From Vision to Reasoning

What it takes to model a computer game opponent

Enkhbold Nyamsuren

Printed by Gildeprint Drukkerijen,
Enschede, the Netherlands.

ISBN printed version: 978-90-367-6879-5
ISBN digital version: 978-90-367-6880-1

university of groningen

From Vision to Reasoning

What it takes to model a computer game opponent

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus, Prof. E. Sterken
and in accordance with the decision by the College of Deans.

This thesis will be defended in public on

Friday 28 March 2014 at 14:30 hours

by

Enkhbold Nyamsuren

born on 1 September 1985
in Ulaanbaatar, Mongolia

Supervisor
Prof. dr. N.A. Taatgen

Assessment committee
Prof. dr. W.D. Gray
Prof. dr. P. Roelfsema
Prof. dr. J.B.T.M. Roerdink

Contents

CHAPTER 1

Introduction

First studies, first questions.

The original goal of the research presented in this dissertation was to study techniques and methods for creating believable artificial opponents for computer games. Computer games have a significant entertainment factor. It is hard to find anyone who did not occasionally play them in some form or other whether it is on a desktop computer, tablet or smart phone. Games are one of the widest spread recreational activities most people are familiar with (Entertainment Software Association, 2013). Using games as experimental tasks provides a more familiar environment to participants and allows them to be more engaged in a task.

More importantly, computer games are great representatives of real-life tasks. They require from a player a wide range of cognitive skills used in real world tasks on a daily basis. Many studies show that skills acquired during gameplay are transferred to real-world problems. Playing video games can help to improve various aspects of cognition such as attention, vision, memory or even executive control (Boot, Kramer, Simons, Fabiani, & Gratton, 2008; Glass, Maddox, & Love, 2013). Furthermore, positive effects of video games are present in children, adults and elderly people alike (Anguera, Boccanfuso, Rintoul, Al-Hashimi, Faraji, Janowich, Kong, Larraburo, Rolle, Johnston, & Gazzaley, 2013; Basak, Boot, Voss, & Kramer, 2008; Mackey, Hill, Stone, & Bunge, 2011).

It is relatively easy to adapt games to controlled laboratory environment without overly simplifying to a degree that it is no longer a true representative of real-world tasks. As a result, games allow studying roles of individual cognitive processes as well as how those processes work together to solve complex problems.

In this thesis, the emphasis is put on the latter issue: how cognitive processes collaborate. There are plenty of existing studies of individual cognitive processes, but there is a lack of understanding how those processes fit together. For example, it is relatively well understood how feature-based visual search works (Treisman, & Gelade, 1980; Wolfe, 2007; Wolfe, J., & Horowitz, 2004). Holistic perception of human faces is also a well-researched field (Maurer, Grand, & Mondloch, 2002; Wenger & Ingvalson, 2003). However, studies of these two types of cognitive processes are done separately and in isolation from one another. Neither of them can provide a feasible explanation of how we find and recognize a familiar face in a crowd of people, something we do quite often in a real world. Furthermore, connecting two theories of visual search and holistic perception is not a trivial task. At least, it requires an understanding of how a top-down goal of finding a particular person's face is combined with bottom-up processes of visual search such processing and recognition of visual features either individually or as whole. A similar argument can be invoked with respect to cognitive processes other than holistic processing and feature-based search.

Chapter 2 describes a study of players' behaviors in the card game of SET. The game has a highly perceptual component that requires a skill to identify quickly combinations of cards defined by the game's rules. As such, top-down strategy in the game is heavily influenced by bottom-up visual processes. Chapter 2 tackles the issues of how we combine bottom-up visual processes with top-down meta-control, and the roles of two types of processes in defining participants' performance.

Chapter 5 builds on Chapter 2 and further extends study about SET. Any complex task can be defined in terms of its structural and presentation components. Chapter 5 explores how the overall SET strategy revealed in Chapter 2 is defined and influenced by these two components. In this follow-up study, the visual presentation style of the original game of SET was manipulated without any change in the structure of the game. This paradigm allows studying how overall strategy and individual steps in the strategy change due to different presentation style. It is shown how a change in a presentation style without any change in the isomorphic structure of a game can significantly affect subjects' performance.

Models or modules?

Although we were successful in modeling player behavior in SET using ACT-R cognitive architecture (Anderson, 2007), the study revealed a major obstacle in achieving our initial goal of developing a human-like AI using a cognitive architecture. Existing architectures are simply not sophisticated enough to tackle real-world problems. On the one hand, cognitive architectures such as ACT-R, EPIC (Kieras & Meyer, 1997; Meyer & Kieras, 1997a, 1997b) or SOAR (Laird, Rosenbloom, & Newell, 1987; Newell, 1990) do not have automated reasoning mechanisms. On the other hand, computational reasoning systems based on architectures such as the Procedural Reasoning System (Ingrand, Georgeff, & Rao, 1992) are not proper representatives of human cognition. Significant amount of overhead coding and major simplifying assumptions are required to create a plausible model of any complex problem-solving task. The obstacle reflects one of the major problems cognitive modeling community faces today.

The main objective of cognitive modeling is not to create isolated models of separate tasks, but to build a computational architecture of human embodied mind (Anderson, 2007; Newell, 1973, 1990) that explains its internal workings not only (1) from the perspective of constituent cognitive processes and theories, (2) but also as a single coherent system. So far, the cognitive modeling community excelled in achieving the first objective. However, the second objective more often than not is ignored. As a result, development of ACT-R as a cognitive architecture has seen very little significant advances in recent years. The conclusion is simple: ACT-R in its current implementation is simply not ready for modeling complex real-world problems.

Current and past studies in cognitive modeling resulted in a proliferation of isolated models of human cognition. 80% of all articles published in major theoretical journals of Cognitive Science involve cognitive modeling (Busemeyer & Diederich, 2010; Farkaš, 2012). These are the models of human behavior in very specific tasks with little regard of how it fits within theory of human mind as whole. This isolation gives a rise to a very challenging and daunting problem of how these models can be reused and combined (Newell, 1973, 1990). Some propose establishing a common repository of cognitive models (Addyman & French, 2012; Myung & Pitt, 2010). Common repository makes it easier to find existing models and reuse them. However, it provides little solution to problems of compatibility among models (even if they use the same cognitive architecture). As a result, there remains a major challenge of combining those models to simulate human cognition in complex tasks such as SET.

Others adopt an approach of supermodels (Salvucci, 2010). A supermodel is a model based on a fixed cognitive architecture that already has a necessary knowledge that can be used across range of tasks. The major problem with supermodels is that it does not make such distinctions as task-specific versus task-general knowledge and innate versus acquired knowledge (Carey & Spelke, 1994; Kirkham, Slemmer, & Johnson, 2002; Markram & Perin, 2011; Perin, Berger, & Markram, 2011). A person born in Asia might have a very different acquired knowledge than the person born in Europe. However, it is quite likely that they have the same innate knowledge. Similarly, our capability of solving a problem by using an analogy is task-general (Gust, Krumnack, Kühnberger, & Schwering, 2008). However, individual steps within analogy-based strategy might be dictated by task-specific knowledge.

I have adopted an approach that is different from the two previously mentioned. In any study of human cognition it is important to differentiate between task-specific and task-general knowledge and processes. Ideally, any model of a specific task should contain knowledge and simulate cognitive processes specific to that task. Task-general knowledge and processes should be a part of cognitive architecture. Therefore, any cognitive modeling effort should be separated into two mandatory steps: (1) making any necessary additions and changes into a cognitive architecture to reflect task-independent components of experimental findings and (2) developing an architecture-based model to reflect task-specific components of experimental findings. Current practices of cognitive modeling do not make, at least, any explicit differences between the two steps. This is a major cause why we have largely incompatible numerous cognitive models that contribute little to overall understanding of human cognitive system.

ACT-R consists of a collection of modules each representing a particular type of cognitive resources that are independent of any task (Anderson, 2007; Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). New modules can be easily added into the architecture for the purpose of either adding a new type of cognitive resources or extending the functionality of the existing ones. Starting from Chapter 3, all modeling works in this dissertation adopt a strategy of developing both models and modules. When necessary, a task-general module for ACT-R architecture was developed or changes to existing ones were made alongside the development of a task-specific model.

Because of being part of the architecture itself, modules are reusable across different models directly addressing the problem of reusability mentioned earlier. Furthermore, any changes in the architecture were extensively tested for validity via fitting models across range of behavioral tasks. Such testing ensures that new module or module changes are compatible with existing theories of human cognition. This, at least, partially addresses the problem of compatibility among separate cognitive modeling studies. Finally, development of modules directly contributes to the advancement of the ACT-R cognitive architecture. This is an approach different to traditional modeling. It puts emphasis on the architecture rather than the model. Consecutively, it also promotes understanding of human mind as a single coherent system without sacrificing the granular view of individual cognitive processes within distinct models.

In ACT-R community, there are isolated instances where modelers put emphasis on development of modules rather than models. Those include modules of integrated threaded cognition theory (Salvucci & Taatgen, 2008), short-term declarative inhibition (Lebiere & Best, 2009), integrated theory of eye movements and encoding (Salvucci, 1999), time duration estimation (Taatgen, Van Rijn & Anderson, 2007), etc. These and other studies shows feasibility of module-centric modeling as way of explaining cognitive and psychological theories even in isolated experimental tasks.

What modules do we need to play a game?

As was previously mentioned, the original goal of this research was to study techniques and methods for creating believable artificial opponents for computer games. However, the study described in Chapter 1 revealed that ACT-R is not sophisticated enough to model complex tasks such as SET. At the same time, the study clearly indicated that two major components were lacking in ACT-R. Following these results, the original goal was redefined as research into minimum components of a cognitive architecture necessary to develop plausible cognitive models for complex tasks such as computer games.

Human vision

Firstly, any complex task that requires making a decision based on the real-time knowledge of the world needs a reliable means of gathering information. Among five senses, the visual system is arguably the most important medium of gathering such information. Most of modern technology is designed around visual input as the primary source of communicating information to the user. Studies of efficient information communication mainly revolve around information visualization. ACT-R has a vision module as one of the core modules (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Byrne & Anderson, 2001; Salvucci, 2001). Unfortunately, the module provides only a bare bone implementation of the human visual system. Guidance of visual attention is a very complex process that has both top-down and bottom-up components (Orban, Fiser, Aslin, & Lengyel, 2008). On the one hand, visual attention is guided by bottom-up inherent properties of the visual scene such as contrast-based saliency of its constituent parts. On the other hand, visual attention is also guided by top-down components such as immediate goal and a context defined by previous experience. The default vision module provides no support to bottom-up attentional guidance. It also lacks several other fundamental functionalities such as long- and short-term visual memories, definition of visual objects along several feature dimensions and imagery capability.

Chapter 3 introduces the **P**re-**A**ttentive and **A**ttentive **V**ision (PAAV) module that is an extension to ACT-R's default vision module. As the name suggests, PAAV introduces a significant amount of pre-attentive functionality to ACT-R's visual system. PAAV is essentially an implementation of a collection of well-studied theories of human vision ranging from visual memory to contrast-based saliency maps for guiding visual attention (Itti, Koch, & Niebur, 1998).

As a module, PAAV is completely task-independent. It was designed to be able to handle not only tasks commonly used to test computational models of human vision, but also more general and complex tasks such as SET. Several task-general elements that were part of the SET model described in Chapter 2 are incorporated into PAAV. A new refined model of the game of SET that uses PAAV module is described in Chapter 3. The new SET model inherited only task-specific components (game rules and strategies) from the old SET model. PAAV was also used in all cognitive models described in chapters following Chapter 3.

Chapter 4 contains an interesting case study based on a task of identifying a **M**ost **A**bundant **V**alue (MAV). Experiments and models based on the task show how top-down control can directly override bottom-up saliency-based attentional guidance. Creating a plausible model for the task would have been impossible without the PAAV module.

Human reasoning

Any problem-solving task requires some degree of reasoning. It can be any form of reasoning: reasoning by analogy, reasoning based on rules or simply based on associations. Individual steps in the process of reasoning can be tied to specific task context, but our general ability to reason is a fundamental process independent of any specific task

(Gigerenzer, Todd, & the ABC Research Group, 1999; Johnson-Laird, 1983). For example, reasoning based on analogy is perceived to be fundamental to human cognition (Gust, Krumnack, Kühnberger, & Schwering, 2008). Cummins (1996a, 1996b) argued for innateness of deontic reasoning. On the other hand, there is evidence that infants' knowledge acquisition is largely dependent on innate concepts and principles (Baillargeon, 2008). It should be clarified that the general reasoning ability is not necessarily innate. Innate knowledge is not synonymous with task-general knowledge. We have an extremely specialized ability to recognize and process human faces holistically (Richler, Gauthier, Wenger, & Palmeri, 2008). Similarly, acquired knowledge is not necessarily task-specific knowledge. For example, in his paradigm of skill transfer, Taatgen (2013) argues for a learning mechanism that produces partial task-general production rules from a specific context. It is very much possible that we acquire general reasoning skills in early stages of our development.

In terms of ACT-R, there should be a task-independent general set of production rules that provide schematic rules for reasoning based on a given context. For example, rules in SET dictate that if two candidate cards are green and blue then the third card should be red in order to form a valid set. This rule can be written as *(Blue, Green) => (Red)*. This is a task-specific rule. However, the ability to reason based on this rule should be task general. ACT-R should have a task-general knowledge of conjunction of concepts. It also should have knowledge that conjunction of certain concepts can imply another concept.

Chapter 6 describes the **H**uman **R**easoning **M**odule (HRM) developed by me for ACT-R. The HRM introduces a basic set of declarative and procedural knowledge to ACT-R that allows it to reason based on task-specific instructions. The HRM introduces an explicit notion of a *concept*. It also knows that individual concepts can be combined to form more complex *statements*. In turn, those statements can be combined into declarative rules that can be used to encode task-specific instructions. Concepts, statements and rules based on statements form the declarative part of the HRM knowledge. The HRM's procedural knowledge is represented by a set of task-general production rules that describe how declarative and other forms of knowledge can be used for reasoning. Ideally, if the HRM is used and a proper set of instructions about the task is given in the declarative memory, the modeler will have to write only few task-specific production rules responsible mostly for meta-control.

One of the features that set apart HRM from the traditional view of human mental logic (Rips, 1983) is that reasoning in the HRM is not purely top-down. Facts and evidence necessary for reasoning can be extracted on the fly from information sources other than declarative memory. An example of such alternative source is the visual memory in PAAV. If I am asked about a position of a fork relative to the plate, and I can see both of them in front of me, then I do not need to recall a propositional statement from my declarative memory. I can just directly and quickly extract the fork's location cue from my spatial memory. Similarly, PAAV can take advantage of bottom-up information in visual memory during the reasoning. For example, it can extract raw spatial information from visual memory and directly translate it into a declarative statement.

The PAAV and the HRM combo

The combination of PAAV and the HRM within the ACT-R architecture provides a powerful toolset that allows modeling human behavior from the simplest of tasks to complex problem solving. Chapter 6 describes a model of a spatial reasoning task that is highly dependent on both the HRM and PAAV. With the help of this model, we try to explain the connections between the two competing theories of deductive reasoning: mental models (Johnson-Laird, 1983) and mental logic (Rips, 1983).

Chapter 7 provides a description of incomplete but very promising and interesting study of players' strategy in the Qwirkle game. The game requires a player to make explicit comparisons and weigh alternative moves to maximize the score. At the same time, the game retains a significant perceptual component since a significant amount of reasoning is based on matching or mismatching colors and shapes. Although we do not have a cognitive model yet, the analysis of human data already indicates performance is highly dependent on both the visual system and reasoning capability to make an optimal decision. As such a plausible model of a Qwirkle player will require both the PAAV and the HRM modules.

CHAPTER 2

Set as an Instance of a Real-World Visual-Cognitive Task

This chapter has been published as:

Nyamsuren, E. & Taatgen, N.A. (2013).
Set as instance of a real-world visual-cognitive task.
Cognitive Science, 37 (1), 146-175.

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. This article 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.

Introduction

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 it 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 article 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 [DM], 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 SET. Please, refer to Appendix A for a detailed description of SET.

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.

