, the scanpath from each trial was transformed into labeled fixation sequences. Each card in a trial was assigned one area of interest with four different labels (see Figure 4).

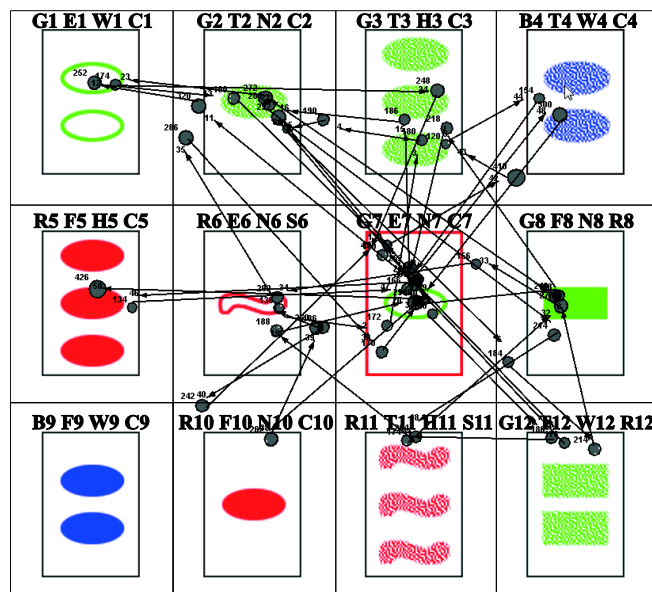


Figure 4. One of the problems shown to a subject. Card 7 is the highlighted card. Also shown are the fixations (circles) and saccades (arrows) produced by the subject. The outer thin black borders indicate 12 areas of interest. The four combinations of letters and numbers on top of each

card represent four labels for each of areas of interest. A set is formed by the fourth, fifth and seventh cards.

Each label describes one of the attribute values in a card and the position of the card in an array. For example, “G1”, “E1”, “W1” and “C1” are four labels describing the first card with values as green-open-two-oval. Then each fixation was tagged with four labels of an area of interest within which it falls. The consecutive fixations on the same area of interest are considered as a single fixation, and the corresponding fixation durations are summed. Combining all labeled fixations of a common attribute type into fixation sequences produces four distinct sequences for each trial.

An analysis of the fixation sequences revealed the existence of a pattern of fixations related to the usage of dimension reduction. We will demonstrate this using the example problem from Figure 4.

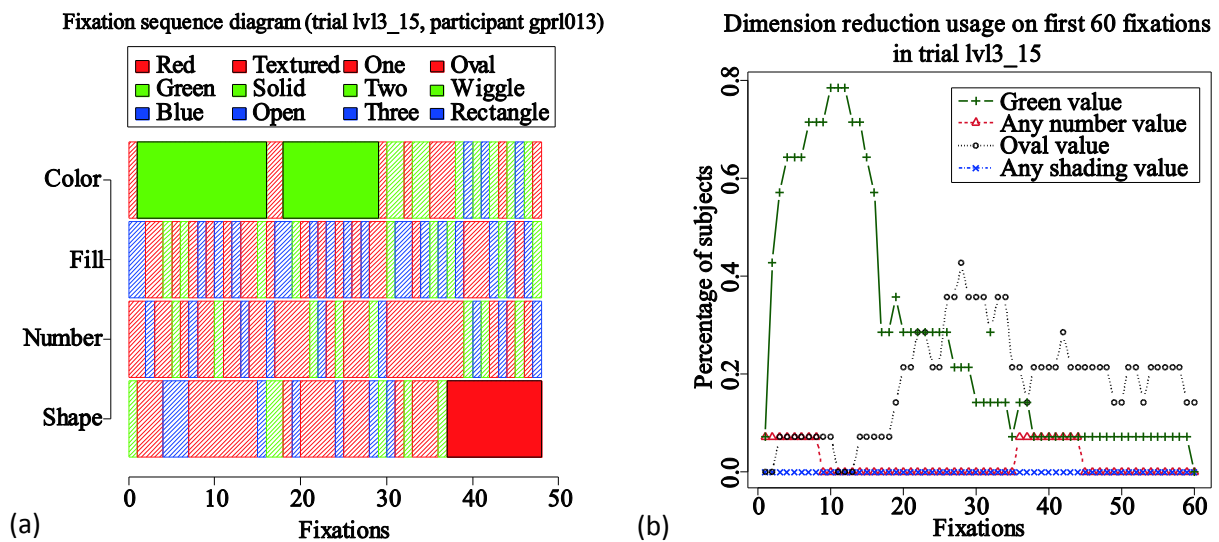


Figure 5. (a) Single subject's fixation sequence diagram for trial "lv13_15"; (b) Changing proportion of subjects who used dimension reduction in trial "lv13_15" as a function of fixation position in the sequence and attribute value.

Figure 5a shows a fixation sequence diagram produced from the scanpath shown in Figure 4. Each horizontal lane in the diagram shows a subject's fixation sequence with respect to the particular attribute type. One unit on the x-axis represents a fixation on one particular card, while the corresponding bars on four lanes represent attribute values of that card. In the diagram the labels are color coded according to the corresponding attribute value. The consecutive fixations on the cards with the same attribute value are shaded with a solid color if the probability of such a fixation subsequence occurring by chance is equal to or below 0.05 (refer to the Appendix for details of calculating the probability).

From the figure we can see that at the beginning of the trial, the subject looked at green cards, and, by the end, at cards with an oval shape. We can conclude that the subject used dimension reduction strategy at least two times and each time with respect to a different attribute value: green and oval. The fixation pattern for this trial is not unique for this particular subject. Figure 5b shows the proportion of all subjects that used dimension reduction with green and oval values. This proportion is also contrasted against proportions of subjects that used dimension reduction on any of the three values from either number or shape attributes. The figure shows that at the start of the trial subjects preferred to search for a set among green cards and later switched to a group of cards with an oval shape while mostly ignoring all other values.

Effects of an attribute type on dimension reduction. According to Jacob and Hochstein (2008), dimension reduction primarily occurs with the most abundant value. However,

it can be observed from Figure 5b that majority of subjects prefer the group of green cards to the group of cards with an oval shape despite the fact that the latter has the most abundant value. This suggests that the type of the attribute also plays a role in deciding the value to be used for dimension reduction.

To find an effect of an attribute type, we have calculated an average proportion of fixation sequences where all subjects used dimension reduction for all problems. The result indicates that blocks of fixations with the same attribute value occupy on average 46% and 35% of an overall fixation sequence in trials with and without highlighted card respectively. Note that these estimates are on the conservative side, because some sequences may not have been recognized because they cannot be distinguished from a random sequence, either due to wandering fixations, sequences that are too short, or inaccuracy in the eye tracker.

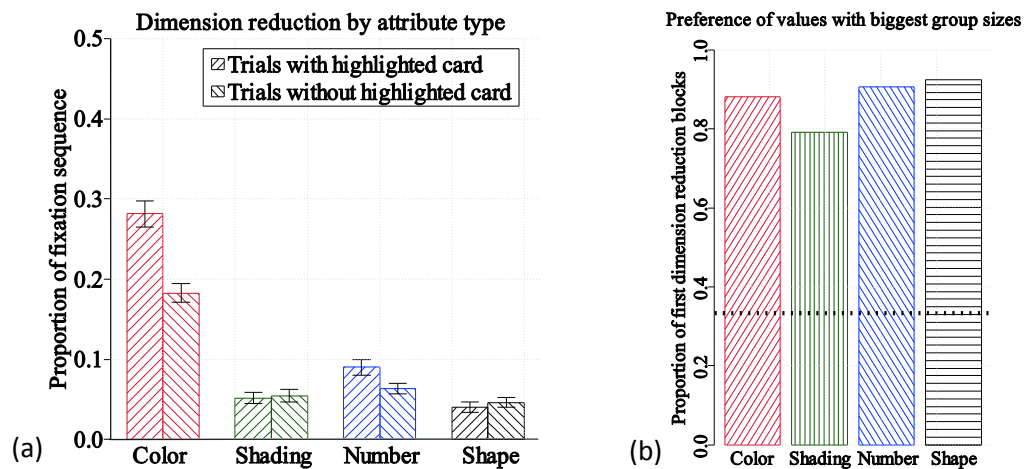


Figure 6. (a) Mean proportions of attribute types used in similarity-based scanning. Proportions are shown separately for trials with and without highlighted card. (b) Proportion of trials where subjects preferred to use a value for dimension reduction with the biggest group size among

other values of the same attribute type. The horizontal dashed black line indicates the expected proportion if the choice was made randomly.

Figure 6a shows how use of dimension-reduction distributes over the four attribute types and reveals an effect of overall attribute preference. Subjects are two times more likely to look at the group of cards with the same color than any other attribute. The distributions of the most abundant values in the 60 trials among Color, Shading, Number and Shape were 28, 27, 19 and 26 respectively². According to such, the corresponding bars on Figure 6a should have nearly equal height if choice of a value was dependent only on group size. This is not the case. The results suggest that the four attribute types have different saliency properties with color being the most salient while shape and shading being the least salient attributes.