There are several advantages of choosing SET as a target game of study. First, 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, & 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 $\binom{81}{12}$ (approximately $7*10^{13}$) possible combinations of 12 cards, which makes it highly unlikely that players will play through the same 12 cards again. There are also 1,080 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, a player may also consider combinations of two cards as 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 $\binom{12}{2} = 66$ 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, a player still has to consider a maximum of 54 pairs before finding a set. This is a significant decrease in complexity compared with a strategy where a 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 us assume that a player is using similarity in color to find a set. Blue and green cards cannot have a set as there are only two cards in each group. There are $\binom{8}{3} = 56$ combinations of three cards among red cards and 32 combinations of three cards with different colors. It is already a significant decrease in complexity from 220 to 88 possible combinations and a 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 article. 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 experiments aimed at 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 10 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 Adaptive Control of Thought–Rational (ACT-R) model. However, the data they collected did not have enough detail to determine whether subjects used such a 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 to 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. As 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 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 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 3×4 array similar to one shown in Figure A.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 the 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 seven or eight 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 220 to 55, and the number of possible pairs from 66 to 11. As there are two pairs that lead to a set, in the worst case, a 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 500 Hz and spatial resolution of <0.01° RMS. The card images were shown on a 20-inch LCD monitor with screen size of 1,024 × 768 pixels and screen resolution of 64 pixels/inch. The card images had a size of 124 × 184 pixels, or 4.02° × 5.95°. 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 cm as subjects were given a certain freedom for head movement. The gaze position was calculated using the eye's corneal reflection captured using 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 2.1 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 (Jacob & Hochstein, 2008; Taatgen et al., 2003).

As it is shown in Figure 2.2a, 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 2.2b indicates that there is only a small difference in speed among three groups when it comes to finding a level 1 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.

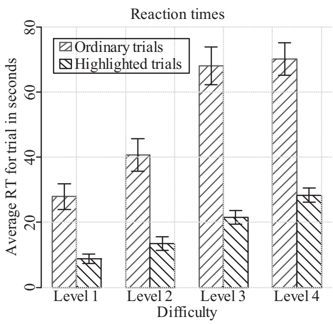


Figure 2.1: The mean reaction times with standard errors in ordinary and highlighted trials clustered by the levels and averaged over all subjects.

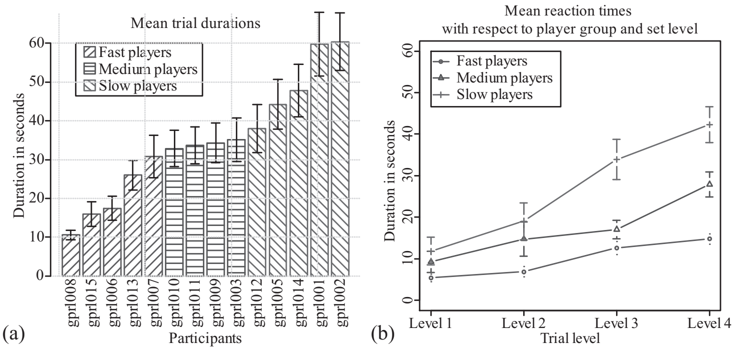


Figure 2.2: (a) Mean reaction times averaged over all trials for each subject. Subjects are divided into three groups: fast, medium, and slow players. (b) Mean reaction times averaged over trials of the same level and player group.

Fixations

As it is quite obvious from RT plots shown above, trials vary by their lengths of fixation sequences. Figure 2.3 shows proportions of trials by lengths of their collapsed fixation sequences. A collapsed fixation sequence is a sequence where consecutive fixations on the same card are collapsed into one fixation. Figure 2.3a shows that in the ordinary condition around 49% of all trial has 100 or less collapsed fixations. According to Figure 2.3b, around 85% of all trials have less than 100 collapsed fixations in the highlighted condition.

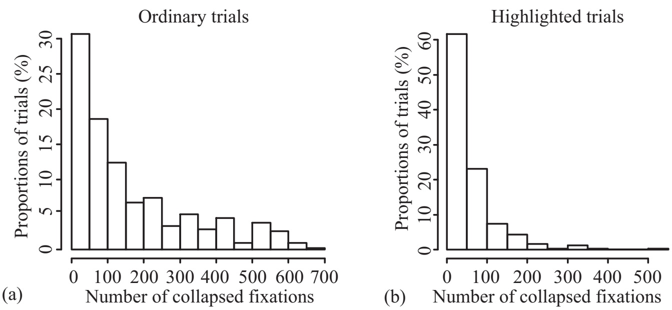


Figure 2.3: Proportions of trials by lengths of collapsed fixations sequences in (a) the ordinary and (b) the highlighted conditions.

There is also a variation in the number of collapsed fixations depending on the level condition. In the ordinary condition, average lengths of collapsed fixation sequences in levels 1 to 4 are 93 (SE = 15), 132 (SE=15), 218 (SE=26) and 224 (SE=26) fixations respectively. In the highlighted condition, those numbers are 28 (SE=4), 43 (SE=6), 67 (SE=7) and 91 (SE=10) fixations.

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 a 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.

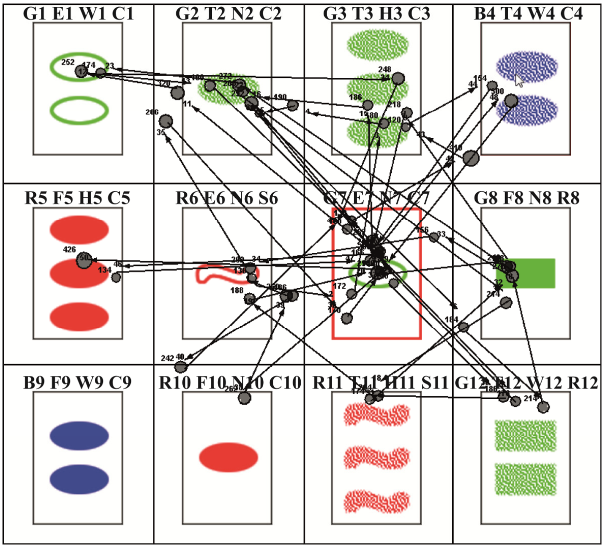


Figure 2.4: One of the problems shown to a subject. Card 7 is the highlighted card. Also shown are the collapsed fixations (circles) and saccades (arrows) produced by the subject. The outer thin, black borders indicate 12 areas of interest. The four combinations of letters and numbers on top of each card represent four labels for each area of interest. A set is formed by the fourth, fifth, and seventh cards.

To explore the existence of such a 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 2.4).

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. We further refer to such fixations and fixation sequences as collapsed fixations and collapsed fixation sequences. Combining all labeled collapsed fixations of a common attribute type into collapsed fixation sequences produces four distinct sequences for each trial.

An analysis of the collapsed fixation sequences revealed the existence of a pattern of collapsed fixations related to the usage of dimension reduction. We will demonstrate this using the example problem from Figure 2.4.

Figure 2.5a shows a collapsed fixation sequence diagram produced from the scanpath shown in Figure 2.4. In this example, the subject needed only 48 collapsed fixations to find a set. Each horizontal lane in the diagram shows a subject's collapsed fixation sequence with respect to the particular attribute type. One unit on the *x*-axis represents a collapsed 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 collapsed fixations on the cards with the same attribute value are shaded with a solid color if the probability of such a collapsed fixation subsequence occurring by chance is equal to or below 0.05 (refer to the Appendix for details of calculating the probability).

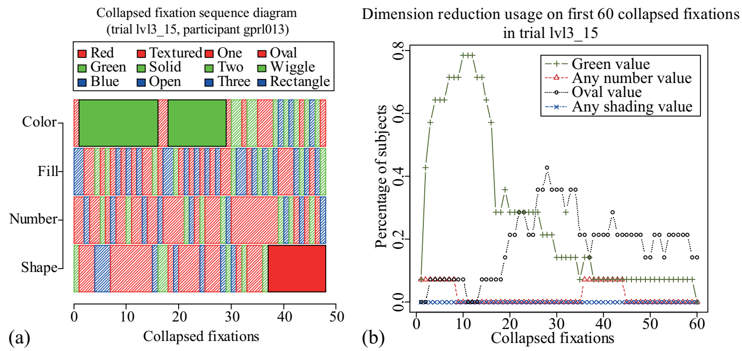


Figure 2.5: (a) Single subject's collapsed fixation sequence diagram for trial "lvl3_15." (b) Changing proportion of subjects who used dimension reduction in trial "lvl3_15" as a function of fixation position in the collapsed sequence and attribute value.

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 2.5b 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 2.5b that a 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 collapsed fixation sequences where all subjects used dimension reduction for all problems.

The result indicates that blocks of collapsed fixations with the same attribute value occupy on average 46% and 35% of an overall collapsed 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.

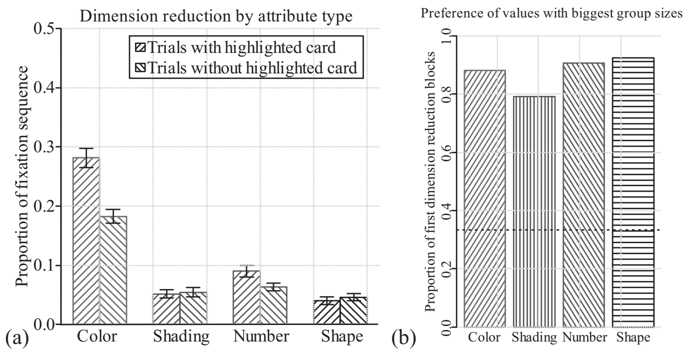


Figure 2.6: (a) Mean proportions of attribute types used in similarity-based scanning. Proportions are shown separately for trials with and without highlighted card. (b) Proportion of trials where subjects preferred to use a value for dimension reduction with the biggest group size among other values of the same attribute type. The horizontal, dashed black line indicates the expected proportion if the choice was made randomly.

Figure 2.6a 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, respectively[1]. According to such, the corresponding bars on Figure 2.6a 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, and 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 value is preferred for dimension reduction. As Figure 2.6b indicates, in 85% of all trials, subjects prefer the most abundant value over the other two values of the same attribute type[2]. The trend is consistent among all four attributes.

[1] 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.
[2] To eliminate possible influence of the highlighted card, only trials without highlighted cards were considered in calculating the proportion.

Effect of dimension reduction on performance

There is also a difference between fast and slow players in how they use dimension-reduction strategy. Figure 2.7a 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.

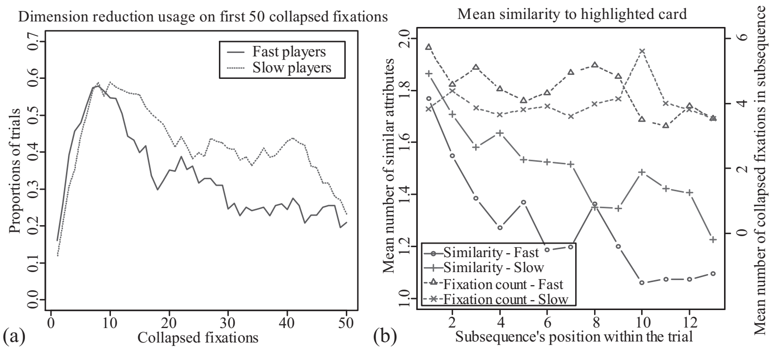


Figure 2.7: (a) Changing proportion of trials in which dimension reduction was used. The proportions are calculated as a function of the collapsed fixation position within a trial. The proportion on collapsed fixation *x* is calculated by counting the trials that have a dimension-reduction block that includes *x*. (b) The x-axis shows subsequences' positions within trial's overall collapsed fixation sequence. The y-axis on the left measures overall similarity of a subsequence to the highlighted card. The decreasing trend indicates that, with each new subsequence, subjects looked at cards less similar to the highlighted card. The y-axis on the right measures the mean (and standard error) number of collapsed fixations in each subsequence calculated separately for fast and slow players.

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, we have discussed earlier that subjects used dimension-reduction strategy only in 46% and 36% of the time in trials with and without highlighted card. Figure 2.7a 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 collapsed 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 collapsed fixations, presumably to refresh their memory and to restart a new search subsequence. The following labeled collapsed fixation sequence (provided for a purpose of example only) "**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. Fixations in each subsequence can be labeled with respect to four attribute dimensions in the same way described in Dimension Reduction section. Given the labeled subsequence, we estimated similarity to the highlighted card for each fixated card within the subsequence. Similarity is the number of attributes that share the same value between two cards. Finally, by taking a mean of all similarity values, we estimated subsequence's overall similarity to the highlighted card. 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 2.7b). 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 2.7b 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 2.7b indicates. Fast players appear to abandon similarity-based search earlier than slow players. On average, fast and slow players made 4.8 (SE = 0.47) and 4.0 (SE = 0.52) collapsed fixations per subsequence. According to Figure 2.7b, fast player may made fewer fixations with consecutive subsequences. However, the number of fixations per subsequence remains stable for slow players suggesting that the decrease in the mean similarity is not caused by the variability in the number of collapsed fixations over consecutive subsequences.

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 2.7b). 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 collapsed fixation within a subsequence. Variable *RT* is subject's mean reaction time in seconds shown in Figure 2.2a. In addition, two random effects on an intercept, *Subject* and *Trial*, were added, each representing subjects and trials, respectively.

Resulting coefficients for fixed main and interaction effects are shown in Table 2.1. 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 2.2.

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 collapsed fixation sequence as a 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 as 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 the 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.

Table 2.1: 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 2.2: 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

Experiment discussion

Experiment results' summary

The main effects of Fixation and Subsequence in the mixed-effect regression analysis indicates that the subjects' starting strategy of playing SET is similarity based. Subjects prefer to look for a set among the cards that are similar to each other and then gradually switch to groups of more dissimilar cards.

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 (Figures 2.5 and 2.6) 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: (a) the size of the group of cards that share the value and (b) its attribute type. The importance of group size (Figure 2.6b) 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 2.6a) 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 & 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 2.7a) nicely coincides with gradual reduction in reliance on similarity (Figure 2.7b). 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 2.7b and Table 2.1). 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 an 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 an 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 & 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, at the beginning of the game, most players tend to focus on the group of cards that share particular color or number values, as 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 suggest 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 may 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, not only collaborating but also competing to achieve the same goal.

Competitive parallelism assumes that all players have two parallel processes independent 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 sets and differ significantly in finding higher level sets.

An ACT-R model of an SET player

We have implemented the model using the ACT-R cognitive architecture (Anderson, 2007). Please, refer to Appendix B for a detailed description of ACT-R's internal workings.

Model design decisions

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. V_{DM} does not have to be part of HC.
   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 2.8).
7. If a set is not found then go back to step 1.

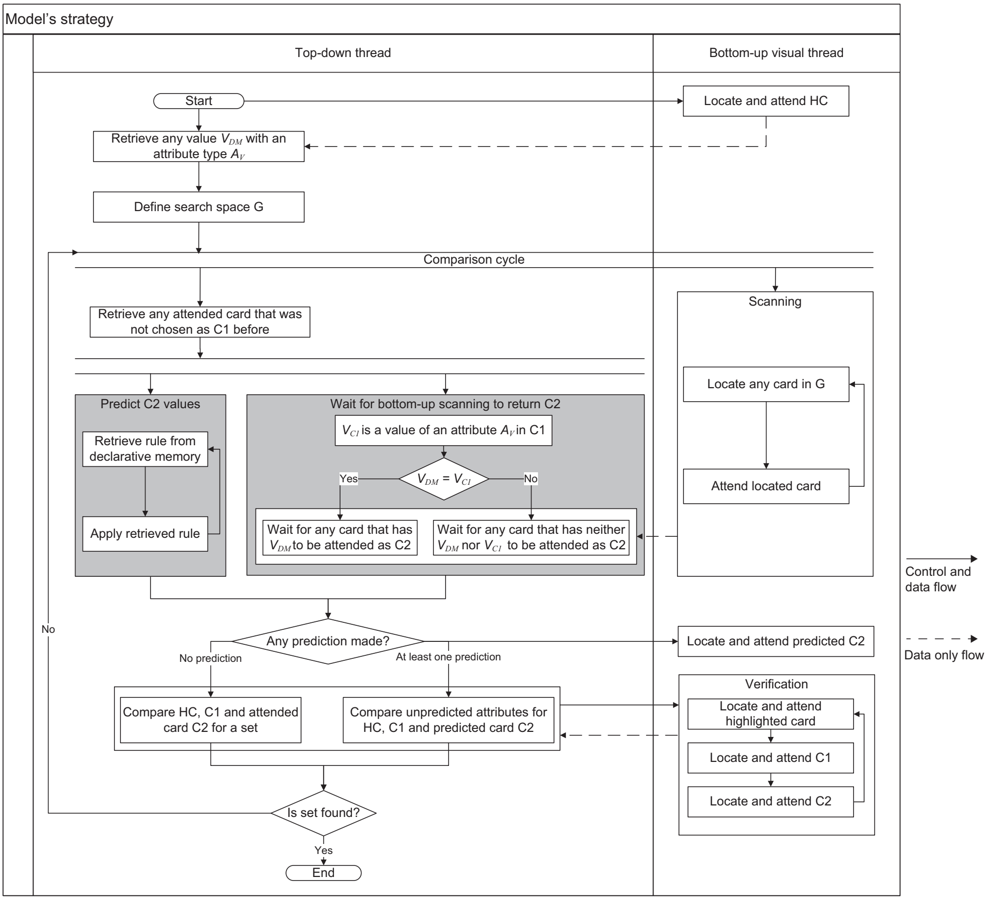


Figure 2.8: An algorithm for searching for a set, given a specified group of cards G. Two shaded boxes represent two approaches that model uses in parallel to find a set. The shaded box on the right shows the bottom-up approach to find a set, and the shaded box on the left shows the top-down prediction approach.

Implementation of both strategies in the model is emergent 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 2.3).

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 declarative memory (DM) in the same manner as the chunks in buffers do. This feature was added to model the effect of the group size. We used a logarithmic function (see Table 2.3) 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 2.3.

Values for most of the constants mentioned in Table 2.3 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 2.3 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.

Table 2.3: Parameters for calculating activation for an attribute value *i*.

Parameter	Influence On Activation	Implementation method
Attribute type	positive	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): • Color chunks: 40 • Number chunks: 36 • Shape chunks: 32 • Shading chunks: 28 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.
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 the value in the 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 the variance σ^2. σ^2 is calculated as $\sigma^2 = \frac{\pi^2}{3} s^2$, where s is an instantaneous noise 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.

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 2.8. 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 a higher chance of being chosen as C1. The model is free to choose its own scanpath 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 2.8 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 2.8 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 depend 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 favored.

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 the 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 the model can fail to find a set even if search space contains it.

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 *Given(Textured, Solid) ⇒ Expected(Open)*. It should be noted that *Given(Solid, Textured) ⇒ Expected(Open)* and the previous rule are treated as different ones. There are also rules for similarity such as *Given(Red, Red) ⇒ Expected(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 DM. The second model had 12 prediction rules for predicting similarity of the corresponding 12 attribute values (e.g., *Given(Red, Red) ⇒ Expected(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 2.9a shows reaction times for all eight models averaged over four difficulty levels. As was hypothesized previously, the model's reaction time gradually decreases as the 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 2.9a suggests that the main boost in performance through predictions is produced by trials with high-level sets.

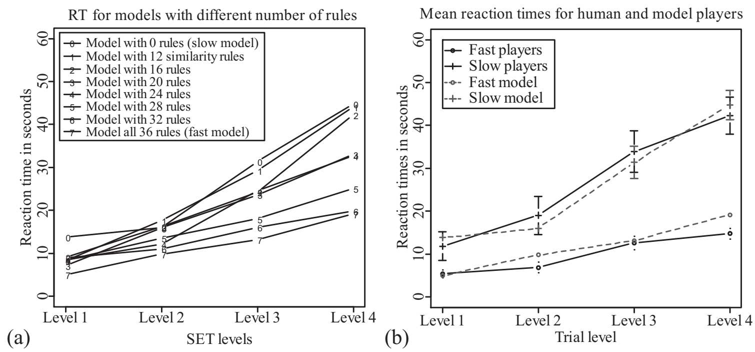


Figure 2.9: (a) Reaction times of eight models averaged over four difficulty levels. (b) Reaction times of the slowest and fastest models compared to the reaction times of the human players in trials with highlighted cards.

In Figure 2.9b, 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 collapsed fixation sequences produced by these two models were further compared to human data from fast and slow groups.

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 2.10a) and the subject's (Figure 2.5a) collapsed fixation sequences from the same trial are compared. The model's collapsed fixation sequence in Figure 2.10a closely resembles the sequence in Figure 2.5a 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 the same manner as human subjects did. This decrease is consistent with behavior of the human subjects.

It is obvious from multiple model runs that half of the times, the model prefers to look at the green cards at the beginning of the trial, although they form the second-largest group after cards with an oval shape (Figure 2.10b). 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 type.

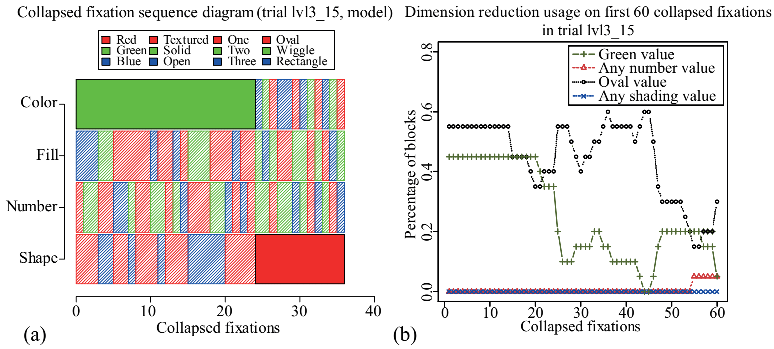


Figure 2.10: (a) Model's collapsed fixation sequence diagram for trial "lvl3_15." The model needed 36 collapsed fixations to find the set. (b) Changing proportion of blocks where the models used dimension reduction in trial "lvl3_15." Proportions are calculated from both slow and fast models' data. Proportions are shown as a function of collapsed fixation position in the sequence and attribute value.

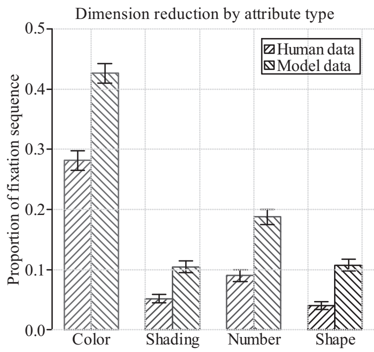


Figure 2.11: Mean proportions of attribute types used in similarity-based scanning. The overall values of all subjects' trials with the highlighted card are compared to the overall values of both models' trials.

It can be observed from Figure 2.10 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 2.10b to Figure 2.5b).

Overall, the saliency and relevancy mechanisms work well in modeling subjects' strategy to use dimension reduction. Combined data from both models show similar order of preference for the attribute types as the human subjects. Figure 2.11 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 2.12a). 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, the model is precise and does not produce such fixations.

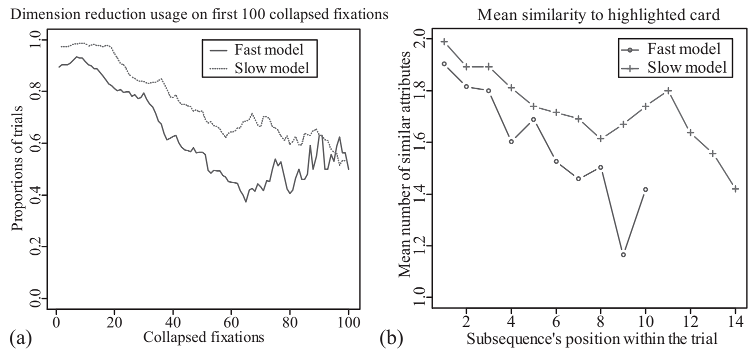


Figure 2.12: (a) Changing proportion of trials in which dimension reduction was used. The proportions are calculated as a function of collapsed fixation position in the sequence. Results are shown separately for slow and fast models. (b) A mean overall similarity of all cards in a subsequence to a highlighted card shown separately for slow and fast models.

Finally, as Figure 2.12a shows, the slow model is more likely to use dimension reduction in the latter part of the trial than the fast model. Overall, the fast model is less biased toward dimension reduction than the slow model, showing an effect similar to one produced by the fast human players (Figure 2.7a).

Dissimilarity-based search

Our experiment revealed that the subjects gradually switch from dimension-reduction strategy to dissimilarity-based search (Figure 2.7b). To test whether the model exhibits the same pattern of behavior as the human players, the same type of analysis was done on collapsed fixation sequences produce by the model. The results can be observed in Figure 2.12b. 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 the graph for the fast model comes to an abrupt end at the 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, a 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, whereas 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 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

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 a deliberate explicit decision, but rather an 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 a 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 reach decision threshold, after which a 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 mechanisms to deal with the limitation. However, recent studies suggest that prediction also plays an 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. A recent study showed that predictability of the environment has significant influence on the decision-making process (Domenech & Dreher, 2010).

In this article, 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.

First, 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.

CHAPTER 3

Pre-Attentive and Attentive Vision

This chapter has been published as:

Nyamsuren, E. & Taatgen, N.A. (2013).
Pre-attentive and Attentive Vision Module.
Cognitive Systems Research, 24, 62-71.

Abstract

This paper introduces a new vision module, called PAAV, developed for the cognitive architecture ACT-R. Unlike ACT-R's default vision module that was originally developed for top-down perception only, PAAV was designed to model a wide range of tasks, such as visual search and scene viewing, where pre-attentive bottom-up processes are essential for the validity of a model. PAAV builds on attentive components of the default vision module and incorporates greater support for modeling pre-attentive components of human vision. The module design incorporates the best practices from existing models of vision. The validity of the module was tested on four different tasks.

Introduction

This paper introduces a general-purpose vision module called PAAV, which stands for **P**re-attentive **A**nd **A**ttentive **V**ision. As the name suggests, the new module incorporates a greater support for bottom-up visual components that are considered pre-attentive in nature, such as multiple feature dimensions to describe visual objects, peripheral vision with differential acuity, iconic visual memory and a decision threshold. The module was developed as an integral part of ACT-R cognitive architecture (Anderson, 2007) that provides a necessary top-down, attentive layer. The reader can refer to Appendix B for a more detailed description of the ACT-R architecture. Here, we will describe only essential details necessary for introducing the PAAV module. By being part of ACT-R, PAAV should be able to model wide range of tasks where both top-down and bottom-up visual guidances are important. ACT-R already has a default vision module and a few extensions for it. However they have drawbacks that PAAV is aimed to solve.

ACT-R's default vision module can be described in terms of a *visicon* and two buffers: *visual-location* and *visual*. *Visual-location* and *visual* buffers essentially represent WHERE and WHAT components of a visual system. The *visicon* represents the visual scene containing visual objects with which an ACT-R model can interact. The *visicon* is considered to be a part of the environment (a monitor screen) rather than part of the model. A model can send a WHERE request to the *visual-location* buffer to find the location in the *visicon* of a potential visual object to encode. Within this request, the model can specify criteria for visual object such as its kind, color, coordinates or size. Given this request vision module randomly chooses one of the visual objects from the *visicon* that exactly matches the given criteria and puts its location information in the *visual-location* buffer. This entire process is instantaneous with no time cost. Next, model can send a WHAT request to the *visual* buffer to encode the object at the chosen location of *visicon*. A WHAT request assumes fixed execution times for both saccade and encoding that in total require 85 ms. This value, although it can be changed by the modeler through a dedicated parameter, is considered as a de facto standard in ACT-R.

EMMA (Salvucci, 2001) is arguably the most used extension to ACT-R's default vision module. EMMA explicitly models saccades including preparation and execution times, path generation and variable landing points. However, EMMA's major contribution is in its ability to model covert attention shifts through variable encoding time dependent on visual object's frequency and eccentricity.

The disadvantage of the default vision module and EMMA is their optimization toward reading tasks or tasks with a relatively simple visual environment where bottom-up perceptual processes can be ignored without sacrificing the model's plausibility and performance. However, ACT-R's vision module is not suitable for tasks where visual stimuli are described with multiple feature dimensions. Such tasks often require theories of scene perception and visual search that are not part of current vision module. The issue is more pressing if one considers the importance of embodied cognition (e.g., Clark, 1997) in problem-solving tasks (Nyamsuren & Taatgen, 2013b) and in everyday human activities in general (Land, Mennie, & Rusted, 1999). Embodied cognition assumes that cognitive control is not purely goal based, but it is also driven perceptually. The simplest example of it is an interference of the salient feature during the task (Theeuwes, 1992). When subjects are asked to look at the scene they tend to look at the most salient parts first. Those salient parts of the scene can interfere with task even if subjects are explicitly asked to not to look at them.

Architecture of the PAAV module

Figure 3.1 provides an overall view of PAAV's architectural components as it is integrated with ACT-R's default vision module. The figure also shows graph of internal workings of ACT-R's declarative retrieval. Also shown is how the retrieval processes is influenced by spreading activations from the vision buffer and PAAV's iconic memory. Details of each of the constituent components of PAAV and corresponding cognitive theories are discussed in the following subsections.

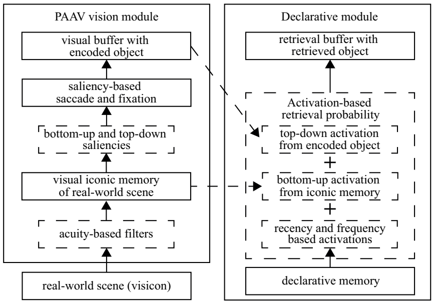


Figure 3.1: Internal workings and external connections between vision, declarative modules.

Feature dimensions

In PAAV every visual object can be characterized by five basic features: color, shape, shading, orientation and size. The features are chosen because of their pop-out nature and importance in guiding visual attention (Wolfe & Horowitz, 2004). Each of those features can have a wide range of values, such as, red and green for color; and oval and rectangle for shape. Currently, PAAV does not support modeler specified custom features. However, it is included as a future implementation milestone.

Peripheral vision

The current implementation of ACT-R’s vision assumes that everything in a visicon is visible to the vision module and consecutively available for information processing. However, human vision is limited in what it can see, especially in the extra-foveal region (Rayner, 1998). PAAV introduces limitations on visibility by assuming that a visual object is only visible if at least one of five features of that object is visible. Visibility of a feature is calculated with an acuity function. We have adopted a modified version of the psychophysical acuity function proposed by Kieras (2010). Kieras’ original acuity function states that for an object’s feature to be visible the object’s angular size s, with some Gaussian noise added to it, must exceed a threshold calculated as a function of eccentricity e:

$$threshold = ae^2 + be + c$$
$$P(available) = P(s + X > threshold)$$
$$X \sim N(0, vs)$$

The free parameters a, b, c and v are to be adjusted for each particular feature. The function works quite well for modeling differential acuity of features. However, the quadratic form in the function makes it less suitable when the object size is particularly small. For example, in their feature search experiment for color, Treisman and Gelade (1980) used visual stimuli of 0.8º×0.6º in size scattered over area of 14º×8º. This feature search experiment cannot be replicated with the above acuity function for color unless parameter a is assigned an extremely low value that is well below the 0.035 used by Kieras (2010).

PAAV uses a modified version of the acuity function to mitigate issue above:

$$threshold = ae^2 - be$$
$$P(available) = P(s > threshold)$$

The constant *c* has been removed since it has no significant influence when object size is reasonably large and too much influence when object size is quite small. Similarly, the Gaussian noise has been removed because of its tendency to introduce too much or too little acuity variation depending on the object size. Next, the coefficient *b* has an opposite sign. It results in less steeper increase in threshold when an eccentricity increases. It also removes the necessity of giving unreasonably small value to coefficient a when object size is small. The free parameter a has been refitted again to 0.104, 0.147, 0.14 and 0.142 for color, shading, size and shape respectively. The parameter *b* has been fitted to 0.85 for color and 0.96 for all other features. We are still in process of fitting parameters for the orientation feature.

Iconic visual memory

Everything PAAV perceives from the *visicon* is stored in iconic memory. Visual features of every object visible via peripheral vision are stored in this memory. As such, the content of iconic memory is not necessarily a complete or even a consistent representation of the objects in the *visicon*.

Information in iconic memory is not treated as consciously perceived visual properties. It is rather perceived as bottom-up visual stimuli on which bottom-up processes can operate. Iconic memory is trans-saccade persistent. Items in iconic memory are persistent for a short duration of time if they are not visible through peripheral vision anymore. The parameter for persistence time is currently set to 4 s, as determined by Kieras (2011), to be a lower bound for a visual memory.

Iconic memory is a model's internal representation of a *visicon*, otherwise visual scene. As such, all WHERE requests are handled with respect to the content of iconic memory via a newly defined *abstract-location* buffer, a replacement to now depreciated *visual-location* buffer. A request may include desired criteria including any of the five feature dimensions or location.

Visual activation

Each visual object in iconic memory is assigned an activation value. The location of the visual object with the highest activation value is returned upon a WHERE request. The activation value is calculated as a sum of bottom-up and top-down activation values. It is adapted from the concept of an activation map used by Wolfe (2007) in his model of a visual search.

Bottom-up activation

The bottom-up activation for a visual object i is calculated based on its contrast to all other objects in iconic memory with respect to each feature dimension k:

$$BA_i = \sum_{j}^{visual\ objects} \sum_{k}^{features} \frac{dissim(v_{ik}, v_{jk})}{\sqrt{d_{ij}}}$$

The $dissim(v_{ik}, v_{jk})$ is the dissimilarity score of two feature values of the same dimension. It is a simplification of a bottom-up activation based on the difference in channel responses used in Guided Search 4.0 (Wolfe, 2007). If two values are the same then $dissim(v_{ik}, v_{jk}) = 0$, otherwise $dissim(v_{ik}, v_{jk}) = 1$. The dissimilarity is weighted by a square root of a linear distance d_{ij} between two objects. Thus the objects farther away contribute less to a contrast-based saliency of the visual object i than the objects closest to it.