There is still an effect of the most abundant value within each attribute type. This means that among the three values of the same attribute type the most abundant values is preferred for dimension reduction. As Figure 6b indicates, in 85% of all trials subjects prefer the most abundant value over the other two values of the same attribute type³. The trend is consistent among all four attributes.

Effect of dimension reduction on performance. There is also a difference between fast and slow players in how they use dimension reduction strategy. Figure 7a shows how the usage of dimension reduction strategy changes over time in trials with a highlighted card. There is a general trend among players to use dimension reduction at the beginning of a trial and gradually stop using it over time. It suggests that players gradually switch from a dimension reduction to some different strategy. Furthermore, the graph suggests that slow players are more likely to stick to dimension reduction longer than fast players.

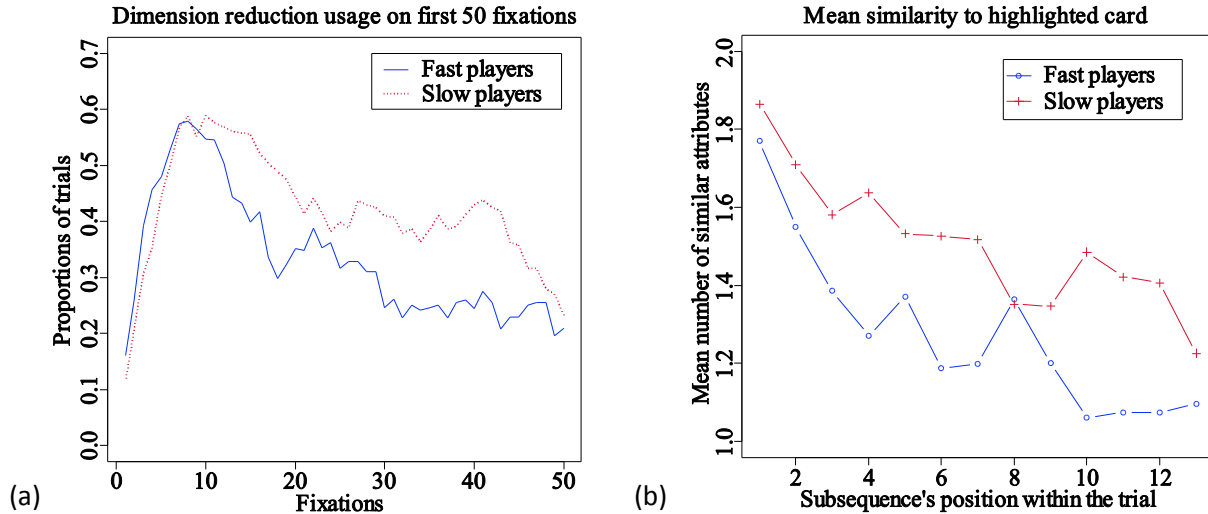


Figure 7. (a) Changing proportion of trials in which dimension reduction was used. The proportions are calculated as a function of the fixation position within a trial. The proportion on fixation x is calculated by counting the trials that have a dimension reduction block that include fixation x . (b) The mean overall similarity of all cards in a particular subsequence to the highlighted card.

Dissimilarity-based Search

In the previous section we have seen that subjects use a dimension reduction strategy to reduce the complexity of finding a set. However, it is not yet clear how a similarity-based approach can eventually find sets with many different attribute values. The fact alone that subjects were able to find level 4 sets, in which all attribute values are different, proves that the strategies they use are not limited to dimension reduction. In fact, Figure 6a has already shown that subjects use dimension reduction strategy only 46% of the time, even though this number

may be conservative given our analysis method. Figure 7a also suggests that players switch to a different strategy.

It is our assumption that subjects gradually switch from a similarity-based strategy to a dissimilarity-based strategy. It should be possible to observe this switch from one strategy to another in fixation sequences produced from trials with highlighted cards.

Search subsequences. The next analysis involves only trials with a highlighted card. Preliminary inspection of the data revealed that subjects refixated on a highlighted card approximately every five fixations, presumably to refresh their memory and to restart a new search subsequence. For example, the following labeled fixation sequence “**4**-7-11-10-3-7-2-11-**4**-3-10-2-5-9-5-6-**4**-7-5-8-**4**”, with **4** being a fixation on a highlighted card, can be broken down into three subsequences. In a similar manner fixation sequences corresponding to other three attributes can be broken down into identical subsequences.

Breaking down a trial into separate subsequences allows us to analyze how a mean perceptual similarity of fixated cards to a highlighted card changes with each subsequence (Figure 7b). The calculations were done separately for slow and fast players. There is a general tendency to look at a less similar card with each new fixation and each new subsequence. When players start a search they seem to prioritize cards based on decreasing similarity to a highlighted card. Furthermore, Figure 7b suggests that with each new search subsequence subjects lower the similarity threshold and include in their visual search less similar cards that were not included in the previous subsequences. Finally, there may be a difference between fast and slow players in terms of bias to similarity-based search as Figure 7b indicates. Fast players appear to abandon similarity-based search earlier than slow players.

Using a mixed-effect regression analysis (Baayen, Davidson & Bates, 2008), we have further investigated how the tendency to look at perceptually similar cards changes during the trial. The dependent variable in the regression is the perceptual similarity of each fixated card to the corresponding highlighted card (the values on the y-axis in Figure 7b). The following fixed effects were used: *Subsequence* is a log-transformed position of a subsequence in a fixation sequence. *Fixation* is a log-transformed position of a fixation within a subsequence. Variable *RT* is subject's mean reaction time in seconds shown in Figure 3a. In addition, two random effects on an intercept, *Subject* and *Trial*, were added each representing subjects and trials respectively.

Table 1. The fixed effects' coefficients, t and p values.

Fixed Effects	Coefficients	Standard Errors	t values	p values
Intercept	1.7376	0.0754	23.036	0.0001
Fixation	-0.1761	0.0158	-11.113	0.0001
Subsequence	-0.1634	0.0110	-14.812	0.0001
RT	0.0046	0.0011	4.106	0.0012
Fixation:Subsequence	0.0334	0.0072	4.647	0.0001

Table 2. Variances and corresponding standard errors of random effects.

Random Effects on Intercept	Variances	Standard Errors
Trial	0.1037	0.3221
Subject	0.0024	0.0495

Resulting coefficients for fixed main and interaction effects are shown in Table 1. The

table also presents corresponding t and p values for fixed effects. The variances and standard errors of the random effects are depicted in Table 2.

In the interpretation of coefficients we are mainly interested in their signs. Positive coefficients increase perceptual similarity to the highlighted card. Hence, the corresponding independent variables promote the similarity-based search. The negative coefficients decrease perceptual similarity. Therefore, the corresponding independent variables facilitate the transition from the similarity-based search to dissimilarity-based search.

Both *Fixation* and *Subsequence* have negative coefficients supporting our assumption that over time cards that subjects look at decrease in similarity to the highlighted card. The significant main effect for *Fixation* indicates that transition occurs not only within fixation sequence as whole, but also within individual subsequences. An interaction effect between *Fixation* and *Subsequence* has positive coefficient. The interaction effect provides a threshold for the main effect of *Subsequence* after which subject cannot look at less similar cards anymore. It makes sense since it is impossible to look at cards that have more than four dissimilar attributes.

There is a strong correlation between subjects' mean reaction time and the tendency to look at cards similar to highlighted card. The variable *RT* serves as a strong predictor. Its coefficient's sign indicates that slower players are more biased toward similarity-based search than faster players. And this bias increases as mean reaction time increases.

Experiment Discussion

Experiment Results' Summary

The mixed-effect regression analysis of the fixation sequences indicate that the subjects' basic strategy of playing SET is similarity based. Subjects prefer to look for a set among the cards that are similar to each other.

One specific instance of a similarity-based strategy is the dimension reduction strategy (Jacob & Hochstein, 2008). The dimension reduction strategy can be used more than once (Figure 5 & Figure 6) within the same trial and each time with different attribute value. The player chooses one attribute value, to which we refer as a *guiding value*, and starts looking for a set among the cards that share that value. If a player fails to find a set with the current value, then another guiding value is chosen, and the new group of cards is defined as a next search space.

The overall strategy of dimension reduction is top-down, but the choice of a guiding value is heavily influenced by two bottom-up elements: (1) the size of the group of cards that share the value and (2) its attribute type. The importance of group size (Figure 6b) was also found by Jacob and Hochstein (2008). However, contrary to their conclusion, we have found that an attribute type also plays an important role (Figure 6a) in choosing a guiding value. In particular, color is preferred to any other attribute type, while shape and shading are the least preferred attribute types. This result coincides with other studies concluding that people prefer to operate on colors rather than on shapes (Kieras, 2010; Kim and Cave, 1995; Pomplun et al., 2001). The number attribute also seems to be preferred to shape and shading, at least in trials with highlighted cards. The presence of a highlighted card can bias players to values of that card. Such bias can override an effect of a group size or even attribute type.