Top-down activation

In a WHERE request a model can provide feature values as desired criteria for the next visual object to be located. Those feature criteria are used to calculate the top-down activation value for each visual object in iconic memory. Given a feature criteria k, the top-down activation for visual object i is calculated as:

$$TA_i = \sum_{k}^{feature\ criteria} sim(f_{ik}, f_k)$$

$sim(f_{ik}, f_k)$ is a similarity score of the feature value f_k in WHERE request to a value f_{ik} with the same feature dimension in visual object i. This similarity score is 1 for an exact match and 0 for a mismatch. If the value f_{ik} is not accessible from iconic memory then the similarity score is 0.5. Thus uncertainty is preferred to certain dissimilarity.

Total visual activation

The total activation for visual object i is the sum of bottom-up and top-down activations:

$$VA_i = W_{BA} * BA_i + W_{TA} * TA_i + N$$

W_{BA} and W_{TA} are the weight parameters for the bottom-up and top-down activations respectively. They are set to 1.1 and 0.45. In correspondence with Wolfe (2007), those weights control the unintentional and intentional attentional captures. The bottom-up activation is given a higher weight to compensate for the distance d_{ij} adjustment, which results in the lower bottom-up activation value in comparison to the top-down activation value. N is noise from a logistic distribution with variance σ^2 calculated as a function of a parameter s: $\sigma^2 = s^2\pi^2/3$. s is set to 0.2 by default.

Saccade and encoding

After a visual object has been located with a WHERE request, a model can send a WHAT request. This is essentially the same encoding processes of a visual object from the *visicon* as in ACT-R's default vision module. However, PAAV assumes that the saccade that precedes the encoding has a variable execution time dependent on the saccade's amplitude. Prior to a saccade execution, PAAV calculates its duration and landing point. Salvucci (2001) described a set of formulas to calculate those variables. For calculating the execution duration, we used EMMA's default parameters: 20 ms as a base execution time plus additional 2 ms for an every degree of angular distance covered by a saccade. Differently from Salvucci (2001), we have used two Gaussian distributions around the center of the object to calculate saccade's landing position. The standard deviation for distribution along X axis is calculated as s_g times of the object's linear width, where s_g is a gaze noise parameter set to 0.5. In a similar manner, the standard deviation for Y axis is calculated using object's linear height. Such implementation is in accordance with theory that the saccade's landing position depends on the size of a visual stimulus (Rayner, 1998).

Upon completion of a saccade, PAAV starts encoding. The parameter for encoding time is 50 ms. It is in line with findings that the sufficient information is encoded in the first 45–75 ms of a fixation for an object identification to occur (van Diepen, De Graef, & d'Ydewalle, 1995). Except eccentricity, Salvucci (2001) used word frequency to calculate variable encoding time. However, we believe this approach is not applicable to PAAV where visual object is defined along multiple dimensions. Hence, further study is needed to investigate the object's encoding process in more details sufficient for proper computational modeling.

Visual decision threshold

One of the challenging problems in a visual perception is how does the visual system recognize the absence of a desired visual object. For example, humans can spot the absence of a salient object as fast as its presence in a visual field (Figure 3.2). Similarly, given a WHERE request with specific criteria, how does PAAV know that the desired object is not in iconic memory. One obvious solution is to attend every object in *visicon* and stop when there are no more objects to attend. However, visual search paradigms, such as feature search, show that it is not the case. The visual system is much more efficient and does not require fixation on every item to detect an absence of a target (Treisman & Gelade, 1980; Wolfe, 2007).

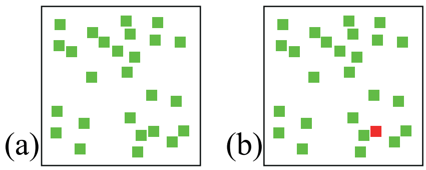


Figure 3.2: Humans can spot an absence (a) of a red object in field of green objects as fast as its presence (b).

PAAV incorporates the concept of a visual decision threshold to decide whether any of the objects in iconic memory will match a given WHERE request. A partial solution is to ignore every object that has zero top-down activation due to complete mismatch. However, results from tasks, such as conjunction search, show that a visual search can be efficient even when distracters partially match the target. PAAV should also be able to filter out objects that match only partially. This is done via simulation of visual grouping based on top-down activation. Given a WHERE request, PAAV returns some object i. Let's assume that, at the time of WHERE request, the distance between object *i* and the gaze position was d_{Th}, and object *i*'s top-down activation was TA_{Th}. When object *i* is encoded these two values are stored and used as a threshold for the consecutive WHERE requests. In the following WHERE requests PAAV completely ignores every object *j* in iconic memory that has $TA_j \leq TA_{Th}$ and $d_j \leq d_{Th}$ where d_j is a distance between object *j* and gaze position. Top-down activation serves as a natural threshold for object selection. Every time a model encodes an incorrect object, the acceptance threshold for the next WHERE request increases up to the activation value of that object. The distance d_{Th} provides a measure that PAAV uses to judge whether it can reliably compare two top-down activation values. It is a simulation of a visual grouping where a cluster of similar objects is grouped together. The d_{Th} can be viewed as an approximate radius of the cluster.

Step by step example

Let us consider an example in which the model is looking for a red square, but there are only three gray squares in the iconic memory. The example is depicted in Figure 3.3. In this example the model is able to notice the absence of a red square after only one fixation.

When the model sends the first WHERE request, the module calculates distance d_j between each object *j* in iconic memory and model's current gaze position (depicted as a black cross). It also calculates the top-down activation TA_j for every object (for the sake of simplicity the bottom-up activation is ignored). All objects receive a top-down activation of one for matching the requested shape feature. Since all objects have the same activation values, let us assume that the module randomly returns the location of the second object as the next object to be fixated. The state of iconic memory after the first WHERE request is depicted as state *A* in Figure 3.3.

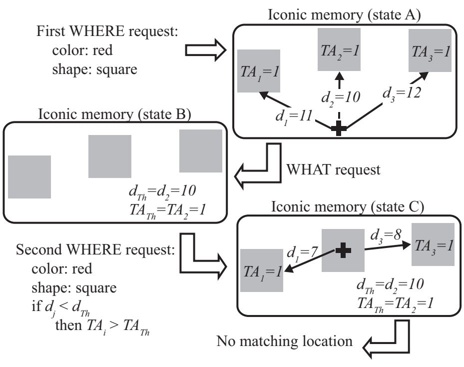


Figure 3.3: Example usage of a visual decision threshold.

The first WHERE request is followed by WHAT request. Given this request, the module stores the value of d_2, the distance between the current gaze position and the second object, as the distance threshold d_{Th}. The module also stores the second object's top-down activation value TA_2 as an activation threshold TA_{Th}. After those steps, the module triggers a saccade execution, changes the gaze position to the location of the second object and encodes the object. At this point iconic memory transitions into state *B*.

Since the encoded object is not a red square, the model sends a second WHERE request. However, this time the model includes the distance and top-down activation thresholds as request parameters along with the color and shape values. The threshold parameters state that if the object's distance from current gaze location is less than distance threshold d_{Th} then the object's top-down activation should be higher than the activation threshold TA_{Th} for the object to be considered a next valid destination to be attended. So, in the current example, there are two unattended objects in iconic memory (state *C* in Figure 3.3). The distances to both objects from the current gaze location, seven and eight respectively, are less than threshold distance of 10. Therefore, both objects should have a top-down activation that is higher than activation threshold of one. This is not the case since both objects again have a top-down activation of only one because of the color mismatch. Hence, the PAAV module lets the model know that there are no more locations to attend. In turn, the model knows that there is no red squared object in iconic memory.

In the example, three gray objects are treated as a cluster of similar objects rather than three individual objects each needing separate attention. The distance threshold d_{Th} can be viewed as a maximum radius of the cluster, while activation threshold TA_{Th} is a maximum dissimilarity threshold within which objects can be considered members of the cluster.

Spreading activation from iconic memory

Lastly, PAAV module introduces spreading activation from visual iconic memory to declarative memory. It has long been observed that visual stimuli can influence the result of a memory retrieval (Wais, Rubens, Boccanfuso, & Gazzaley, 2010). PAAV's spreading activation mechanism was developed to replicate this cognitive process.

ACT-R's declarative memory is a long term memory where knowledge is stored in the form of chunks with slots. One chunk usually represents one concept, while concept properties can be described through values assigned to chunk slots. The model can retrieve only one chunk at the time, and, when there are several chunks that match the retrieval request, the one with the highest activation value has the highest probability of retrieval. There are can be different sources of activation for a chunk, and chunk's total activation is a sum of activations from all available sources.

In the PAAV module, visual objects in iconic memory also serve as sources of activation. Visual feature values from all visual objects spread activation to all chunks in declarative memory that have the same visual feature values as slot values. For example, each green object in iconic memory spreads activation to all green objects in declarative memory. Let us assume there is a chunk k in declarative memory, and it receives a total spreading activation of S_k from iconic memory. Then S_k is calculated as:

$$S_k = W * \sum_{i \in V} (S + ln\ (1 + fan_{ik}))$$

V is a set of all slots from chunk k that have any visual feature value as a slot value. fan_{ik} is a normalized value indicating a number of visual objects in iconic memory that have the same feature value as the chunk k in its slot i. In ACT-R fan_{ik} has to be normalized because a chunk, technically, can have infinite number of slots and the same value in two or more slots. We will not go into the details of normalization since it is ACT-R specific. S, a parameter for the minimum associative strength, indicates the minimum amount of activation that should be spread. W, a parameter for association weight, is a weight of total spreading activation from iconic memory. By default, S and W are set to 0 and 0.7

respectively. With the addition of S_k the default activation equation for declarative memory changes to:

$$A_k = B_k + S_k + P_k + \varepsilon_k + S_k$$

Validation models

This section describes two models that do visual search tasks and a more complex model of a player for a game of SET that requires both top-down and bottom-up cognition. All models are based on ACT-R with the default vision module replaced by the PAAV module. The tasks are simple, yet demand complex cognitive and perceptual processes, and require most of the components of the PAAV module described in this paper. Hence, those tasks serve as a good way to validate the PAAV module. All models use the same default values for PAAV parameters described in this paper with the only exception that top-down activation weight W_{TA} is increased to 3.0 in the model of SET to account for a higher top-down cognitive load.

A model of feature and conjunction searches

The first model was created to do feature and conjunction searches. Both of these visual search tasks involve finding a target among a set of distracters. In a feature search task the target differs from distracters by a single feature such as color (Figure 3.4a). In a conjunction search the target can differ from distracters by either of two features (Figure 3.4b). A feature search is usually an efficient search with reaction time being independent of a number of distracters. On the other hand, reaction time in a conjunction search increases with a number of distracters. Those results are consistent among different studies (e.g., Treisman & Gelade, 1980; Wolfe, 2007; Wolfe, Cave, & Franzel, 1989).

The goal in feature search was to find a red rectangle among green rectangles. In a conjunction search, the model had to find a red rectangle among green rectangles and red ovals. In each trial values for both shape and color were present in near equal amount.

The following experimental conditions were set for the model. In both types of visual search tasks, the set size ranged from 1 to 30. For each set size, there were 500 trials where a target was present and another 500 trials where a target was replaced with a distracter. In total, there were 6000 trials in each of feature and conjunction search tasks. The screen size was 11.3°×11.3°, and the size of each object was 0.85° both in width and height. Within the screen, objects were positioned in a random pattern with the constraint that they should not overlap. The model had to press either "P" or "A" for target being either present or absent. The time of key press was considered as trial end time. The model was reset after each trial.

(a) (b)

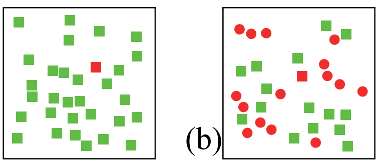


Figure 3.4: Examples of feature search (a) and conjunction search tasks (b). In both tasks the red rectangle is a target.

Figure 3.5b shows the model's mean reaction times in both feature and conjunction search tasks each averaged over trials of the same set size. The black solid line is for feature search task where target was present, and black dashed line is for feature search task where target was absent.

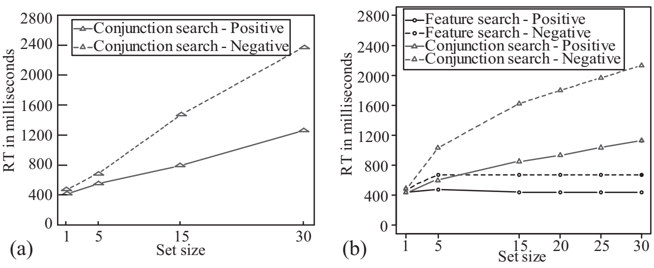


Figure 3.5: (a) Mean reaction times of human subjects in conjunction search as reported by Treisman and Gelade (1980); (b) mean reaction times in feature and conjunction search tasks produced by our model.

In feature search task the model was asked to find any red object. The resulting RT is mostly independent of set size and averages to 446 ms when a target is present and 641 ms when a target is absent. It is consistent with experimental findings where RT for positive trials is also around 430 ms and for negative trials is 550 ms (Treisman & Gelade, 1980; Wolfe, 2007). The model RT remains the same in positive trials due to very high bottom-up activation the target receives due to its color contrast to homogeneous surrounding objects. Top-down activation from the matching color also contributes to the overall saliency of the target. However, bottom-up activation alone is enough to make the target salient enough to attract almost immediate attention. In negative feature search trials all objects in iconic memory have zero top-down activation. It takes the model few fixations to realize absence of a top-down activation after which the model stops searching. As a result, model also produces flat RT line independent of a set size, although slightly higher than in positive trials.

In a conjunction search task the model was asked to find any red rectangle. Figure 3.5 compares the RT produced by the model to the RT[3] obtained by Treisman and Gelade (1980) from their experiment with human subjects. The standard errors for the model RT are too small, and thus are not shown in Figure 3.5b. As the blue lines in Figure 3.5 indicate the RT in both positive and negative trials rise as the set size increases. The slopes, however, are different with negative trials having a significantly higher slope. Linear regression of model's RT on set size gives intercept of 459 ms and 646 ms for positive and negative trials respectively. The slopes are around 23.2 ms/item and 53.8 ms/item. The model results can be compared to those obtained in previous studies (Table 3.1).n this task the distracters are not homogenous. They vary by both color and shape. As a result, there is no guarantee in positive trials that a target will have a higher bottom-up activation than distracters. However, the target always receives higher top-down activation than any other object in iconic memory since it has both matching color and shape. When a set size is small the target's top-down activation is enough to compensate for smaller bottom-up

[3] Confidence intervals or standard errors are not available for human data in feature, conjunction and comparative visual search tasks due to lack of the data in original papers.

activation, and the target almost immediately attracts attention as the most salient object. When the set size is big, there is a higher chance that the target will get significantly lower bottom-up activation than a distracter, which then cannot be compensated by higher top-down activation. Consequently, those distracters with a higher overall activation are attended first which results in RT increasing with set size.

Table 3.1: Comparison of the results of the model's linear regressions of RT on set size to results of linear regression from similar experiments with human subjects.

	Trial type	Slope (ms/item)	Intercept (ms)
Model data	Positive	23.2	459
	Negative	53.8	646
Treisman and Gelade, 1980	Positive	28.7	398
	Negative	67.1	397
Wolfe, Cave and Franzel, 1989	Positive	7.5	451
	Negative	12.6	531

In negative conjunction trials the model should know when to stop the search and report the absence of the target. Since most of the distracters either match color or shape with a target, there are few objects that have zero top-down activation. Hence, the model had to rely on visual decision threshold to filter out partially matching distracters. The model requires on average 53.8 ms/item in negative trials indicating that the model does not need to fixate on every object to realize the absence of a target. Hence, top-down activation serves quite well as a visual decision threshold.

Considering the variations between different studies, the model gives a good fit to experimental findings from previous studies with a slightly higher intercept for negative trials than that found in experiments with human subjects. This is probably due to the fact that the corresponding RT line (Figure 3.5b) is not strictly linear, and as a result has an elevated intercept for an entire linear function. We are still in process of investigating what causes the slightly increased RT for those trials.

A model of comparative visual search

The second model does a comparative visual search, a paradigm proposed by Pomplun et al. (2001). The task involves detecting a mismatch between two, otherwise equal, halves of a display referred to as hemifields (Figure 3.6). The task is a simplified version of the traditional picture matching task (Humphrey & Lupker, 1993) with a major difference that it does not require image processing.

For the model of comparative visual search, we set the screen size to 24º×16º, and the size of each object was 0.6º both in width and height. Those are the same conditions used in the original experiment (Pomplun et al., 2001). The screen was divided vertically in two halves, hemifields. Each hemifield contained 30 objects varying in shape (rectangle, oval and triangle) and color (red, green and blue). Each color and shape value was represented in a trial in an equal quantity. Positions of the objects were generated randomly with minimum margin of 10 pixels from the boundaries of the screen. Two hemifields were identical except one object, the target, which mismatched in either color or shape. The target was chosen at random among 30 objects as well as the type of mismatch. In total, the model had to do 10,000 trials where half of the trials had targets that mismatched color and the other half that had targets with mismatched shape. The model was not aware of the type of mismatch it had to find in a trial. The model was reset after each trial.