Another interesting finding is the fact that within a trial subjects decrease the use of dimension reduction strategy. This reduction (Figure 7a) nicely coincides with gradual reduction in reliance on similarity (Figure 7b). As the game progresses, players increasingly look at more dissimilar cards more suitable for finding higher-level sets.

It seems that all players follow more or less these strategies. However, there are subtle differences between fast and slow groups of players. We found that fast players are less

dependent on similarity than slow players (Figure 7b and Table 1). Fast players are initially less likely to use dimension reduction and switch faster to the dissimilarity-based search than slow players.

Additional Assumptions

There are still open questions that were not answered by the data analysis. For creating a plausible model of a SET player it is essential that we have a complete picture of a player's behavior. In this section we address the essential but missing aspects of a SET player's strategy by referring to relevant literature or making our own assumptions.

The two critical aspects of finding a set are reducing the search space by selecting an appropriate guiding value, and the search strategy itself once a guiding value has been selected.

Choice of a guiding value. Although the decision to choose a guiding value is top-down, the choice itself, we assume, is not top-down. This choice is defined by two components: a static task-independent component that defines the saliency of an attribute value in the visual field and task-dependent factors, some of which change while the search for a set progresses.

Task-independent components include attribute type and group size. The four attributes have different inherent saliency properties. The color is the most salient attribute type, and the number is more salient than shape or shading (Kieras, 2010; Kim and Cave, 1995; Pomplun et al., 2001). On the other hand, six green cards are more salient than four red cards because of an effect of group size on the saliency. These factors are not dependent on the current goal and are inherent properties of the visual object and the visual scene as a whole.

Task-dependent components include the presence of a highlighted card and the current progress within a trial. The task for the player is to find a set that includes the highlighted card (if it is present). This connection of a highlighted card to the current task increases the relevancy of

the attribute values in the highlighted card. The relevance of an attribute value, however, decreases once we have already tried to find a set with that attribute value. So, if the player has not been able to find a set among the green cards then the task relevancy of the green value decreases. This decreasing relevance can explain why the similarity of attended cards to the highlighted card decreases: once particular attribute values have been tried as a guiding value, their relevance decreases and other, more dissimilar values are selected to guide search.

For example, in the beginning of the game, most players tend to focus on the group of cards that share particular color or number values, since color and number are the most salient attribute types. However, their relevancy will decrease over time, and eventually a player will focus on other attribute types.

Strategies and within-trial strategy shifts. As described earlier, the data suggests a gradual shift from dimension reduction to a dissimilarity-based strategy.

However, so far we have no concrete evidence for the mechanisms behind such a strategy shift. One option is that there is an explicit meta-cognitive process tracking the current state of the game and timing the strategy shifts. However, a far more elegant and simpler explanation would be one in which a strategy shift occurs implicitly as a result of changing relevance of the attribute values as they are used as a guiding value. The second option does not require an explicit process of tracking current state and timing strategy shifts. The mechanism that chooses the guiding values, outlined in the previous section, does exactly that: initially, the attribute values of the highlighted card will dominate the choice of guiding value and will therefore lead to similarity-based search. However, once those values have been tried, their relevance diminishes, and other values are chosen that are not attributes of the highlighted card. This will lead to a dissimilarity strategy in which a third, dissimilar card will be necessary to complete the

set.

Strategy implementation. Once a guiding value is chosen, a search process is needed to try to find a set using the guiding value. There are two basic strategies to do this: the first is to, in addition to the highlighted card, pick a second card on the basis of the guiding value, and then pick a third card that is perceptually similar to the second card. At that point, the three cards can be compared to see whether they constitute a set. Even with a highlighted card this search process is potentially expensive, because there are still 55 possible combinations to check. The use of a guiding value is helpful to look for the most promising combinations first, especially combinations that are potential lower-level sets.

The second strategy is to select a second card in addition to the highlighted card, and predict what the third card should be. After making the prediction, the predicted card may or not be present among the remaining cards. If it is, it completes the set. This strategy is much more efficient, because there are only 11 combinations of the highlighted card with a second card, and two of those will complete a set. Even when there is no highlighted card, the prediction strategy is more efficient than the similarity strategy, because there are only 66 possible pairs, three of which are part of the set, but 220 combinations of three cards. However, the prediction strategy is more effortful and requires at least some experience with the game to be successful.

Competitive parallelism of the two strategies. Even though we can identify two distinct strategies, several hybrid combinations are possible. For example, instead of predicting all attribute values of the third card, it is possible to only predict two values and use these two to guide the similarity strategy. In fact, both strategies and all possible hybrids can be produced if we assume two parallel processes, a bottom-up process that scans cards based on similarity, and a top-down process that makes a prediction for the third card. This idea is consistent with the

threaded cognition theory of multitasking (Salvucci & Taatgen, 2008), a bottom-up visual scanning and a top-down prediction task run in parallel collaborating but also competing to achieve the same goal.

Competitive parallelism assumes that all players have two parallel processes independently of player's proficiency. Slow players know how to predict, but they are not good at it, so typically the visual-scanning process will dominate performance. Faster players are proficient enough to make fast and accurate predictions, so the prediction process can keep up with visual scanning, making targeted search of a predicted card possible rather than just scanning on the basis of similarity.

Competitive parallelism provides advantages over a pure sequential strategy. It provides a means for a more objective comparative evaluation of efficiency of one process over another. It prevents a one-sided choice of one process over another even if one is less efficient. The less efficient process has a chance to become more cost effective with training and rehearsal. Competitive parallelism actually provides an opportunity for slow players to become faster, because even a partial prediction (i.e., two attributes instead of all four) already provides an advantage over pure similarity-based search.

Prediction works at a more conceptual level, and therefore requires a certain degree of proficiency that slow players may lack. Prediction is more beneficial in finding higher level sets in contrast to sequential perceptual comparison. However, it may provide little leverage against parallel bottom-up similarity detection in lower level sets. Those differences can explain why slow and fast players differ little in finding lower-level set and differ significantly in finding higher-level sets.

An ACT-R Model of a SET Player

Briefly About ACT-R

We have implemented the model using the ACT-R³ cognitive architecture (Anderson, 2007). ACT-R has a modular organization where each module is dedicated to a distinct type of cognitive resources (visual, motor, etc.). Factual knowledge in ACT-R is represented by chunks with slots where other chunks serve as slot values. Each module has its own buffer where either new chunks can be created or existing chunks can be passed on. Three modules that are important for this paper are described next.

The visual module handles visual mechanisms such as perception, attention shift and encoding of visual stimuli. Visual stimuli are represented in form of chunks within the visicon, a virtual imitation of a screen visible to model. This module cannot create new chunks, but rather “perceives” chunks within the visicon by placing them in its buffer.

Every chunk that has been cleared from any buffer is stored in declarative memory (DM) module and can be retrieved again. The DM module can retrieve only one chunk at a time, which is stored in the module’s buffer. Each chunk in DM has a base-level activation value, which represents frequency and recency of use (e.g., Anderson and Schooler, 1991). A chunk’s activation in DM can also be influenced by chunks contained in buffers at the time of retrieval, by a spreading activation mechanism. Based on activation the module computes the probability and time cost of retrieving a chunk from memory. We have implemented an additional extension to the ACT-R visual module, which enables chunks in visicon (i.e., the whole visual field) to spread activation to chunks in DM in the same manner as the chunks in buffers do.

Lastly, there is a problem state module that serves as a working memory. This module is unique since it can create new chunks that are not perceived in the environment nor retrieved from DM. Slot values from chunks in other buffers can be used as values for the new chunk’s

slots. However, creating a new chunk is a time costly process that takes 200 ms, a parameter in the architecture that is typically not changed.

The architecture provides an essential set of parameters by default including, but not limited to, times it takes to move the mouse, retrieve a chunk from memory or encode a visual stimulus. It also provides set of adjustable parameters and range of recommended values for each of those parameters. These elements of the architecture have received extensive experimental support (e.g., Anderson, 2007 and see <http://act-r.psy.cmu.edu/>).

Model Design Decisions

We will now describe the solution method that we outlined in the previous section in more detail.

Threads. The model consists of two parallel processes (threads; see Salvucci & Taatgen, 2008) reflecting both top-down and bottom-up nature of a task. A bottom-up thread is responsible for visual processes such as choosing a scanpath or shifting attention from one card to another. The top-down thread is responsible for higher-level processes such as deciding a guiding value and comparing cards. Both threads can influence each other's processes indirectly. For example, the top-down thread chooses a guiding value based on what has already been tried earlier in the trial. However, bottom-up features such as what cards are visible or which card is being fixated also influence the choice.

Algorithm for general strategy. The model largely follows strategies that we have deduced from the data and the assumptions we made in the previous section. The following is the description of model's general strategy:

1. Focus attention on the highlighted card HC.
 - a. Let $Card_{HC}$ be a set of four attribute values in the highlighted card.
2. Retrieve any attribute value V_{DM} from declarative memory.

- a. Let A_V be the attribute type of V_{DM} .
3. Pick the attribute value V_{HC} from $Card_{HC}$ that also has A_V as attribute type.
4. If $V_{DM} = V_{HC}$ then use dimension reduction.
 - a. Define search space G as a group of cards that have V_{HC} .
5. If $V_{DM} \neq V_{HC}$ then use dissimilarity strategy.
 - a. Define search space G as a group of cards that does not have V_{HC} .
6. Start comparison cycles on G to search for a set (depicted in Figure 8).
7. If a set is not found then go back to step 1.

Implementation of both strategies in the model is liberal in a sense that model behavior is not hardcoded. There is no explicit control over the guiding value choice. Neither there is an explicit top-down control over strategy shift. The model decides all specific details of those steps on the fly based on a visual scene and progress of a current trial. Steps 2 and 6 are most important. The outcome of step 2 defines the strategy to be used, while in step 6 bottom-up and top-down threads run in parallel each trying to find a set separately.

Saliency and relevancy. This subsection describes how model takes step 2 of the algorithm. The attribute value that is the most salient and relevant at the time is chosen as the guiding value V_{DM} . Saliency is a constant feature within a trial, however, relevancy is not and calculated each time a new V_{DM} needs to be chosen. Within the model we have used ACT-R's activation mechanism to mimic both saliency and relevancy. Activation depends on several parameters such as values of a highlighted card, number of times the attribute value was used previously, the last time it was used, etc.

The two main parameters defining saliency are attribute type and the size of the group of cards that have that value. Color is generally the most salient attribute type followed by number,

while shape and shading are the least salient types. Attribute type saliency is simulated using ACT-R's chunk referencing mechanism (Table 3).

In order to model the effect of the group size, we used a logarithmic function (see Table 3) to map the number of occurrences of an attribute value i in the visual field onto a group size factor fan_i . This mapping is similar to the spreading activation mechanism in ACT-R's declarative memory.

The relevancy of a value depends on whether it appears on a highlighted card and whether it was used previously. The highlighted card spreads additional activation to each value it has. The relevancy of a value is temporarily inhibited after it has been used and no set was found. The time and duration of the inhibition are calculated according to Lebiere and Best's (2009) short-term inhibition equation. The complete description of the parameters used in calculating the activation is shown in Table 3.

Table 3. Parameters for calculating activation for an attribute value i .

Parameter	Influence On Activation	Implementation method
Attribute type	positive	<p>Base-level activation B_i is calculated for each attribute value based on initial number of references it is assigned. An initial number of references (n) is set for each attribute type as following (higher number results in higher activation):</p> <ul style="list-style-type: none"> • Color chunks: 40 • Number chunks: 36 • Shape chunks: 32 • Shading chunks: 28 <p>An exact calculation was used with the decay rate of base-level learning (d) set to default value of 0.5. (t_j) is the elapsed time since the chunk has been used for the j-th time.</p>
Group size	positive	Custom extension for ACT-R that spreads activation from the visual field to the DM. The associative weight parameter (W) is set

		to 0.7. fan_i - is a measure of how many chunks in the visual field are associated with chunk i . Higher fan_i results in more activation spread to value i : $G_i = W * \ln(1 + fan_i)$
Highlighted card	positive	ACT-R's equation for a spreading activation from a visual buffer. (j) indicates to a value in j -th slot of a chunk that is in visual module buffer. (fan_{ji}) is a measure of how many chunks in DM are associated with value in j -th slot. Higher fan results in less activation spread to value i . Maximum associative strength (S) is set to 4, a sufficiently high value to prevent negative spreading activation. (W_j) is the amount of activation to be spread from in value j -th slot to value i if two are associated and set to 0.13. $S_i = \sum_j W_j * (S - \ln(fan_{ji}))$
Frequency of use	negative (inhibitive effect)	ACT-R extension for a base-level inhibition is used with short-term decay rate (d_s) and time scaling (t_s) parameters set to 1 and 10 respectively as recommended by Lebiere and Best (2009). $I_i = \log(1 + \frac{t_n}{t_s})^{-d_s}$
Latency of use	negative (inhibitive effect)	
Random noise	positive	ε_i - ACT-R's transient noise generated from logistic distribution with mean 0 and with ans parameter set to 0.1. This noise ensures that model's behavior differs each time even if presented with exactly same trial and starting conditions.

Values for most of the constants mentioned in Table 3 are taken from the range of recommended values mentioned in ACT-R literature (see <http://act-r.psy.cmu.edu/publications/>). However, we fitted the four initial numbers of references for attribute types. Two other parameters that required fitting are the associative weight parameter W and spreading activation amount W_j . The first parameter defines scale of influence of a group size, and the second one defines scale of an influence of a highlighted card.

Combining all parameters from Table 3 results in following equation for calculating activation for attribute value i : $A_i = B_i + S_i + G_i - I_i + \varepsilon_i$. The value with the highest activation is chosen for retrieval from declarative memory. The time cost of retrieval is calculated via ACT-R

equation: $Time = Fe^{-A}$ where A is an activation value and F is the latency factor set to 0.2, a value most commonly used in other models.

Top-down versus bottom-up processes in comparison cycles. After deciding which strategy to use, the model proceeds by scanning a chosen group of cards. This is described as a step 6 in the algorithm. Individual steps of scanning are described in Figure 8. The entire scanning can be divided into comparison cycles. In each cycle the model picks two cards, further referred to as C1 and C2, to compare to the highlighted card. The model first chooses C1, and then C2. In each cycle the model picks as a C1 a card that was not chosen as C1 before. Hence, the number of cycles is the same as the number of cards that match the scanning criteria.

The order in which cards are chosen as C1 is mostly defined by the order in which those cards were fixated since the scanning began. Earlier fixated cards have higher chance of being chosen as C1. The model is free to choose its own scan path with the only restriction that it will not refixate on the cards it fixated before until all other cards have been fixated.

Two different approaches are used in parallel to make the decision about C2: bottom-up and top-down. In the bottom-up approach, the model continues scanning the search space and compares the first fixated card with the highlighted card and C1 (a box in Figure 8 denoted *Wait for bottom-up scanning to return C2*). At the same time, the top-down approach tries to make a prediction about C2 based on the available rules (a box in Figure 8 denoted *Predict C2 values*). It generates the abstract representation of C2 and asks the visual thread to find the card matching that representation. The success and completeness of the prediction depends on availability and accessibility of prediction rules. Both approaches compete with each other. The approach that requires less time is favored over the other. In other words, if the model is able to make a prediction before the visual thread fixates and encodes some card as C2 then prediction is

avored.

Given all three cards, the model verifies if the cards really make a set. If cards do not make a set then the model goes back to visual scanning. If a set is still not found then model interrupts the scanning and refixates on the highlighted card to choose another guiding value. Due to limited number of cycles and the liberal way the model chooses C2, the search is not exhaustive, and model can fail to find a set even if search space contains it.