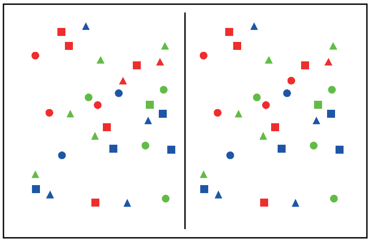


Figure 3.6: An example comparative visual search task where targets are red triangle and red oval in left and right hemifields respectively.

Table 3.2: Comparison of model's mean RTs to those reported by Pomplun et al. (2001). All RTs are in ms.

	Color	Shape	Total
Model	9051	9197	9124
Pomplun et al. (2001)	9903	11997	10950

The model used a very simple algorithm to do visual search. The model starts from a top-left corner of a screen and does following steps:

1. Fixate on any unattended object (further referred to as O1) in the current hemifield.
2. Fixate on any object (referred as O2) in the opposite hemifield that has the same y coordinate as the O1.
3. If O1 and O2 are the same then go to step 1.
4. If O1 and O2 are different then:
   a. Fixate on an object NO2 nearest to O2.
   b. Fixate on O1.
   c. Fixate on an object NO1 nearest to O1.
   d. If NO1 and NO2 are the same then end the trial.
   e. If NO1 and NO2 are not the same then go to step1.

The steps 4a to 4e are necessary to ensure that the module is comparing a correct pair of objects. This uncertainty comes from the fact that when locating a target's twin in the opposite hemifield the model knows only its *y* coordinate and not the *x* coordinate. Therefore, it is possible for the model to fixate on a wrong object that by chance had the same *y* coordinate. To detect such mistakes model also compares two objects from two hemifields that are closest to respective target objects.

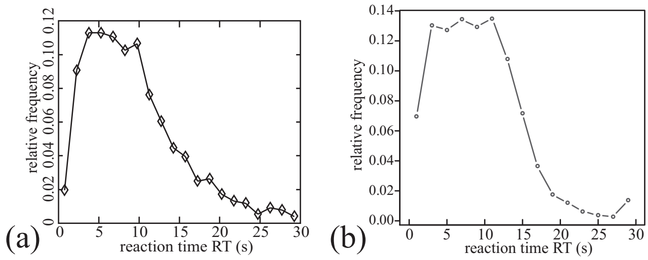


Figure 3.7: (a) Histogram of reaction times in original comparative visual search experiment (Pomplun et al., 2001); (b) histogram of reactions times from 10,000 model trials in comparative visual search.

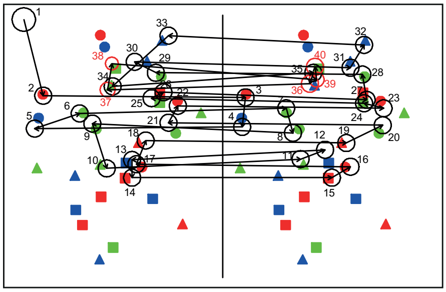


Figure 3.8: Example scanpath produced by the model. Open circles indicate fixations while arrows indicate saccade directions. Numbers are positions of fixations in the fixation sequence. Targets are blue and green triangles at 36th and 37th fixations.

The model's mean RT over all trials was 9124 ms (Table 3.2). On average, the model needed 9051 ms (SE = 79) and 9197 ms (SE = 80) to finish trials where the difference was either in color or in shape respectively. This is a reasonable fit to reaction times reported by Pomplun et al. (2001). The current model was unable to show difference between trials where the mismatch was either in color or in shape.

Figure 3.7a shows a histogram of reaction times from original experiment done by Pomplun et al. (2001). This histogram can be compared to a histogram of reaction times produced by our model depicted in Figure 3.7b. Both graphs show a plateau of short RT between 3 and 10 s, indicating that the distribution of RT produced by the model closely fits the distribution from the original experiment. On average, the model made 37.4 (SE = 0.23) fixations during a trial. This is a close match to 39.6 fixations reported by Pomplun et al. (2001). The model produces nicely structured scanpath (Figure 3.8) even though there is no explicit control of which object should be chosen as O1.

A model of a SET player

In our previous study (Nyamsuren & Taatgen, 2013b) we have described how human players play the card game of SET and how human behavior in that game can be replicated and further explained by an ACT-R model. In that study we have used ACT-R's default vision module and compensated for lacking functionalities with custom code specifically written for that model. In this study we have changed the original model to work with PAAV module. We show how PAAV module helps to describe and explain one of the interesting effects found in original study. Refer to Chapter 2 (Nyamsuren & Taatgen, 2013b) for a more detailed description of the study.

SET players have a tendency to use a dimension reduction strategy while playing a game (Jacob & Hochstein, 2008). That is, they prefer to look for a set among cards that share a common feature value thus effectively reducing the search space by one feature dimension. For example, subjects might look for a set among the cards that have the color green. The choice of a common value heavily depends on an attribute type. For example, an analysis of fixations (Nyamsuren & Taatgen, 2013b) indicates that color, as shown in Figure 3.9a, is used for dimension reduction twice as much as any other feature. The new model easily explains this effect using PAAV's spreading activation from iconic memory and differential acuity. The model also serves well in validating these two functionalities of PAAV module.

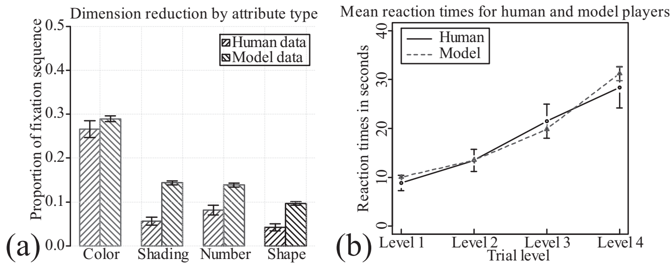


Figure 3.9: (a) Dimension reduction usage by feature type shown both for human players and model; (b) mean reaction times for human players and the model.

In the new model we used three different values for size feature to mimic number of shapes on a card. The actual size of a stimulus representing a card in the visicon also varied based on the number of shapes on the card. Sizes were 9.67º, 23.43º and 37.04º for one, two and three shapes respectively. The model chooses a feature value for dimension reduction by retrieving any of 12 possible values from declarative memory. This retrieval is heavily influenced by a spreading activation from iconic memory. For example, if oval shape is a dominant feature value in iconic memory then the model is more likely to retrieve oval. However, availability of feature values in iconic memory is limited by feature's differential acuity. Therefore, even if the shape value is the dominant value in the visicon, the color value can become the dominant value in iconic memory because it has lower visibility threshold. Therefore, overall color is used more often by model for dimension reduction than other features (Figure 3.9a). The model is not only able to replicate the effect of dimension reduction, but also provides a nice overall fit to human players' mean reaction times (Figure 3.9b).

The Figure 3.10 shows an example collapsed fixation sequence (consecutive fixations on the same card were collapsed into one fixation) produced by the model during one of the trials. The numbers indicate fixations' positions in the sequence. Although we can exactly calculate why the model has chosen to fixate at each location, the the first half of the scanpath looks very chaotic. The model has not fixated on some cards at all. Fixations jump from one end of the scene to other. The absence of visible consistency is very much similar to scanpaths produced by subjects. Finally, there are repetitive fixations on the set cards at the second half of the scanpath. It is very much similar to repetitive fixations subjects produce at he end of the trial when they try to verify if cards indeed form a valid set. So scanpath shows models ability to do two distinct types of search: the one guided by bottom-up visual features and the one guided by a top-down goal.

As our model shows, the tendency of human players to prefer color can be explained with embodied cognition, influence of an external world on our decision making, and the limitations of human peripheral vision.

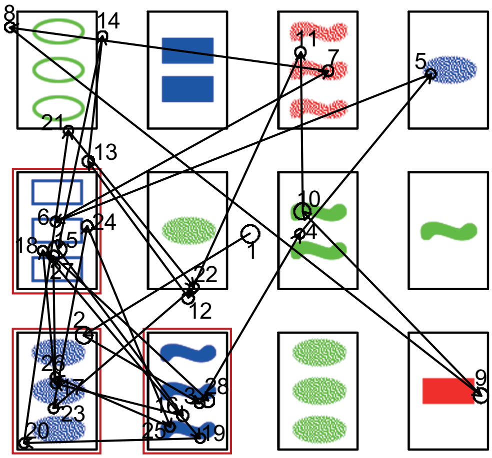


Figure 3.10: An example array of 12 cards where cards with red borders make up a set. Also shown is an enumerated fixation sequence produced by the model.

Conclusion

There are many existing models of the human visual system. We have greatly leveraged from those models by adopting different concepts and integrating them into one module that became PAAV. Our main goal is not to reinvent the wheel, but to create a tool that allows modelers to create cognitively plausible models of tasks that require comprehensive visual system. This is the major difference between PAAV and existing models of a visual system. Models, such as a three-level model of comparative visual search (Pomplun & Ritter, 1999) or Guided Search 4.0 (Wolfe, 2007), were created to perform very specific set of tasks. On the other hand, PAAV was developed to be general enough to model a wide range of tasks. For example, PAAV is highly customizable due to the possibility to adjust any parameter mentioned in this paper. This is why we prefer to call PAAV a module rather than a model. Furthermore, PAAV is not a stand-alone tool, but rather a part of a cognitive architecture. For example, Guided Search 4.0 excels at modeling feature and conjunction search tasks. However, an absence of a general cognitive theory makes it hard to investigate top-down influence in these tasks. On the other hand, ACT-R imposes limitations on what PAAV is allowed to do, but it also gives additional layer of plausibility. The source code for the PAAV module and the models of the visual search tasks described in this paper can be downloaded via http://www.ai.rug.nl/~n_egii/models/.

CHAPTER 4

The synergy of top-down and bottom-up attention in complex task

This chapter has been published as:

Nyamsuren, E. & Taatgen, N.A. (2013).
The synergy of top-down and bottom-up attention
in complex task: going beyond saliency models.
Proceedings of the 35th Annual Conference of the Cognitive Science Society.
(pp. 3181-3186), Austin, TX: Cognitive Science Society.

Abstract

This paper studies how visual perception of a scene is affected by cognitive processes beyond the scene's bottom-up saliency. The game of SET is taken as an example where contrast-based salient parts of a scene are ignored in favor of a larger group of similar elements. Using results from a laboratory experiment and a model simulation we explain how three cognitive mechanisms, differential acuity, visual iconic memory and declarative retrieval, considered together help to explain player's visual perception in SET.

Introduction

Many studies describe how perception of a visual scene is governed by visual bottom-up mechanisms (Rayner, 1998). The conclusions derived in those studies are often based on results from relatively simple tasks involving free scanning or target search. It is widely accepted that visual attention is drawn toward a scene's salient parts (Egeth & Yantis, 1997). This bottom-up saliency is commonly used to explain pop-out effect of items that are increasingly different from its surroundings (Theeuwes, 1992). However, these findings alone may lead to incorrect conclusions if used within a context of more complex problem-solving tasks. It is important to consider a relationship between salience and other cognitive mechanisms to properly understand the inner workings of human mind in such tasks. We use the game of SET as an example of a problem-solving task that gives results that can be interpreted initially as contradictory to the visual pop-out effect. Next, we describe how the same results can be explained within a framework that combines bottom-up saliency with top-down goal-directed attention. For a description of the game of SET refer to Appendix A.

Jacob and Hochstein (2008) studied how bottom-up components of the game, such as attribute value distribution among cards, influences player's strategy. They concluded that players prefer to search for a set inside the largest group of cards that share at least one common value. They referred to a common value as the Most Abundant Value (MAV) and the group of cards that contained it as a MAV group. Sets that were inside MAV group were found sooner than sets outside of the group with an observed probability being significantly higher than a chance probability.

According to the bottom-up saliency mechanism it is expected that players should start a search with visually unique, hence most salient, cards. However, Jacob and Hochstein's finding suggests that player's visual attention is drawn toward larger group of cards that are visually similar. From a perspective of a bottom-up saliency, this is a highly counterintuitive result. Furthermore, another study by Nyamsuren and Taatgen (2013b) revealed that a similarity along particular attribute dimension plays more important role in players' strategy than the saliency of any individual card. Players are more likely to search for a set among larger group of cards with the same color than to attend a card, for example, with a unique shape.

In this paper, we describe a more controlled experiment with set cards with an aim of more in-depth exploration of underlying cognitive processes. In order to use the MAV strategy, subjects must be able to recognize very quickly, which attribute values are most common. The goal of the study is to focus on this particular aspect of SET: to answer the question what cognitive processes facilitate such quick recognition in players. Based on experimental results and model simulations, we describe how three cognitive mechanisms that include visual acuity, visual memory and declarative memory retrieval help to explain MAV effect and bias toward similarity in color attribute.

Experiment

Design and Procedure

14 subjects participated in the experiment. All subjects were students of University of Groningen. Subjects' age ranged from 18 to 27 (M=22). Subjects started each trial by looking at the center of a computer screen. Next, they were shown a 3×4 array of SET cards

for a predetermined duration of time. After image of cards disappeared, subject was prompted to select one of 12 possible attribute values subject perceived as being the most abundant. The experiment consisted of 336 unique trials generated semi-randomly. Trials were divided into a short and a long condition block. The array of cards was shown to subjects for 600 and 2000 ms in the short and long conditions respectively. For half of the subjects, blocks were presented in a reverse order. Within a block, trials were presented in a random sequence. In each block, the MAV group size varied from 6 to 12. There were six trials in each combination of MAV group size and attribute type. Prior to experiment, subjects were asked to do eight, four from each block, trials to let them get familiar with an experiment setup. Results from those trials were not included in the analysis. In addition, subjects' eye movements were recorded. We used the EyeLink 1000, a desktop-mounted remote eye tracker with monocular sampling rate of 500Hz and spatial resolution of < 0.01° RMS. Exactly the same experiment setup and stimulus sizes as in Nyamsuren and Taatgen (2013b) were used in this study.

Experiment Results

Scanpaths

The difference in trial durations also results in quite clear difference in scanpaths. Subjects on average make 8.8 (SE=0.38) fixations in the long condition compared to 2.9 (SE=0.17) fixations in the short condition. Figure 4.1 provides a more detailed look on the trials' fixation counts. There is an 87% probability that subject will make from seven to 11 fixations in the long condition. In contrast, subjects are likely to make only 2 to 4 fixations in 94% of all trials in the short condition.

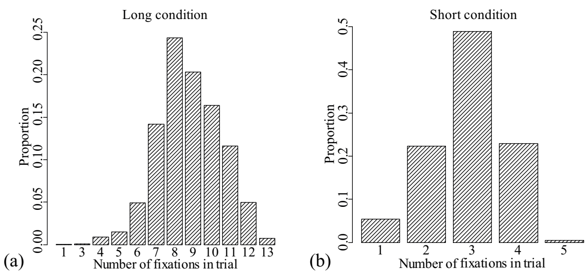


Figure 4.1: Frequencies of fixation counts subjects made during a trial. Frequencies are calculated separately for the (a) long and (b) short conditions.

Figure 4.2a shows mean durations of fixations in a trial. All durations are measured in milliseconds. The last fixations are excluded from the calculation of these means since it is likely that those fixations were interrupted when the time limit was reached. The first two fixations do not show much difference between the short and long conditions. The durations for consecutive fixations in the long condition does not change much. In contrast, durations of third and fourth fixations in the short condition gradually become lower. There can two explanations to this. It may be an artifact of averaging. Smaller number of trials with three or four fixations may be resulting in lower mean. On the other hand, it is possible that shorter durations are deliberate. To test this hypothesis we have also calculated the average duration of fixations in the short condition trials with exactly four fixations. As we have expected, fixations in these trials have much shorter durations than respective fixations in the long condition trials. Therefore, it is indeed possible that subjects were deliberately making shorter fixations in the short condition.

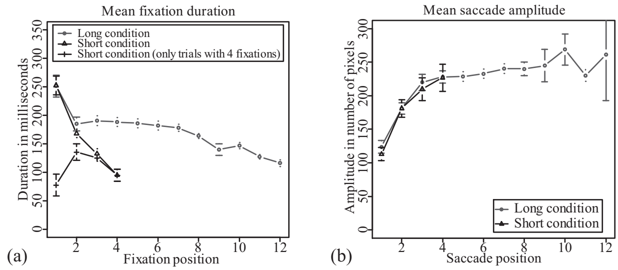


Figure 4.2: (a) Changes in mean fixation durations over course of a trial in the short and long conditions. (b) Changes in saccade amplitude over the course of a trial in the short and long condition.

Figure 4.2b shows how saccade amplitude changes over the course of a trial in both long and short conditions. Amplitude is measured in number of pixels that the saccade covers. There is not much difference between the two duration conditions. However, there is an obvious gradual rise in saccade amplitude as trial progresses. It suggests that there is a specific pattern in subjects' scanpaths.