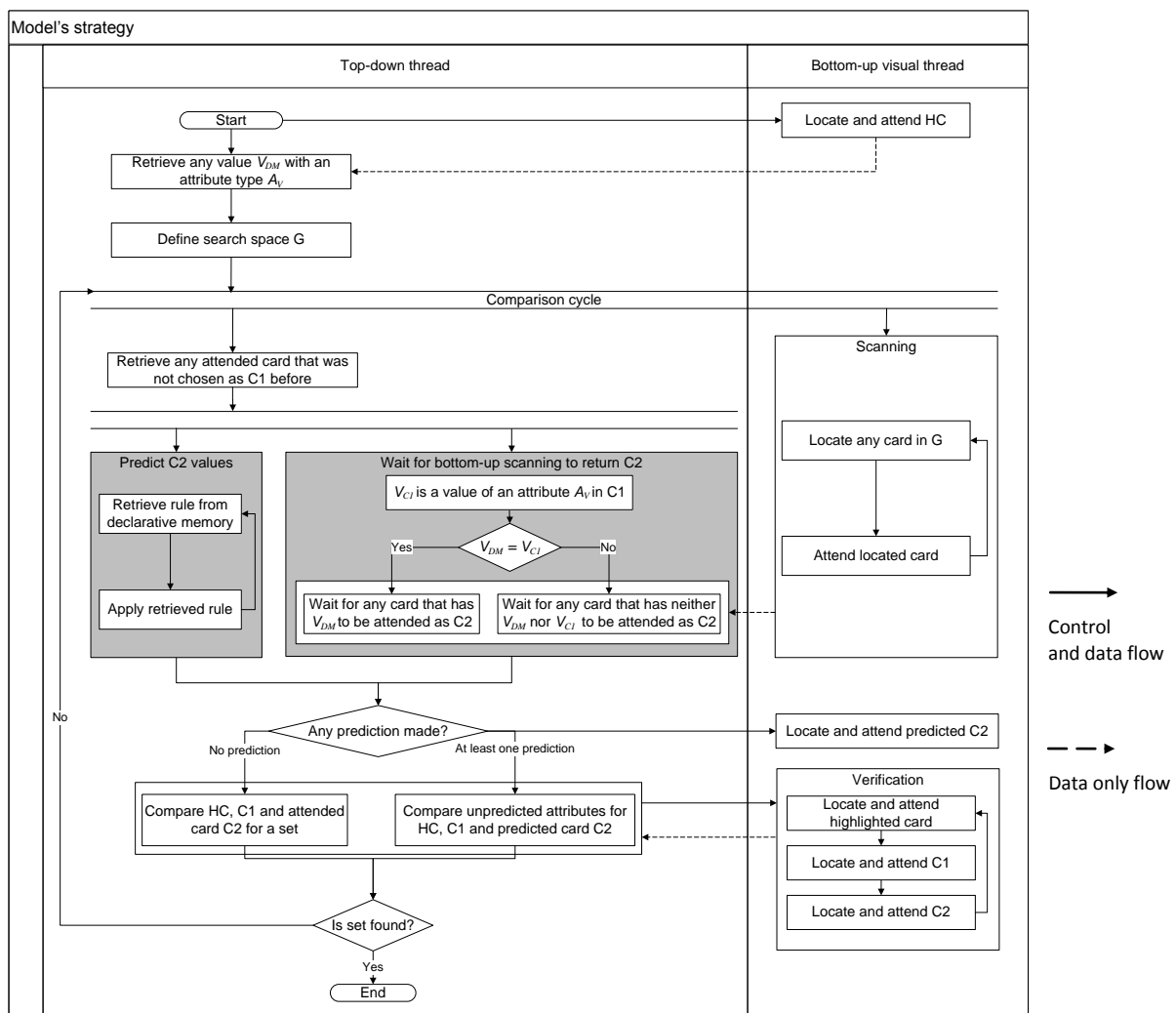


Figure 8. An algorithm for searching for a set given specified group of cards G . Two shaded

boxes represent two approaches that model uses in parallel to find a set. The right shaded box shows the bottom-up approach to find a set, and the left shaded box shows the top-down prediction approach.

Prediction rules. Predictions are made based on prediction rules. Rules are declarative chunks that have to be retrieved from memory when necessary. An example of such a rule is: $\text{Given}(\text{Textured}, \text{Solid}) \Rightarrow \text{Expected}(\text{Open})$. It should be noted that $\text{Given}(\text{Solid}, \text{Textured}) \Rightarrow \text{Expected}(\text{Open})$ and the previous rule are treated as different ones. There are also rules for similarity such as $\text{Given}(\text{Red}, \text{Red}) \Rightarrow \text{Expected}(\text{Red})$. In total, the model can have 36 rules: nine rules for each attribute.

Model Results

In each trial, the model is presented with 12 cards. One card is always highlighted indicating that it belongs to a set. The model has to find the other two cards forming a set. The same 30 trials from the experiment with human subjects were used.

We created eight versions of the model. The only difference between model versions was the availability of prediction rules. The first model had no prediction rules in declarative memory. The second model had 12 prediction rules for predicting similarity of the corresponding 12 attribute values (e.g., $\text{Given}(\text{Red}, \text{Red}) \Rightarrow \text{Expected}(\text{Red})$). The third model had 16 rules: 12 similarity rules and four rules for predicting dissimilarity, one for each attribute. The number of available rules in subsequent models was increased in a similar manner by four. Each version of the model was run 10 times.

Reaction times. Figure 9a shows reaction times for all eight models averaged over four difficulty levels. As was hypothesized previously, the model's reaction time gradually decreases

as model becomes better at making predictions. The model with zero rules is the slowest model and the model with all rules is the fastest model. For low-level sets, there is little difference in RT between different versions of the model. However the difference is quite high for trials with high-level sets. This effect resembles the one found in human data. Overall, Figure 9a suggests that the main boost in performance through predictions is produced by trials with high level sets.

In Figure 9b, the mean reaction times (dashed lines) of the fastest and slowest models are compared to the mean reaction times (solid lines) of corresponding fast and slow groups of human players. As it can be seen, the models closely reproduce reaction times of both slow and fast human players. The fixation sequences produced by these two models were further compared to human data from fast and slow groups.

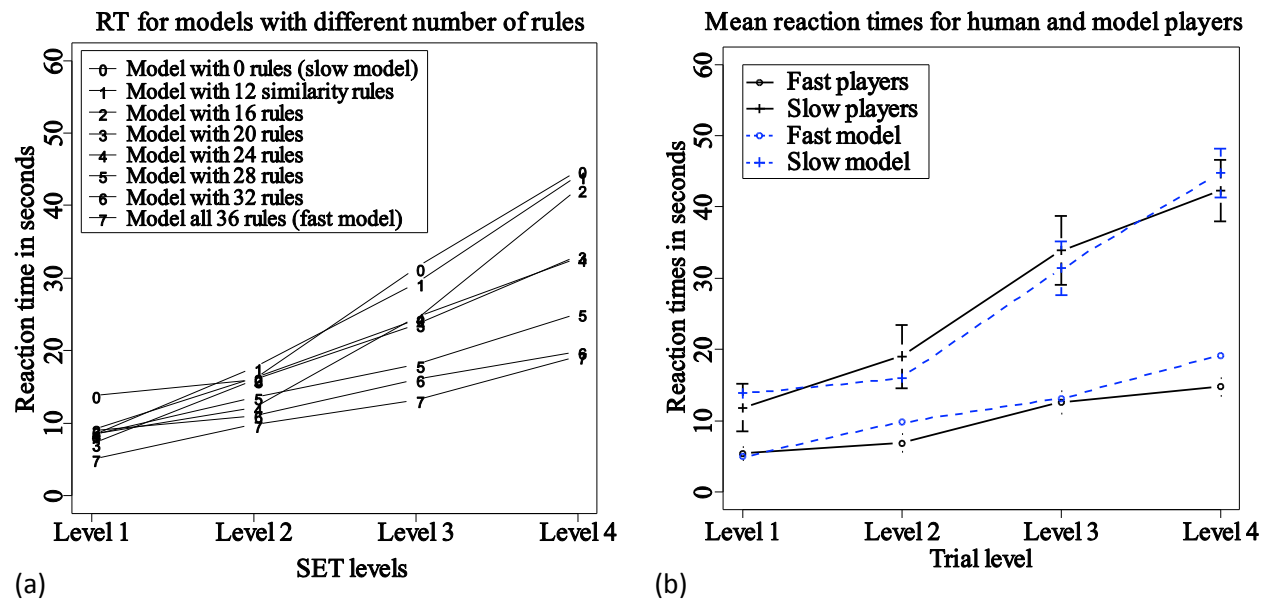


Figure 9. (a) Reaction times of eight models averaged over four difficulty levels. (b) Reaction times of the slowest and fastest models comparing to the reaction times of the human players in trials with highlighted cards.

Dimension reduction. Both the fast and slow models are quite good at replicating the subjects’ tendency to use dimension reduction and preference to certain attribute values. As an example, the fast model’s (Figure 10a) and the subject’s (Figure 5a) fixation sequences from the same trial are compared. The model’s fixation sequence closely resembles the sequence produced by the subject. At the beginning of the trial, the model also preferred to look at the green cards and later on switched attention to a group of cards with oval shape in a same manner as human subjects did. This decrease is consistent with behavior of the human subjects.

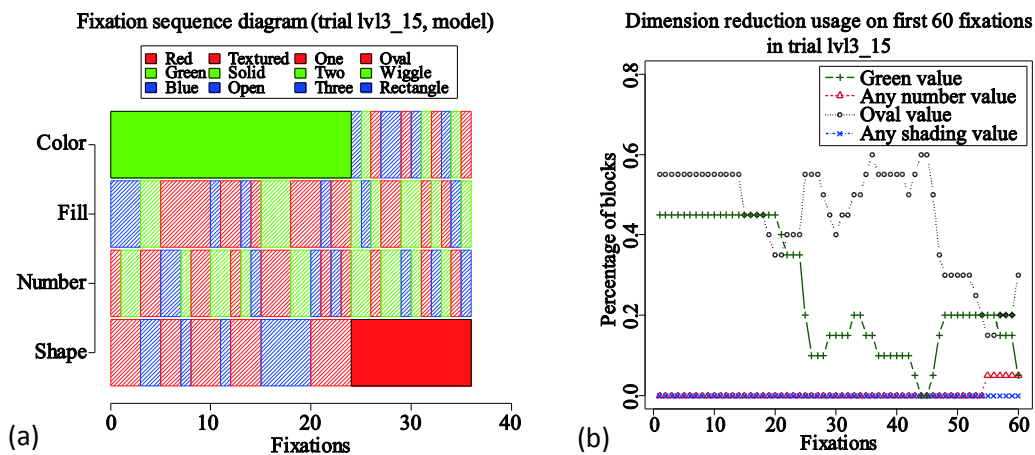


Figure 10. (a) Model’s fixation sequence diagram for trial “lv13_15”. (b) Changing proportion of blocks where the models used dimension reduction in trial “lv13_15”. Proportions are calculated from both slow and fast models’ data. Proportions are shown as a function of fixation position in the sequence and attribute value.