Accuracy

As Figure 4.3 shows, the overall accuracy increases as MAV group size increases. This is true for both short and long conditions. A test of proportions on pooled data indicate that subjects were more accurate in the long condition than in the short condition, $\chi 2(1, N=4704) = 35.63$, $p < 0.001$. However, as Figure 4.3 shows, there are remarkably small differences in accuracies with respect to group sizes in two duration conditions.

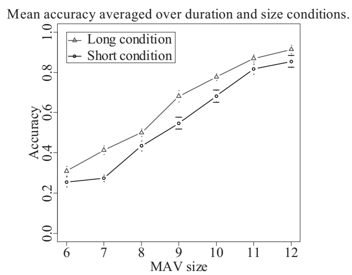


Figure 4.3: Mean accuracies averaged over all combinations of MAV group sizes and duration conditions.

Figure 4.4 shows a boxplot of accuracy variations based on attribute type and duration. We did logistic mixed-effect regression analysis using the duration condition, attribute type and the interaction between the two as predictors. The intercept in the regression model reflects expected accuracy in a short condition trial where the MAV belongs to shading. Relative accuracy increased when MAV belonged to color ($z = 3.19$, $p = 0.001$) and decreased when MAV belonged to either number ($z = -4.142$, $p < 0.001$) or shape (z = -

2.577, p = 0.01). Overall performance in the long condition increased significantly (z = 2.093, p < 0.036). However, there were no significant interactions between duration conditions and attribute types.

Chi-square tests confirmed that subjects were significantly better at identifying the MAV with a color attribute than any other attribute type. Subjects showed little difference in accuracies in the short and long conditions with respect to color ($\chi2(1, N=1176) = 2.91$, $p = 0.088$). It is surprising that, despite the significant difference in average number of fixations made, subjects are equally good at identifying color value in both duration conditions. In contrast, accuracies in the long condition were significantly higher for number ($\chi2(1, N=1176) = 15.283$, $p < 0.001$), shape ($\chi2(1, N=1176) = 16.94$, $p < 0.001$) and shading ($\chi2(1, N=1176) = 4.12$, $p = 0.04$) than in the short condition.

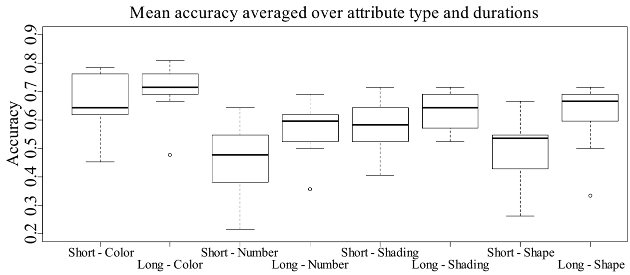


Figure 4.4: Mean accuracies averaged over all combinations of attribute types and duration conditions.

Experiment Discussion

Effect of MAV Group Size on Accuracy

This effect can be explained by the priming of declarative memory by the visual system. There are several studies indicating that the human visual system has some form of iconic memory (Kieras, 2011). It is a low-resolution high-capacity memory where visual information is stored pre-attentively for a short duration of time. The process of gathering information is massively parallel and almost instantaneous. However, information about a visual object is stored as a collection of separate feature channels (such as color or shape) rather than single coherent object (Treisman & Gelade, 1980). Therefore, iconic memory has just enough resolution to guide further attention shifts and encoding.

There is evidence that visual perception can influence processes of memory retrieval (Wais, Rubens, Boccanfuso, & Gazzaley, 2010). It is reasonable to assume that visual stimuli can facilitate memory retrieval of items that are in some form related to the stimuli. Furthermore, we assume the same process applies to iconic and declarative memories. Items in iconic memory facilitate retrieval of similar or related items in declarative memory. In other words, items in declarative memory get activated by items in iconic memory. The strength of such activation depends on the number of items in iconic memory that are related to the item in declarative memory.

This interaction between iconic and declarative memories can explain why subjects find it easier to identify the MAV among larger group of cards. Subjects need to do two tasks: (1) gather visual information through attention shifts and (2) retrieve the MAV from memory when prompted. The second retrieval step is influenced by the content of iconic memory that was gathered during the first step. When MAV group size is large, more values enter iconic memory, and corresponding MAV value in declarative memory receives a higher activation during the retrieval.

Effect of Attribute Type and Duration on Accuracy

The exchange of activations from iconic to declarative memories also helps to explain why subjects are better at identifying color values than values from any other attribute type.

There are studies showing that an ability to capture finer details of a visual scene becomes worse as the distance from a foveal region increases (Nelson & Loftus, 1998). This introduces limitations on what visual features can be gathered into iconic memory. As an object is further away from the foveal region it becomes more likely that some of its features will not enter iconic memory due to limitations of peripheral vision. A feature's acuity threshold defines the maximum distance from a foveal point at which the feature is still recognizable (Kieras, 2011). Compared to other features, color has a higher threshold making it easier to recognize in the peripherals. Thus, color values have a higher chance of entering iconic memory thereby spreading more activation to the same values in declarative memory.

When features, such as shape and shading, have a limited acuity, subjects need to fixate closer to respective visual objects to bring them within threshold distance. This explains why subjects perform better in the long condition trials. Subjects can make more fixations and gather a more complete gist of the visual scene in iconic memory, which then facilitates a more accurate declarative retrieval. This probably explains why subjects tried to make shorter fixations in the short condition trials. Instead of completely encoding every fixated card, the subjects were probably attempting to make short successive fixations and gather as much information as possible into iconic memory via peripheral vision rather than foveal vision alone.

Scanpaths

There are two interesting effects in subjects' scanpaths. Firstly, subjects seem to react to time pressure in the short condition by having shorter fixation durations. This behavior also supports our assumption that iconic memory and peripheral vision play an important role. It is possible that subjects compensate for a shorter duration by making as many fixations as possible and accumulating in iconic memory as much visual information as possible. The pattern of increasing saccade amplitudes provides a clue about preferences of possible fixation locations. Subjects start by fixating on the cards closest to the center of the screen and gradually switch to the cards on the peripherals. These fixations from inwards toward outwards should result in increasing saccade amplitudes shown in Figure 4.2b. In addition to providing more clues about subjects' behavior, scanpaths provide additional measurements besides accuracy against which model fit can be evaluated.

Cognitive Model

Cognitive Architecture

We have used ACT-R cognitive architecture (Anderson, 2007) to develop the model. This section provides a bare essential description of ACT-R features relevant to this study. Appendix B provides a more detailed description of the architecture. We also used Pre-attentive and Attentive Vision (Nyamsuren & Taatgen, 2013a) module as an extension to

ACT-R's default vision module. The PAAV module provides several extra functionalities that are otherwise not supported by ACT-R.

PAAV can pre-attentively capture the gist of a visual scene and store it in iconic memory. The content of iconic memory is updated before and after each saccade and before each time the memory is accessed. The update process is instantaneous from a perspective of model's timeframe. Iconic memory may contain complete information for some visual objects, such as an object's color, shape, shading and size. However, for most visual objects the iconic memory will contain incomplete information (e.g. color only) due to limited acuity. PAAV recognizes that not everything in a visual scene can be resolved by model's peripheral vision at any given moment. In PAAV two parameters, a and b, define differential acuities of color, shape, size and shading with color having the highest acuity. Fitness of these parameters was tested on models of three different visual search tasks and the updated model of game of SET (Nyamsuren & Taatgen, 2013a). An object's feature in iconic memory, although persisting through saccades, decays after a short period of time (currently 4 sec) if not recognizable in peripheral vision anymore. Chapter 3 provides a detailed description of PAAV module.

The content of iconic memory is used to guide the model's visual attention. Visual objects with the highest saliency values are prioritized for visual attention and further encoding. In PAAV, the bottom-up saliency is a sum of saliency values calculated for each feature dimension as a function of contrast to its surrounding. For example, a single red card among green, otherwise similar, cards will be the most salient one and draw the model's attention. PAAV uses a binary measure of similarity: 1 for exact match and 0 otherwise. No adjustable parameters are used in calculation of bottom-up saliency (Nyamsuren & Taatgen, 2013a). It is a simplified version of Wolfe's (2007) saliency function.

In ACT-R knowledge chunks are stored in declarative memory. Each chunk has an activation value that usually reflects chunk's recency and frequency of use by a model. A chunk with the highest activation has the highest probability of retrieval. Besides frequency and recency, a chunk's activation can be increased by the content of iconic memory. Each visual object in iconic memory spreads activation to every declarative chunk with the same features. So depending on the content of iconic memory at the time the results of two same retrievals can differ. The model uses exactly the same set of parameters for declarative retrieval as in the original model of game of SET. Details of those parameters are described by Nyamsuren and Taatgen (2013b).

Model of MAV Task

Model Strategy

Model performed 50 times the same two blocks of trials subjects did. Model starts each trial while fixating at the center of the screen. When cards are shown, models need some time to create a working memory before the first saccade is made. At the same time, model updates its iconic memory with representations of cards. Then model follows with free scanning using bottom-up saliency values to calculate consecutive fixation points. Each fixation is followed by encoding of an attended card. Free scanning stops when time limit is reached and representations of cards disappear. At this point model retrieves any one of 12 possible attribute values from declarative memory. Result of this retrieval depends on content of iconic memory the model has built up during the free scanning. The retrieved value is recorded as model's response for the trial.

Model Accuracy

Model is quite good at replicating subjects' accuracy. Figure 4.5 shows that model's accuracy increases linearly as the MAV group size increases. This effect is present in both the short and long condition. However, just like subjects, the model shows a better performance in the long condition.

The model is also good at reflecting subjects' accuracy depending on combination of attribute types and duration conditions. Firstly, as Figure 4.6, there is a general increase in model's accuracy in the long condition. Except in color, the model clearly benefits from additional time in all other three attributes. Next, Figure 4.6 shows that model is much better at identifying MAV belonging to color attribute than to any other attribute type. Similar to human performance, model's accuracy for color in the short condition is higher than the accuracies for other three attribute types in the longer trials.

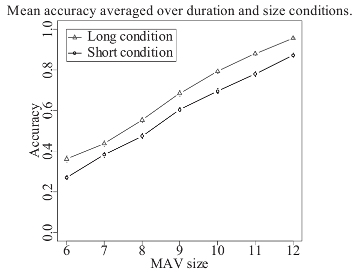


Figure 4.5: Mean accuracies averaged over all combinations of MAV group sizes and duration conditions.

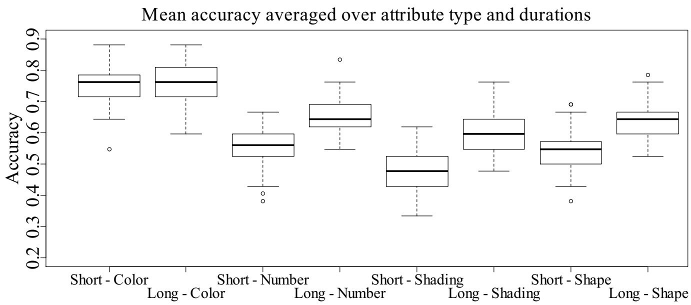


Figure 4.6: Mean accuracies averaged over all combinations of attribute types and duration conditions.

Model Scanpaths

Comparison of model's scanpaths to that of subjects should give additional measure of how well the model fits human data at the level of raw eye movements. Figure 4.7 shows distributions of fixation counts the model made in the long and short conditions. In 99% of all long condition trials, the model made 9-10 fixations. It is within a range of 7-11 fixations subjects made. In the short condition, the model made either two or three fixations. It is also within a range of 2-4 fixations subjects made. As Figure 4.8a shows, model's fixation durations do not differ in the long and short conditions. The lower duration for the third fixation in the short condition is a result of interruption due to duration limit.

The model was able to reproduce a pattern of increasing saccade amplitudes in long condition trials, as it is shown in Figure 4.8b. It was not completely expected since we have not incorporated into the model any deliberate mechanisms to promote this behavior. Because the model makes only one or two saccades in a short condition trial, it is hard to make any conclusive statements about the pattern of amplitude changes. The same model is used in both duration conditions. Hence, there is no reason to expect the model to show different scanpath pattern in the short condition. The lower amplitude for the second saccade in the short condition is most likely due to smaller number of observations from which the mean is calculated. For exactly the same reason, amplitudes for the 9th and 10th saccades drop in the long condition since there are fewer trials that have more than 10 fixations.

The fact that the model shows the same pattern of increasing saccade amplitudes in the long condition suggests that such behavior may be a result of the task environment. Since the model starts the task by fixating at the middle of the screen, the model cannot make saccades longer than the distance between the middle point and a corner of the screen. However, as trial progress the model has a higher chance of making longer saccades (from one corner to another). The same environmental conditions are probably responsible for the increasing amplitude of saccades produced by subjects (Figure 4.2b).

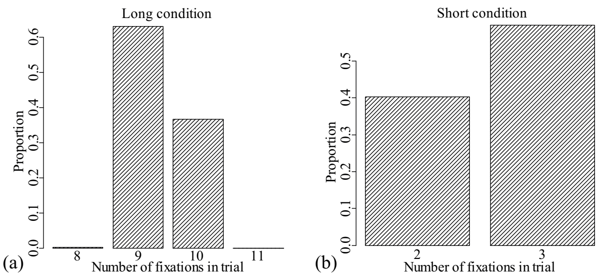


Figure 4.7: Frequencies of fixation counts model made during a trial. Frequencies are calculated separately for (a) long and (b) short conditions.

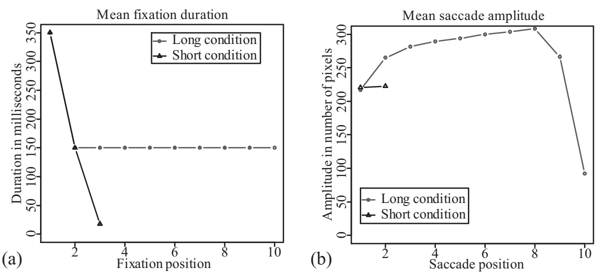


Figure 4.8: (a) Changes in model's mean fixation durations over course of the trial in the short and long conditions. (b) Changes in model's saccade amplitude over the course of the trial in the short and long condition.

Discussion on Model Results

The point at which model has to decide on a choice of MAV is the retrieval from a declarative memory. As model shows, the spreading activation from iconic memory is a major factor deciding the result of this retrieval.

However, it is possible to counter-argue that spreading activation from iconic memory is not necessary, and items in declarative memory are activated directly through visual encoding of similar items. Such mechanism is possible and used in our model. Cards with the MAV have a higher chance probability of getting visual attention and being encoded. As a result, the MAV in declarative memory receives more activation and is retrieved. Although activation through visual encoding would explain subjects' behavior in the long condition, it does not explain why there is a similar effect of MAV group size in the short condition. Neither subjects nor model can encode more than two cards in the short condition, and it is not enough to influence the retrieval. Therefore, bottom-up spreading activation from iconic memory is necessary to explain results in the short condition, and activation through visual encoding cannot explain alone the experimental results. It is likely that subjects rely on visual information in peripheral regions for choosing MAV. Furthermore, the fact that subjects are quite good at identifying the MAV even within 600 ms implies that process of gathering information from peripherals is very efficient. The model simulation suggests that it may be massively parallel and instantaneous.

In the other side, acuity limitations of visual features in peripheral vision can result in incomplete and inaccurate iconic memory. This imperfect internal representation may explain why subjects fail to reach 100% accuracy. It also explains why subjects get better given opportunity to do more fixations in the long condition. More fixations negate the effect of low acuity and result in a more complete representation of the scene inside iconic memory. Furthermore, giving a higher acuity to color in model simulation increases model's accuracy in identifying the most abundant color values in both conditions. This result is similar to the result from the experiment, and, therefore, supports the assumption that human vision is affected significantly by different acuity properties of visual features.

The model produces the same pattern of increasing saccade durations in the long condition without any deliberate mechanisms. It suggests that spatial arrangement and the bottom-up salient parts of the visual scene define the topology of fixation points, more specifically the characteristic fixations from inwards to outwards. In the model, cards around the edges of the screen are not fully visible due to limited acuity. Those cards have reduced bottom-up activation compared to cards around the center of the screen. As a result, the model prefers to fixate on cards closer to the screen center at the early stages of the trial. We were not able to simulate the deliberate reduction in fixation durations subjects have shown in the short condition. Visual processes currently implemented in ACT-R do not provide appropriate mechanisms to simulate this effect.

Discussion and Conclusion

The model fits subjects' accuracies and scanpaths well supporting the hypothesis that the same cognitive processes simulated in the model may also be used by human subjects. More specifically, a combined effect of differential acuity, pre-attentive visual iconic memory and implicit communication with declarative memory can influence our visual perception of the world.

The results from this study can explain player's behavior in game of SET. Player has to decide on a group of cards to be searched for a set. This choice is made through a declarative retrieval of an attribute value that is common among group cards. Similar to the experiment's task, this retrieval is influenced by a content of iconic memory introducing a bias toward a larger group of cards and cards with same color. The retrieved value is used to target attention to specific cards with that value. This top-down control over eye movements overrides the bottom-up saliency of the scene. It explains both why players are better at finding set within a group with many similar cards (Jacob & Hochstein, 2008) and the general preference toward cards with a similar color (Nyamsuren & Taatgen, 2013b). The model of SET player implemented on the same principles described here was able to simulate player's behavior (Nyamsuren & Taatgen, 2013a, 2013b). It is a good example of a case where cognitive mechanisms beyond bottom-up saliency can influence the behavior in a reasonably complex problem-solving task. It implies that not every eye movement pattern can be attributed to bottom-up salient components of the scene.

Subjects are far better in identifying the MAV even in the most difficult conditions. In 600 ms condition with smallest MAV group size, subjects show much higher accuracy than 8% chance probability of success. This result indicates that capabilities of human visual system may be higher than previously expected. The ability to capture a gist of a visual scene from first few fixations is known for a long time (Loftus & Mackworth, 1978). However, it is commonly viewed that functionality of such gist is limited to attentional guidance and providing early structural information for encoding, a preview effect (Rayner, 1998). On the other hand, our study suggests that gist, in form of iconic memory, may be involved in decision-making. It is possible through connections between memories in human brain. In this study, we talked about similarity-based cross-memory activations between iconic and declarative memories. However, it may be possible that similar cross activations exist between other forms of memory.

The model code and the data can be downloaded via following link: http://www.ai.rug.nl/~n_egii/models/.

CHAPTER 5

The effect of visual representation style in problem-solving.

This chapter has been published as:

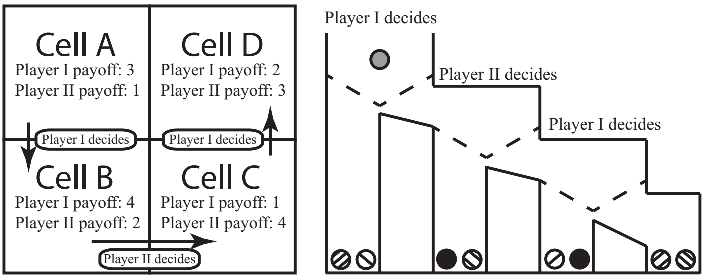
Nyamsuren, E. & Taatgen, N. A. (2013).
The effect of visual representation style in problem-solving:
a perspective from cognitive processes.
PLoS ONE 8 (11): e80550. doi:10.1371/journal.pone.0080550.

Abstract

Using results from a controlled experiment and simulations based on cognitive models, we show that visual presentation style can have a significant impact on performance in a complex problem-solving task. We compared subject performances in two isomorphic, but visually different, tasks based on a card game of SET. Although subjects used the same strategy in both tasks, the difference in presentation style resulted in radically different reaction times and significant deviations in scanpath patterns in the two tasks. Results from our study indicate that low-level subconscious visual processes, such as differential acuity in peripheral vision and low-level iconic memory, can have indirect, but significant effects on decision making during a problem-solving task. We have developed two ACT-R models that employ the same basic strategy but deal with different presentations styles. Our ACT-R models confirm that changes in low-level visual processes triggered by changes in presentation style can propagate to higher-level cognitive processes. Such a domino effect can significantly affect reaction times and eye movements, without affecting the overall strategy of problem solving.

Introduction

More often than not, the study of problem solving is approached from the perspective of logical and rational thinking. In an early study, Weitzenfeld (1984) defined the *isomorphic structure* of a task in terms of its elements and the relationship between those elements. Weitzenfeld further claimed that the isomorphic structure defines the strategy for accomplishing the task. However, Weitzenfeld ignored the significant impact presentation style can have on problem solving even when the isomorphic structure is preserved. Weitzenfeld took two games as an example of structure preservation: Tic-Tac-Toe and Number Scrabble. In Number Scrabble, players select in turn one of the remaining numbers from a pile that contains the numbers from 1 to 9. A player who is first to collect a triad of numbers adding up to 15 wins the game. Tic-Tac-Toe and Number Scrabble are valid isomorphic tasks. Isomorphism is obvious if numbers in Number Scrabble are arranged into a magic square where each column and row adds up to 15. Although it is highly likely that the two games require the same strategy, they are fundamentally different in terms of cognitive processes applied due to differences in visual presentation. While Number Scrabble requires top-down addition and subtraction, Tic-Tac-Toe requires more intuitive spatial reasoning (Michon, 1967). Furthermore, such a difference in presentation styles may affect a player's performance independently of the strategy applied. For example, Michon speculates that JAM, another game isomorphic to Tic-Tac-Toe, is easier to learn than Tic-Tac-Toe due to the fact that it has a different presentation.



(a) (b)

Figure 5.1: The Matrix game (a) used in Hedden and Zhang (2002) and its Marble Drop equivalent (b) described in Meijering, Van Maanen, Van Rijn and Verbrugge (2010).

In more recent work, Meijering, Van Maanen, Van Rijn and Verbrugge (2010) showed that performance can differ significantly in two isomorphic tasks due to a change in visual presentation only. They did a comparative study of subjects' performances in Matrix and Marble Drop games. Hedden and Zhang (2002) originally developed the Matrix game to study higher-order reasoning. Marble Drop is isomorphic to the Matrix game, albeit having a very different presentation style (Figure 5.1). In the Matrix game, each cell contains two separate payoffs for players. The game starts in cell A. Players make decisions in turns and can choose to either switch to a next cell or stay in a current cell. The game finishes when a player chooses to stay, or when cell D is reached. A player's goal is to finish the game in a cell with a maximum possible payoff. The Marble Drop game replaces numeric payoffs with color-graded marbles and cells with bins of decreasing height. Through manipulation of the trapdoors, a player has the choice to drop a marble to either the current bin or to the next set of trap doors controlled by the other player. Although the same chain of reasoning is required in both tasks, subjects showed superior reaction times and accuracy in the Marble Drop game.

This isomorphic structure may be appropriate in explaining players' strategies at a high level, but it is certainly not enough to explain the performance difference shown by Meijering et al. (2010). So how does presentation style change the way humans approach a problem-solving task? After all, it is possible that a change in presentation style imposes a completely different strategy. However, in light of previous studies, it is an unlikely explanation. Alternatively, it can be the case that the overall strategy is the same, but specific actions within that strategy are performed in different ways depending on presentation styles. It is possible that the effects from those relatively small changes can accumulate and result in a significant difference in performance.

There is evidence that individual steps within a strategy in the same task can be done differently, depending on a player's experience. For example, part of a common strategy in Tetris is to rotate and move a token to check where it fits best. This manipulation of tokens is done either physically or mentally, depending on a player's experience (Destefano, Lindstedt & Gray, 2011; Kirsh & Maglio, 1994). A similar effect is also observed in players playing Scrabble. Some players prefer to rearrange letters physically to check what valid words the letters can form (Halpern & Wai, 2007). Other players prefer to do the same step mentally. Experienced players who do mental manipulations generally perform better in both games. These examples show that the same actions in the same strategy can result in differences in performance if done in different ways.

In Tetris and Scrabble, we see a straightforward substitution of a physical process with a mental one. However, a change in presentation style while preserving the isomorphic structure may result in a more subtle substitution of one mental process by another mental process. For example, in the Matrix game, a player may be mostly reliant on top-down processes (arithmetic operations), while in Marble Drop game, a player may also leverage from faster visual bottom-up processes (color perception).

All of the above examples show that problem solving is dependent both on the isomorphic structure of a task and its presentation style. Furthermore, a study of human behavior in a problem solving task should be done with respect to both the overall strategy dictated by its isomorphic structure and the individual cognitive processes imposed by its presentation style. However, the extent of the dependency of problem solving on presentation style is still to be investigated.

Research Objective

The main question in this particular study is how the high-level strategy adapts to the perceptual characteristics of a task. The simplest adaptation would be to keep the top-down strategy the same. However, in such a case, some of the processes done originally by low-level perceptual processes should be transferred to top-down cognitive processes (or vice versa).

We are capitalizing on previous work (Nyamsuren & Taatgen, 2013b) done on the game of SET that provides a more or less complete description of the strategy players use. In SET, the rules and the isomorphic structure of the game largely determine the players' top-down strategy. However, the perceptual elements of the game can have a significant impact on how the strategy is implemented. This makes SET uniquely suited for our study of the effects of changes in presentation style at levels of both the overall strategy and the cognitive processes. The reader can refer to Appendix A for more information on SET.

The earliest study with SET (Taatgen, Oploo, Braaksma & Niemantsverdriet, 2003) found that the time required to find a set increases as the set level increases. A more controlled study in which subjects were presented with 12 cards with only one set in it also showed the same pattern (Nyamsuren & Taatgen, 2013b). The reaction times already show a tendency toward a strategy that finds a set with similar cards faster than a set with dissimilar cards.

Further studies revealed that a player's strategy can be divided into two phases of search: dimension-reduction and dissimilarity-based search. Jacob and Hochstein (2008) showed that players often reduce the search space by looking at groups of cards that share at least one attribute value. They referred to it as a dimension-reduction, since players reduce the number of attributes that they need to compare. The choice of a group of cards is highly dependent on group size: larger groups have a higher probability to be chosen. Surprisingly, subjects need as little as 600ms to extract such complex visual information as identifying the largest group of cards sharing a common attribute value among 12 SET cards (Nyamsuren & Taatgen, 2013c). This search is very much dependent on the visual similarity of the cards. Nyamsuren and Taatgen (2013b) further found that the choice of an attribute value for dimension reduction is not random, because players often prefer color over any other attribute. Furthermore, dimension reduction is mostly used at the beginning of the search, and, if a set cannot be found, players gradually transition to looking for increasingly dissimilar cards.

When players fail to find a set using dimension-reduction, they switch to dissimilarity-based search (Nyamsuren & Taatgen, 2013b). Dissimilarity-based search is used for finding higher-level sets with dissimilar cards. Players still focus on a particular attribute to guide the search. However, instead of looking at cards with the same attribute value, their attention is drawn to cards that have different values for the chosen attribute. Dissimilarity-based search does not allow the use of lower-level similarity-based perceptual processes. One can argue that discriminating between two colors can be done purely with bottom-up visual processes. However, identification of three colors that are all different from each other likely requires some form of top-down control. These factors make dissimilarity-based search a cognitively more demanding process.

The strategy already provides clues about the type of cognitive processes involved. At the beginning of the game, subjects use perceptual processes to identify similar regions of the scene. Those processes are fast, efficient and more suitable for finding lower-level sets with similar cards. At the latter stages of the game, subjects use a slower, but more deliberate and controlled search to find higher-level sets. The preference for dimension reduction explains why subjects need less time to find lower-level sets than higher-level sets.

The question remains whether the preference toward similarity is a result of a deliberate strategy choice or an effect imposed by presentation style. The iconic nature of the presentation style in SET makes it easy to identify similar cards using low-level perceptual processes. This advantage may prompt players to choose dimension-reduction over the more demanding dissimilarity-based search. It certainly can explain why players prefer to start the game with dimension-reduction and require less time to find lower-level sets. As such, presentation style may be directly influencing strategy choice. On the other hand, it is still possible that strategy choice is not dependent on presentation style and may be inherent to the structure of the task. The simplest way to test this hypothesis is to change the presentation style in such a way that the identification of similar and dissimilar groups of cards requires an equal effort. Of course, the task structure should be preserved. If strategy choice in SET is indeed defined by presentation style, then the preference for dimension-reduction should disappear. In other words, players should be equally likely to use dimension-reduction and dissimilarity-based searches at the beginning of the game. It is also possible that a new presentation style may even result in a new strategy. However, if strategy is defined by task structure, then we should observe little change in strategy, even if the presentation style of a task has been changed.

In this study, we used a modified version of SET, in which each card has a set of four words describing its four attribute values. The objectives and rules of the game are the same as in the original version. Word set is isomorphic to the original version of the game. However, the textual representation of cards removes most of the advantages inherent to perceptual components of the game. For example, textual representation should effectively deny subjects the ability to quickly identify a group of similar cards reported by Nyamsuren and Taatgen (2013c). On the one hand, it is interesting to analyze how problem-solving strategies change based on changes in presentation style. On the other hand, it might be the case that the strategy is still the same, and subjects prioritize similarity, despite the absence of a perceptual leverage. In this second case, the question is how cognitive processes are changed and adapted to apply the strategy to different visual presentations.

In our previous studies (Nyamsuren & Taatgen, 2013a, 2013b), we have described a cognitive model of a SET player. The model simulates a player's behavior at the level of individual cognitive processes involved during the game. Those processes include both high-level planning and visual bottom-up perception. The model uses the same strategies described earlier, and maintains an overall top-down cognitive control over the implementation of the strategy. However, individual steps within the strategy are highly dependent on low-level visual processes. For example, bottom-up activation from visual memory plays a key role in the model's choice of using either dimension-reduction or dissimilarity-based search. Using a new experiment, we can verify whether the model is still valid if most of the bottom-up aspects of perception are taken away. Additionally to providing a certain validation for the theories proposed in the paper, the model can also serve as a useful exploration tool. If players apply different strategies in word set, the model can help to investigate the cognitive processes underlying the new strategies including the primary triggers of strategy shift.

Experiment

Ethics

The Ethical Committee Psychology (ECP) of the University of Groningen approved this study. Written informed consent as approved by the ECP was obtained from each participant before conducting the experiment.

Subjects

In total, 20 subjects participated in the experiment. All subjects were students of the University of Groningen. The subjects' previous experience with SET ranged from a few played games to several years of experience. The results from two subjects were excluded from analysis due to extreme noise in the eye movement data caused by a decreased eye tracking accuracy of the camera.

Design and procedure

The experiment was divided into two blocks with different trial types: a block with picture SET trials, and a block with word SET trials. Each block had 32 trials presented to subjects in random sequence. Each trial consisted of 12 cards shown on a computer screen and arranged in an array similar to the one in Figure 5.2. A trial had exactly one combination of three cards that formed a set. As a hint to the subjects, one of the set cards was highlighted by a red border. All trials were generated semi-randomly ensuring a same number of trials per difficulty level in each block. The order of the four attributes in each word SET trial was chosen semi-randomly from the following four possible combinations: (Shading, Shape, Number, Color); (Number, Shading, Color, Shape); (Color, Number, Shape, Shading); (Shape, Color, Shading, Number). It was ensured that all four of the combinations received an equal number of trials. Ten subjects started the experiment with a block of word trials, and eight subjects started the experiment with a block of picture trials.

The trials were essentially the same in the two blocks, except that attribute values were rotated between two blocks. Subjects were not told of this similarity. For example, while converting a picture trial into a word trial, all greens were replaced with blues, blues with reds and reds with greens. In a similar way, the values for other three attributes were rotated as well. This ensured that the trials in the two blocks were the same, but not recognizable by the subjects as such.

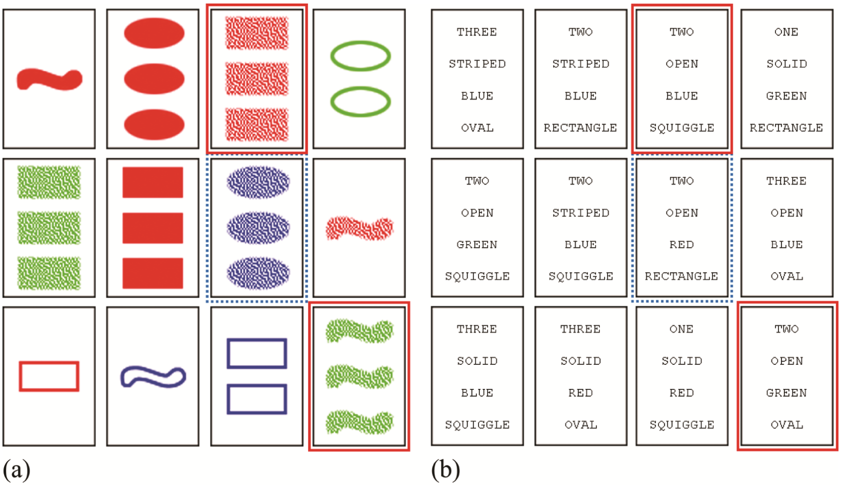


Figure 5.2: An example of a picture trial (a) used in the experiment and its equivalent word version (b). Cards highlighted with a border are the cards that form set (not visible for subjects). The card with a dashed border is a highlighted card.

Prior to the experiment, subjects were asked to do six warm-up trials, three from each block, to let them become familiar with the experimental setup and with picture/word SET. The results from those trials were not included in the analysis. Half of the subjects started the experiment with a block of picture set trials, while the other half started the experiment with a block of word set trials.