It is obvious from multiple model runs that half of the times the model prefers to look at the green cards in the beginning of the trial although they form the second largest group after cards with an oval shape (Figure 10a). Nevertheless, the fact that color is the most salient

attribute type is enough to compensate for a smaller group size. Defining separate saliency values for attribute types works quite well for modeling players' bias to an attribute types.

It can be observed from Figure 10 that the model favored shape in the later stage of the game, which is the least salient attribute type. This is due to the effect of a group size. Oval shape compensates its inherent lack of saliency with bigger number of occurrences. The fact that oval value provides strong competition to green value even at the beginning of the trial suggests that the effect of a group size is stronger than it should be (compare Figure 10b to Figure 5b).

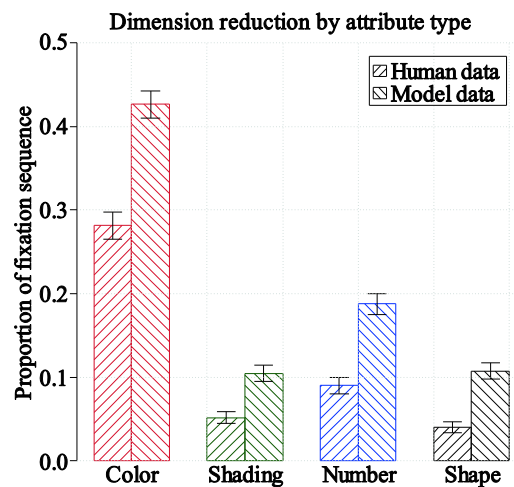


Figure 11. Mean proportions of attribute types used in similarity-based scanning. The overall values of all subjects' trials with the highlighted card are compared to the overall values of both models' trials.

Overall, the saliency and relevancy mechanisms work well in modeling subjects' strategy to use dimension reduction. Combined data from both models shows similar order of preference for the attribute types as the human subjects. Figure 11 shows that, in general, models clearly

prefer color and number while they make little difference between shape and shading. Both models gradually stop using dimension reduction if it fails to find a set (Figure 12a). This behavior is again consistent with behavior of human subjects. However, models are more dependent on dimension reduction strategy than the human subjects. We attribute this difference to the difference in manner of scanning between model and human subjects. We discussed earlier that human subjects can get distracted and produce wandering fixations in the middle of the scan. On the other hand, model is precise and does not produce such fixations.

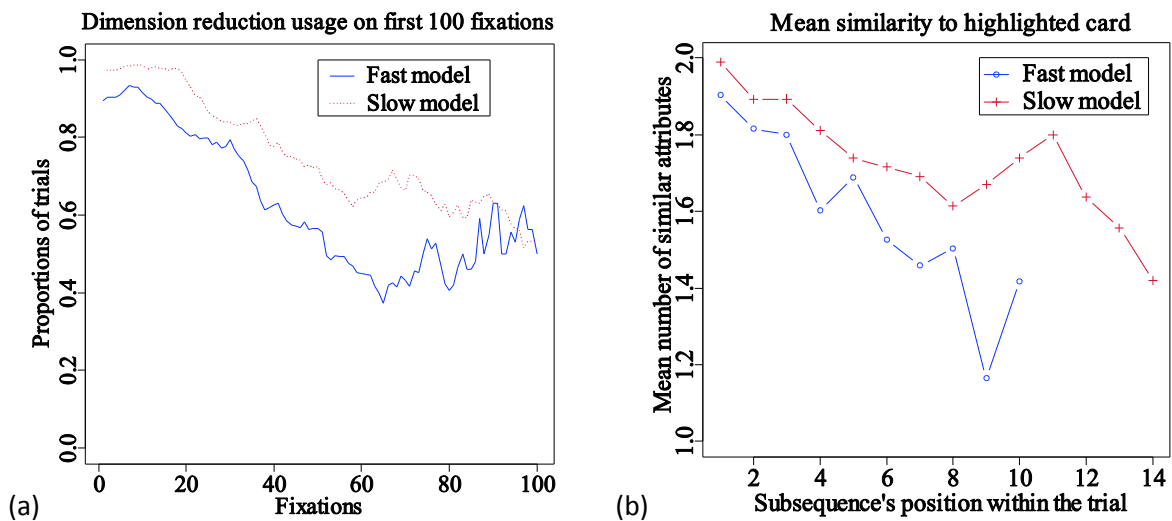


Figure 12. (a) Changing proportion of trials in which dimension reduction was used. The proportions are calculated as a function of fixation position in the sequence. Results are shown separately for slow and fast models. (b) A mean overall similarity of all cards in a subsequence to a highlighted card shown separately for slow and fast models.

Finally, as Figure 12a shows, the slow model is more likely to use dimension reduction in latter part of the trial than the fast model. In overall fast model is less biased toward dimension

reduction than the slow model showing effect similar to one produced by the fast human players (Figure 7a).

Dissimilarity-based search. Our experiment revealed that the subjects gradually switch from dimension reduction strategy to dissimilarity-based search (Figure 7b). To test whether the model exhibits the same pattern of behavior as the human players, the same type of analysis was done on fixation sequences produce by the model. The results can be observed in Figure 12b. There are gradual transitions from the similarity- to dissimilarity-based search for both slow and fast models. The difference between fast and slow models with respect to bias to the perceptual similarity is smaller than in human players, however it is present. It can be seen that graph for the fast model comes to an abrupt end at 10th subsequence. This is due to the fact that the fast model rarely required more than 10 subsequences to find the set.

General Discussion

Bottom-up and Top-down Processes

Improvement in playing SET can be explained by the interplay between the two types of processes. Slow players initially tend to rely on bottom-up processes, because their top-down strategies are too slow to keep up. Improvement in the game is characterized by an increase in efficiency and involvement of top-down processes.

A similar development was found in studies of other games such as Scrabble (Halpern & Wai, 2007). In that study, slow and fast players also differed in the interplay between top-down and bottom-up processes. Slow players prefer to rotate and rearrange the letters physically to check whether they form a word. It makes players very much dependent on bottom-up motor processes and perceptual stimuli representing the letters. On the other hand, fast players prefer to

rotate and rearrange the letters mentally. Hence, an fast player prefers to use top-down processes to manipulate the abstract representations of the letters.

Another example of a shift in balance between bottom-up and top-down processes is observed in Tetris. Initially it was believed that slow players prefer to rotate and translate tokens mentally to check whether that piece will fit at various parts of the screen while more experienced players prefer to rapidly manipulate the tokens physically (Kirsh & Maglio, 1994). However, a later study showed that players with extensive experience prefer to rotate and translate pieces mentally rather than physically (Destefano, Lindstedt & Gray, 2011). This means that they no longer require perceptual input to verify their solution. This is similar to learning in Set, where prediction processes make it unnecessary to “see” the third card in order to infer it is part of the set.

In light of these findings, we conclude that such shift in balance between top-down and bottom-up processes may be a very common learning process.

Implications of This Study

Threaded cognition for bottom-up and top-down processes. In earlier studies, fast players substitute bottom-up with top-down processes through substitution of physical with abstract, but otherwise identical, actions. However, our model showed that fast players can combine bottom-up and top-down processes beyond that of simple substitution. The fast model is able to perform actions, such as prediction, that are otherwise beyond capabilities of bottom-up processes. This capability requires viewing bottom-up and top-down processes as parallel and competing processes. Earlier, we referred to it as a competitive parallelism. This is in contrast to conventional sequential or hierarchical view, but in line with theory of threaded cognition (Salvucci & Taatgen, 2008, 2011). However, in addition to the separation of processes into

threads by tasks, we also have a division of processes into threads by their types within a single task. As such, this study can be viewed as a theoretical and practical example of threaded cognition and can contribute to general understanding of this theory.

Having two threads for the same task, competitive parallelism, has a direct implication in learning. Competitive parallelism ensures that when the same task can be accomplished by both bottom-up and top-down processes training will ensure the most suitable one will be chosen eventually. Competitive parallelism can be a cornerstone for problem-solving tasks. For example, it can explain how Tetris or Scrabble players minimize cost of mental operations while still doing the same task physically. Further study is needed to confirm those assumptions.

Implicit decision making. As we indicate there are several interesting findings in this research. SET players can apply more than one strategy during the game, similarity- and dissimilarity-based. Our model has shown that shifts between those strategies can occur as a result of evolving activations triggered by basic bottom-up elements such as inherent memory associations, inhibition and influence of perceptual stimuli. Such a choice of a strategy is not deliberate explicit decision, but rather implicit bottom-up decision. Perhaps, absence of explicit meta-cognitive control can explain why SET players are often unable to clearly describe their strategy. Furthermore, the similarity-based strategy is bottom-up and dissimilarity-based strategy is top-down. It suggests that there is not only an implicit shift in strategy, but also in type of processes. All together it suggests that bottom-up decision may have bigger role in cognitive processes and should be paid more attention in future studies.

As a possible line of future research in this direction, we would like to draw similarity between the way our models shifts between strategies and perceptual decision making models based on decision threshold. These mathematical models assume integration of sensory evidence

until decision variables reaches decision threshold after which categorical choice is made from alternatives (Smith & Ratcliff, 2004; Usher & McClelland, 2001). Similarly in our model, evolving activation in memory influenced by items in the visual field can be viewed as an accumulation of sensory evidence, and resulting probability of retrieval as a decision threshold. However, mathematical models provide no information about processes that govern the contextual regulation of the perceptual decision making (Domenech & Dreher, 2010). In contrast, our model provides a set of perceptual and cognitive processes backed by theory. As such, integration of mathematical and ACT-R based models may provide much more insight in domains of decision making and problem-solving in general.

Predictability and learning in problem solving task. The ability to predict is a useful, but understated, in our opinion, process of human cognition. There are limitations on the amount of visual information a human brain can process, and many consider selective attention shifts as mechanism to deal with the limitation. However, recent studies suggest that prediction also plays important role in mitigating processing limitations (Alink, Schwiedrzik, Kohler, Singer & Muckli, 2010; Soga, Akaishi & Sakai, 2009). Important parts of visual stimuli that are not processed are predicted based on previous experience. Furthermore, prediction is used to anticipate future stimuli. Recent study showed that predictability of the environment has significant influence on the decision making process (Domenech & Dreher, 2010).

In this paper, we showed how predictability of the environment in combination with a player's proficiency influences decision making. The fact that, in our model, difference in ability and accuracy of prediction was able to explain major difference between fast and slow players suggests that prediction possibly has important role not only in decision making, but also in the learning process.

Conclusion

It is our hope to contribute to the understanding of visual cognition where both internal conceptual knowledge and external perceptual stimuli converge in a goal-driven task. As one step toward this goal we have studied the importance of perceptual and cognitive processes in complex tasks requiring both internal planning and reaction to perceptual stimulus from the environment.

Firstly, there is an interaction between two types of process in accomplishing an immediate task. Such interaction involves both a sequential cooperation and a parallel competition with emphasis on the latter. Such competition gives a chance for top-down processes to gain edge over faster but limited bottom-up processes.

Next, both bottom-up and top-down processes are involved in decision making. On the one hand, bottom-up processes can influence top-down decision. On the other hand, bottom-up process, such as evolving memory activations, can result in decision without need of top-down control. This suggests that decision making may not be an explicit process only.

References

- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus Predictability Reduces Responses in Primary Visual Cortex. *The Journal of Neuroscience* , 30 (8), 2960–2966.
- Anderson, J. R. (2007). *How Can Human Mind Occur in the Physical Universe?* New York: Oxford University Press.
- Anderson, J., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological Science* , 2 (6), 396-408.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-Effects Modeling With Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* , 59, 390-412.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence* , 47, 139-159.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.
- Destefano, M., Lindstedt, J., & Gray, W. (in press). Use of Complementary Actions Decreases with Expertise. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Domenech, P., & Dreher, J.-C. (2010). Decision Threshold Modulation in the Human Brain. *The Journal of Neuroscience* , 30 (43), 14305-14317.
- Green, C. S., & Bavelier, D. (2004). The Cognitive Neuroscience of Video Games. In P. Messaris, & L. Humphreys (Eds.), *Digital Media: Transformations in Human Communication*. New York: Peter Lang.
- Halpern, D. F., & Wai, J. (2007). The World of Competitive Scrabble Novice and Expert Differences in Visuospatial and Verbal Abilities. *Journal of Experimental Psychology: Applied* , 13 (2), 79-94.
- Jacob, M., & Hochstein, S. (2008). Set Recognition as a Window to Perceptual and Cognitive Processes. *Perception & Psychophysics* , 70 (7), 1165-1184.
- Kieras, D. (2010). Modeling Visual Search of Displays of Many Objects: The Role of Differential Acuity and Fixation Memory. *Proceedings of the 10th International Conference on Cognitive Modeling*, (pp. 127-132).
- Kim, M., & Cave, K. R. (1995). Spatial Attention in Visual Search for Features and Feature Conjunctions. *Psychological Science* , 6 (6), 376-380.

- Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18, 513-549.
- Kong, X., Schunn, C. D., & Wallstrom, G. L. (2010). High Regularities in Eye-Movement Patterns Reveal the Dynamics of the Visual Working Memory Allocation Mechanism. *Cognitive Science*, 34 (2), 322-337.
- Lebiere, C., & Best, B. J. (2009). Balancing Long-Term Reinforcement and Short-Term Inhibition. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
- Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G., & Ritter, H. (2001). Comparative Visual Search: A Difference That Makes a Difference. *Cognitive Science*, 25 (1), 3-36.
- Rayner, K. (1995). Eye Movements and Cognitive Processes in Reading, Visual Search, and Scene Perception. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Eye Movement Research: Mechanisms, processes and applications* (pp. 3-22). Amsterdam: North Holland.
- Salvucci, D. D. (1999). *Mapping Eye Movements to Cognitive Processes*. Doctoral Dissertation, Carnegie Mellon University, Department of Computer Science.
- Salvucci, D. D., & Taatgen, N. A. (2011). *The Multitasking Mind*. New-York: Oxford University Press, Inc.
- Salvucci, D. D., & Taatgen, N. A. (2008). Threaded Cognition: An Integrated Theory of Concurrent Multitasking. *Psychological Review*, 115 (1), 101-130.
- Smith, P. L., & Ratcliff, R. (2004). Psychology and Neurobiology of Simple Decisions. *Trends In Neurosciences*, 27 (3), 161-168.
- Soga, R., Akaishi, R., & Sakai, K. (2009). Predictive and Postdictive Mechanisms Jointly Contribute to Visual Awareness. *Consciousness and Cognition*, 18, 578-592.
- Taatgen, N. A., Oploo, M., Braaksma, J., & Niemantsverdriet, J. (2003). How to Construct a Believable Opponent Using Cognitive Modeling in the Game of Set. *Proceedings of the Fifth International Conference on Cognitive Modeling*, (pp. 201-206).
- Usher, M., & McClelland, J. L. (2001). The Time Course of Perceptual Choice: the leaky, competing accumulator model. *Psychological Review*, 108 (3), 550-592.

Footnotes

¹ SET is a game by Set Enterprises (www.setgame.com)

² The sum of distribution numbers exceeds the total number of trials because the same trial can have two or more most abundant value groups of equal size but different attribute types.

³ To eliminate possible influence of the highlighted card, only trials without highlighted cards were considered in calculating the proportion.

⁴ ACT-R stands for *Adaptive Control of Thought - Rational*

Appendix

The probability of k subsequent fixations falling on cards that share at least one value in common if the fixations are assumed to be random is calculated with a following equation:

$$P = \sum_{i=1}^4 \sum_{j=1}^3 \frac{n_{ij}}{11} * \left(\frac{n_{ij} - 1}{11} \right)^{k-1}$$

k – the number of fixation in fixation subsequence

n_{ij} – a number of cards in array of 12 cards that have value j for an attribute i

Before further explanation, one should consider that this analysis is done on collapsed fixation sequence where consecutive fixations on the same card are considered as a single fixation therefore the next fixation always falls on another card.

Let's assume that there are five green cards among 12 cards on the desk. If we assume that subjects is always fixating on one of the cards before fixating on another card then the number of possible cards on which subject can fixate is 11. Probability of randomly fixating on one of those 11 cards is 1/11. Now if we assume that subject started looking at green cards then the probability of the first fixation on any green card is 5/11. However the probability of second consecutive fixation on another green card is 4/11, since subject is already fixating on one of the green cards. The probability of each of next consecutive fixations after the second fixation will be 4/11 as well. If subject did seven consecutive fixations on green cards then the probability of entire block of fixations will be $\frac{5}{11} * \left(\frac{4}{11} \right)^6$. If instead we want to calculate a probability of seven consecutive fixations on cards that share any attribute value (not just green color) then it will be the sum of probabilities for each individual attribute value.

If the calculated probability of the block of k fixations is below 0.05 then it is assumed to be not produced by chance. The blocks are calculated for an each attribute type. If two blocks of fixations from different attributes overlap then the block with the least chance probability is preferred. The other block is cut at the point of an overlap, and its probability is calculated again based on the block's new length. If the two blocks overlap and have an equal chance probability then the longest block is preferred. If the lengths are also equal then one of the blocks is randomly chosen and removed. Finally, Holm-Bonferroni correction was used on initial significance value of 0.05. The correction compensated for the inflation of the chance probability when multiple solid blocks are present in the same trial.

Cover Letter

March 7, 2011

Arthur B. Markman
The University of Texas-Austin
Department of Psychology
1 University Station A8000
Austin TX 78712-0187
Work: 512.232.4645
Fax: 512.471.6175
E-mail: markman@psy.utexas.edu

Dear Professor Markman:

I am enclosing a second revision for the earlier submission to the *Cognitive Science Journal* entitled, "Set as an Instance of a Real-world Visual-cognitive Task." The current version of the manuscript is 49 pages long and includes 3 tables and 12 figures.

We would like to thank editor and reviewer for providing a valuable feedback on this manuscript. According to these comments, the manuscript went through a noticeable revision. A point-by-point description of how we have addresses each comment can be found enclosed in this letter.

Preliminary results of this research were presented at the Annual Conference of the Cognitive Science Society (July 2011). This paper extends on preliminary results and provides more detailed in-depth view of the topic in concern. We would be happy to provide copy of the other manuscript if there should be any concern about duplicate or fragmented publication.

My coauthor and I do not have any interests that might be interpreted as influencing the research, and APA ethical standards were followed in the conduct of the study.

I will be serving as the corresponding author for this manuscript. Niels A. Taatgen listed in the byline has agreed to the byline order and to submission of the manuscript in this form. I have assumed responsibility for keeping my coauthor informed of our progress through the editorial review process, the content of the reviews, and any revisions made. I understand that, if accepted for publication, a certification of authorship form will be required that the coauthor will sign.

Sincerely,

Enkhbold Nyamsuren,
Department of Artificial Intelligence, University of Groningen,
Nijenborgh 9, 9747 AG Groningen, Netherlands

+31 50 363 7450 (voice)
e.nyamsuren@rug.nl

Note: Since figure and table enumerations in the paper have changed, we are using new enumerations in our answers to the comments.

Answers to the comments from editor.

Comment_E_1:

This reviewer's concerns derive from your definition of expertise, both in the human data (splitting continuous data into two groups), and the model (modeling performance using two discrete groups from continuous reaction-time performance). The reviewer sees this as especially problematic, and recommends rejection.

On my own reading, I do agree with this reviewer that this seems like a rather indirect definition of "expertise." As you note on p. 9, the evident differences in strategy suggest that this works as a criterion. Still, speed of responding is not, by itself, directly indicative of expertise; instead, does it reflect high vs. low performance, or perhaps skill? Part of this issue is a terminological one. Some clarification would help.

Another issue is how you analyze the human data. As Reviewer 1 notes, your subjects near the criterion boundary exhibit almost exactly the same reaction time. Perhaps you could compare these subjects and show that strategy composition at the boundary line is similar in predictable ways? This would lend credence to your grouping, and suggest that the deployment of bottom-up/top-down strategies is more different at the opposite ends of this distribution (from Fig. 3).

Answer_E_1:

We have removed the definition of expertise from this paper. Furthermore, we have removed any claim related to expertise or experience. The subsection "Expertise" on page 9 is renamed to "Performance". The reaction times are treated as an indicative of difference in deployment of similarity/dissimilarity based search. We also tried to make it clear that we are interested in subject's performance only in terms of difference in strategy.

We have changed the grouping of subjects to fast, medium and slow subjects (pages 12-13). The subjects that form a plateau (have more or less same reaction times) in the middle (Figure 3a, page 13) are included in the group of medium players. Accordingly, models are compared to corresponding groups of fast and slow players. We tried to emphasize the difference between slow and fast players by contrasting these two groups in Figure 7a and Figure 7b (pages 18-19). Furthermore, to add more credence to grouping based on RT, we have used subjects' mean reaction times (from Figure 3a, page 13) as independent variables (instead of dummy variable indicating the group) in linear mixed-effect regression analysis results of which are shown in Table 1 (pages 21-22). The results suggest strong correlation between reaction times and type of deployment of bottom-up/top-down strategies used by players. Finally, using the ACT-R model, we show that a difference in deployment of bottom-up/top-down strategies can produce similar gradience in reaction times observed with human subjects (Figure 9a, page 36).

Comment_E_2:

A third issue is how you are modeling it. It appears, from p. 35, that the subjects are captured with the model discretely: Two completely different classes of model. The human data suggest a gradience of performance. Is there some way you can address this in a revision?

Answer_E_2:

The novice and expert models are actually completely the same. They have exactly the same set of production rules and parameter settings. The models use exactly the same algorithm. Both models try to do both top-down prediction and bottom-up scanning in parallel. The only difference between two models is the number of rules available in the declarative memory when the models perform the task. The novice model has no rules available. However, it still tries to make predictions, although, all attempts will be unsuccessful. The expert model has all possible rules. Therefore its attempts to make prediction can be successful. Our previous Figure 8 probably gave the impression of two separate models, and we have therefore updated to clarify this point.

Furthermore, current model can show the gradience of performance (pages 35-36). Gradience from novice model to expert model can be achieved by gradually increasing number of rules available in a declarative memory. The

Figure 9a, page 36, shows the model's mean reaction times when it is run with a different number of rules. The number of rules is gradually increased from model 0 to model 7. Correspondingly model's RT decreases as the number of rules increases.

In the first version of the manuscript, we have briefly mentioned about model's ability to learn and transit from novice to expert by generating random rules each time model finds a set. However, parts related to learning were removed in the second version of this manuscript due to reviewers' comments stating that more study is needed with this respect.

In the revised version of the manuscript we have included (in the section *Model Results*) a short description of six other models. The corresponding reactions times are also shown in a Figure 9a. The overall results show how gradual increase number of available prediction rules results in gradual decrease in reaction times.

Comment_E_3:

As an additional (more minor) note from me, I also observe on p. 42 that you equate mental rotation in the Tetris task (for example) with the top-down processes explored in SET; yet these seem to be significantly different kinds of top-down processes (where mental rotation *is* simply rotation, just not in the form of epistemic actions, a-la Kirsh and Maglio). I wonder if that note on p. 42 may be rephrased or explained in slightly more detail.

Answer_E_3:

We have slightly expanded the explanation of the correspondence on page 41. Also, in Tetris players are not just doing mental rotation (also translation, which we added), because they also have to decide whether and where the rotated piece fits in the rest of the puzzle.

Comment_E_4:

Finally, I do notice there are several typos in this manuscript (e.g., stray apostrophe on p. 9, middle; missing determiner "the" in "expert model" on p. 36, etc.). I suggest proofreading it carefully to ensure that there are no errors or awkward phrasings.

Answer_E_4:

We have corrected the grammatical errors.

Answers to the comments from Reviewer #1

Comment_R1_1:

The distinction between experts and novices is solely based on speed. The slowest 7 subjects are novices and the 7 fastest subjects are experts. There are two problems with this. First, there are only 14 subjects in the experiment, and each subject performed only 60 trials so this is a very small amount of data (840 trials). Second, Figure 3a shows that reaction times for the fastest novice and slowest expert are virtually the same. There is no natural separation, and the distinction between novices and experts is arbitrary.

Answer_R1_1:

Please, refer to Answer_E_1.

Comment_R1_2:

On p. 9, the authors state: "[We define] expertise as the relative time it takes a player to find a set." Yet, expertise in ACT-R is modeled by inserting rules in the model. This constitutes explicit knowledge, and there is no measure of whether this explicit knowledge is present in human experts. In fact, the authors state on pp. 43-44: "Perhaps, absence of explicit meta-cognitive control can explain why SET players are often unable to clearly describe their strategy." This is a contradiction between what is assumed of the subjects, and how they are modeled (In Anderson et al., 2004, knowledge structures in ACT-R are explicit).

Answer_R1_2:

There is no clear mapping between the notions of implicit and explicit memory onto knowledge structures in ACT-R. Procedural knowledge typically cannot be considered as explicit knowledge, because the cognitive system has no direct access to them. Declarative knowledge is used to model aspects of both implicit and explicit memory and learning: a particular fact can be considered as explicit knowledge, but a collection of experiences onto which a skills can draw typically has the characteristics of implicit knowledge. In the Set model, an aspect of expertise consists of declarative rules that help in the prediction of the third card. The knowledge itself is explicit in the sense that the model knows that green is the third color if the other two are red and blue. But that does not mean that the model's problem solving process in which this knowledge is used is fully open to introspection and subsequent reporting.

Anderson often draws parallels between ACT-R's vision module and declarative module (In Anderson, 2004, How can the human mind occur in the physical universe: Human associative memory). Chunk in a visual-location buffer is not explicit knowledge until it is encoded into visual buffer. Similarly, chunk in declarative memory is not explicit until it is retrieved.

Comment_R1_2:

Finally, although this paper focuses on the distinction between novices and experts, most analyses pool expert and novice data together. The only exceptions are reaction x set level (figure 3b) and strategy switch (Fig. 8b). If experts and novices are truly different, differences should emerge in other measures, and it would be wrong to perform pooled analyses.

Answer_R1_2:

The distinction between expert and novice players is not the main focus of this paper. However, we do agree that fast and slow players differ in measures other than mentioned in this paper. For example, mean fixation duration for faster players is shorter than fixation duration for slower players. Similarly, slower players need more fixations to verify if three cards make a set. However, there are two reasons why we did pooled analysis in certain parts of this paper and did not mention differences in other measurements. Firstly, it is due to space limit for the manuscript, and, secondly, ACT-R as a cognitive architecture does not provide means to model to level of details necessary for those measurements (fixations durations for example). Nevertheless, in the current version of the manuscript we did separate analysis of fast and slow players for three main effects on which major content of the paper is based: reaction times with respect to set level (Figure 3a, page 13), usage of dimension reduction (Figure 7a, page 19) and transition to a dissimilarity-based search (Figure 7b, page 19).