An EyeLink 1000 eye tracker was used for recording the eye movements. It is a desktop-mounted remote eye tracker with a monocular sampling rate of 500Hz and a spatial resolution of < 0.01° RMS. The card images were shown on a 20-inch LCD monitor with a screen size of 1024×768 pixels and a screen resolution of 64 pixels/inch. The card images had a size of 124×184 pixels, or 4.02°×5.95°. The horizontal and vertical distances between the images were 80 and 70 pixels respectively, which constitutes to 2.59° and 2.27°. Angular sizes were calculated with an approximate viewing distance of 70 centimeters since the subjects were given a certain freedom for head movement. The gaze position, as 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 the eye tracker was performed at the start and during the experiment, if necessary. A calibration accuracy of 0.8° was considered acceptable. Before each trial, subjects were asked to do a drift correction as an additional corrective measure.

Experiment results

This section provides an analysis based on the subjects' reaction times and eye movements. Most of the eye movement analyses are based on collapsed fixation sequences in which consecutive fixations of the same card are collapsed into one fixation. It is explicitly mentioned when raw fixation sequences have been used in the analysis.

Reaction times

The reaction times provide the first clue about possible strategies used in the two types of the game. According to a mixed effects two-way ANOVA done on log-transformed reaction times, the order of two blocks had no significant effect on the subjects' overall reaction times (*F(1, 16)=1, p=0.331)*, nor did it have a significant effect on reaction times in either picture or word trials (*F*<1*)*. Reaction times in word trials were significantly higher than reaction times in picture trials (*F(1, 16)= 158.913, p<0.0001*), independently of the order of blocks.

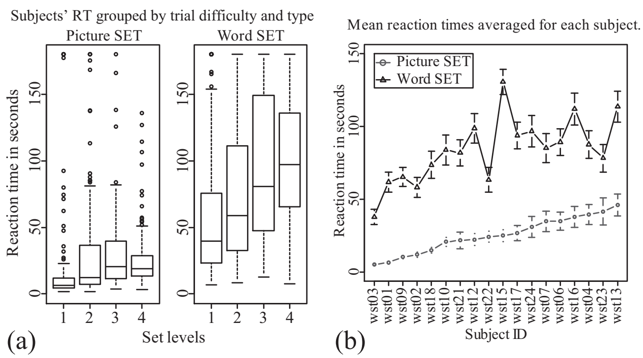


Figure 5.3: (a) Reaction times averaged by trial difficulty and type. (b) Mean reaction times calculated for each subject.

In Figure 5.3, median reaction times for picture trials show the characteristic increase of RT as a result of an increased SET level. We used a mixed-effect linear regression analysis on log-transformed reaction times with set level as a predictor and subjects as a random effect. The analysis showed that reaction time increased on average by 4.54 seconds as the set level increased ($\beta = 0.2939$, $t = 9.774$, $p < 0.001$). This effect is similar to the results from previous studies (Nyamsuren & Taatgen, 2013b).

Reaction times for word trials also exhibit the same effect. However, subjects needed more than twice the amount of time to find sets in word trials than in picture trials. An identical mixed-effect linear regression analysis indicated that reaction time increased on average by 17.2 seconds as the set level increased ($\beta = 0.2591$, $t = 10.44$, $p < 0.001$). There was a positive correlation between the subjects' mean reaction times in picture and word trials: $r(16) = 0.66$, $p < 0.01$. This indicates that subjects who perform well on finding picture sets can be expected to be good at finding word sets as well.

Overall, subjects are better at finding sets with similar cards in both types of the game. It is therefore likely that subjects are using dimension-reduction not only in finding picture sets, but also in finding word sets.

Fixations

It is quite obvious from RT plots above that the number of fixations varies from trial to trial depending on the trial condition and trial level. Figure 5.4 shows how trial length varies in two conditions. The lengths were calculated based on collapsed fixation sequences. A collapsed fixation sequence is a sequence where consecutive fixations on the same card were collapsed into one fixation. Figure 5.4a shows that around 76% of all trials in picture set have 100 or less collapsed fixations. According to Figure 5.4b word trials have significantly more fixations than picture trials.

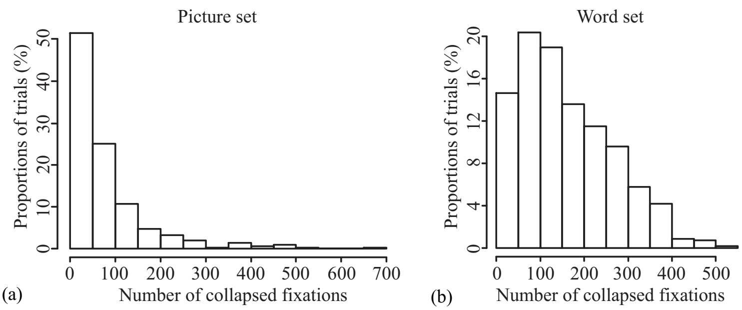


Figure 5.4: Proportions of trials by lengths of collapsed fixations sequences in (a) picture set and (b) word set.

There is also a variation in the number of collapsed fixations depending on the level condition. In picture set, average lengths of collapsed fixation sequences in levels 1 to 4 are 48 (SE = 10), 86 (SE=13), 95 (SE=11) and 82 (SE=8) fixations respectively. In word set, those numbers are 112 (SE=12), 145 (SE=10), 187 (SE=11) and 204 (SE=14) fixations.

Card encoding

The difference in encoding processes can have a significant effect on how information is stored in working/long-term memory, and on how it is further processed. For example, if a card was encoded as a series of visual objects rather than a single object, then it is likely that it will be processed and stored in memory as a series of visual objects.

There is a difference between picture and word trials in terms of the number of fixations required to encode a card. It is hard to quantify exactly how much information about a card is encoded at each instance. However, it is safe to assume that during a fixation in a picture trial, a subject encodes at least as much information as during a fixation in a word trial. Figure 5.5 shows how many consecutive fixations subjects need to encode a card. The proportions were calculated from raw fixation sequences. In picture trials, subjects need one fixation 84% of the time. However, in word trials, occurrences of one fixation per card amount to 43%. Often subjects need two or more fixations to encode a card. This suggests that there is quite a significant difference between picture and word trials in terms of the effort required to encode a card. In word trials, subjects ideally need four fixations, one fixation per attribute, to encode an entire card. Furthermore, in around 4% of the time, subjects had more than four consecutive fixations on the same card. The results suggest that, in word set, a card is encoded as a series of visual objects, as opposed to the single coherent object encoded in picture set.

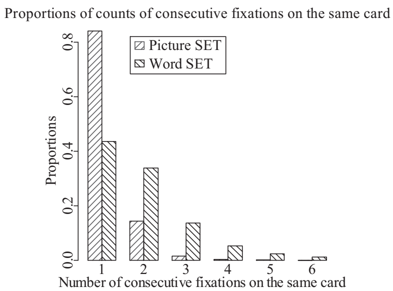


Figure 5.5: Proportions of the counts of consecutive fixations on the same card. Proportions have been calculated separately for picture and word trials.

With respect to picture set, we should be careful in claiming that the process of encoding a card is holistic, even if it results in a coherent visual chunk. Holistic processing is defined as recognition of an entire visual object without any explicit recognition of its constituent parts (Richler, Gauthier, Wegner & Palmeri, 2008). In case of picture set, holistic processing would imply that a card is recognized without explicit recognition of its four attribute values. However, holistic recognition is unlikely because card encoding in picture SET does not violate any of three principles defined by the General Recognition Theory (Wenger & Ingvalson, 2002; 2003). General Recognition Theory states that holistic processing occurs if principles of perceptual independence, perceptual separability or decisional separability are violated in any combination. In picture set, color and shading violate the perceptual independence principle since the two are highly correlated. Therefore, it is probable that the two attribute dimensions are processed holistically. However, other combinations of attribute dimensions violate none of the principles mentioned above. Correspondingly, it is more likely that card encoding is a hierarchical process in which individual attributes are encoded first and then combined into a coherent object.

Processing of word set cards is definitely not holistic. Holistic perception requires a visual object to have a sufficient acuity relative to its distance to the focal point. Text has one of the lowest acuities among common feature dimensions. Kieras (2010) defined a visual angle of one degree as the distance to the focal point within which individual letters are recognizable. Therefore, even if all three GRT principles were violated, holistic recognition of word set cards would be impossible due to the physical limitation of acuity.

The remaining eye movement analyses in this article were based on collapsed fixation sequences, in which consecutive fixations of the same card have been collapsed into one fixation.

Dimension-reduction and dissimilarity-based search

We have calculated the usage of dimension-reduction from eye movement data using the same methods described in our earlier study (Nyamsuren & Taatgen, 2013b). This method finds blocks of consecutive collapsed fixations on cards that have at least one common attribute value. Next, all blocks that have a chance probability above 0.05 were filtered out. Since each card has four attributes, there can be overlapping blocks within the same subsequence of collapsed fixations. Overlapping has been removed by cutting the right-most blocks at the point of overlap. The chance probability has been recalculated for the leftover blocks. Finally, the lengths of the resulting significant blocks were used to calculate the proportions shown in Figure 5.6a. Please, refer to Appendix C for more details on the calculation method.

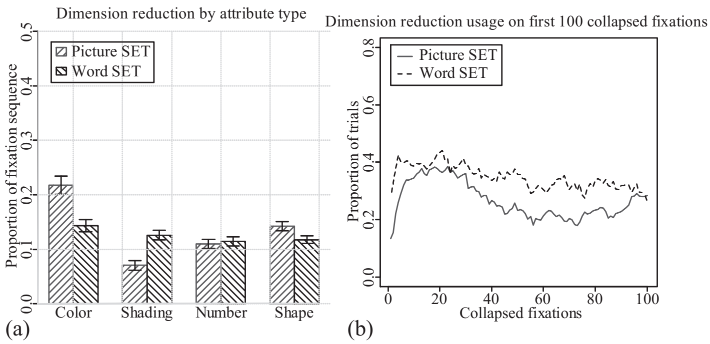


Figure 5.6: (a) The usage of attribute types in similarity-based scanning as a proportion of the trial's collapsed fixations sequence. (b) The probability of using dimension-reduction based on the collapsed fixation's position within trial's collapsed fixation sequence. The probability at collapsed fixation position *x* is calculated as: *N(dr(x))/N(x)*. *N(dr(x))* is a number of trials that have dimension-reduction blocks at position *x*; and *N(x)* is a total number of trials that have at least *x* number of collapsed fixations. Collapsed fixation sequences in word trials are significantly longer than in picture trials. For comparison purposes, sequences and dimension-reduction blocks from word trials were transformed into shorter lengths to match the lengths of corresponding picture trials.

Figure 5.6a shows how often each attribute is used in dimension-reduction during the course of a single trial. Fixation sequences for the trials in which dimension-reduction is impossible with respect to a particular attribute value were removed from analysis. For example, if the highlighted card is green and there are only two other green cards, then the trial is not used for calculating a proportion of dimension-reduction by color. The proportions were calculated separately for picture and word trials. For example, the first bar in Figure 5.6a shows that 22% of the collapsed fixation sequence of a picture trial will be occupied by collapsed fixations in which the subject did dimension-reduction by color. As with the previous study (Nyamsuren & Taatgen, 2013b), in picture trials, subjects show a clear preference to color over any other attribute. However, in word trials, there is hardly any preference to any of the attributes with nearly equal proportions on each attribute type.

Figure 5.6b shows how likely it is that dimension-reduction will be used during the first 100 collapsed fixations of the trial. 76% of all picture set trials had less than 100 collapsed fixations. Therefore, we have chosen to use only first 100 collapsed fixations to avoid increasing noise in the data. Again, the proportions are shown separately for picture and word trials. The probability that subjects will use dimension-reduction during first the 30 collapsed fixations of a picture trial is, on average, around 40%. The probability then goes down with each consecutive fixation. The mixed-effect linear regression analysis done on proportion lines calculated for individual subjects shows that this decrease is significant (Table 5.1). This analysis used all collapsed fixations in positions between 20 and 80. Trials with less than 20 collapsed fixations are not included in the regression analysis since it is reasonable to assume that a set was already found and subjects never switched to a different strategy. Therefore, there is no reason to expect decrease in dimension-reduction in those trials. The number of collapsed fixations is limited to 80 collapsed fixations for two reasons: most trials have less than 100 collapsed fixations; fixations at trial's end usually related to verification (repetitive back and forth fixations to verify validity of a set) rather than visual search. This decreasing pattern is, again, very similar to one found in the earlier study (Nyamsuren & Taatgen, 2013b).

Dimension-reduction also occurs frequently in word trials. In addition, the main and interaction effects of the fixation positions shown in Table 5.1 indicate that there is an overall slow, but significant decrease in the proportion of fixations devoted to dimension reduction as a trial progresses. This indicates that subjects are also using dimension reduction in word trials. However, according to Figure 5.6a there is no clear preference toward a particular attribute value. The visible difference between the two probability lines in Figure 5.6b can be explained by different scanpath structures imposed by differences in visual presentation. This issue is explored further using model simulations. The reader can also refer to Appendix D for additional analysis based on Autoregressive Integrated Moving Average models applied to data on Figure 5.6b. These models treated collapsed fixations as time series and predicted usage of dimension reduction on future unseen fixations. The predictions also showed a decreasing trend in usage of dimension-reduction.

Table 5.1: The result of a linear mixed-effect regression analysis of a predicted proportion of dimension reduction based on a collapsed fixation position and a trial type.

	Estimate	Std. Error	*t* value	*p* value
Intercept (Picture trial)	0.3866	0.0169	22.88	< 0.001
Fixation position	-0.0028	0.0002	-13.84	< 0.001
Word trial	0.0376	0.0152	2.48	0.013
Fixation position and word trial interaction	0.0011	0.0003	3.86	< 0.001

Dimension reduction is a similarity-based strategy. A player searches for a set among cards that are similar with respect to, at least, one attribute dimension. However, subjects gradually stop using dimension-reduction and start looking for higher-level sets. This means that subjects start searching for a set among cards that are increasingly dissimilar. This pattern can be revealed by dividing a trial's collapsed fixation sequence into consecutive series of subsequences, and by calculating the overall similarity of each subsequence to the highlighted card.

An earlier study with SET (Nyamsuren & Taatgen, 2013b) has shown that subjects refixate on a highlighted card approximately every five collapsed fixations, presumably to refresh their memory and to restart a new search subsequence. The following labeled collapsed fixation sequence (given for purpose of example only) "**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. Next, each subsequence's overall similarity to the highlighted card can be calculated.

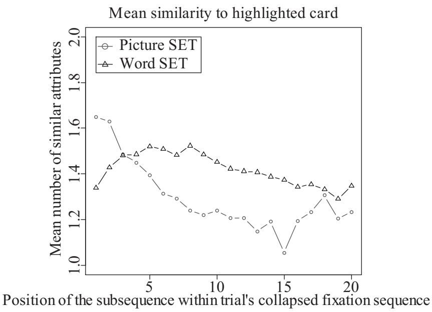


Figure 5.7: The mean overall similarity of all cards in a particular subsequence to the highlighted card. The values are calculated separately for picture and word trials.

Table 5.2: The result of a linear mixed-effect regression analysis of a predicted similarity to a highlighted card based on a subsequence's position and a trial type.

	Estimate	Std. Error	*t* value	*p* value
Intercept (Picture trial)	1.474	0.015	100.3	< 0.001
Subsequence position	-0.020	0.001	-18.9	< 0.001
Word trial	-0.009	0.014	-0.65	0.518
Subsequence position and word trial interaction	0.013	0.001	9.13	< 0.001

The same subsequence-based analysis was done in this study. As shown in Figure 5.7, the mean similarity of fixated cards to the highlighted card decreases over time in picture trials. A linear mixed-effect regression analysis was done on the first 20 subsequences. On average, a subsequence contains 4 collapsed fixations. Therefore, the analysis covers, approximately, 80 collapsed fixations of a trial which close in scale to analysis done on dimension-reduction. The analysis indicates that the decrease is significant (the main effect of subsequence's position on Table 5.2). The decrease is very similar to the one found in previous study (Nyamsuren & Taatgen, 2013b). The same effect is also present in word trials. However, the decrease in similarity, although significant, is very slow (the interaction effect on Table 5.2). This slight decline is nowhere near as big as in picture trials.

In picture set, subjects are clearly transitioning into a dissimilarity-based search as their trials progress. However, the same effect is not conclusive in word set. Nevertheless, considering that subjects were able to find level 4 sets, it is reasonable to assume that dissimilarity-based search was applied in word set trials, despite the lack of evidence in eye movement data.

Systematic versus unsystematic scanpaths

Both dimension-reduction and dissimilarity-based strategies require visual searches. The spatial characteristics of the scanpaths can give insights into differences in visual searches between the two types of tasks.

Figure 5.8a shows a density plot based on the saccades' raw angles. The plot reveals four very distinct distributions centered around 0/360, 90, 270 and 360 degrees. It indicates that in both picture and word trials, subjects prefer to make horizontal and vertical saccades. Such a preference can be partially explained by the grid-like presentation structure of the scene. However, higher peaks in distributions of word trials indicate that preference for vertical and horizontal saccades might be higher in word trials. This difference cannot be accounted for by presentation structure, since this structure is identical in both types of trial.

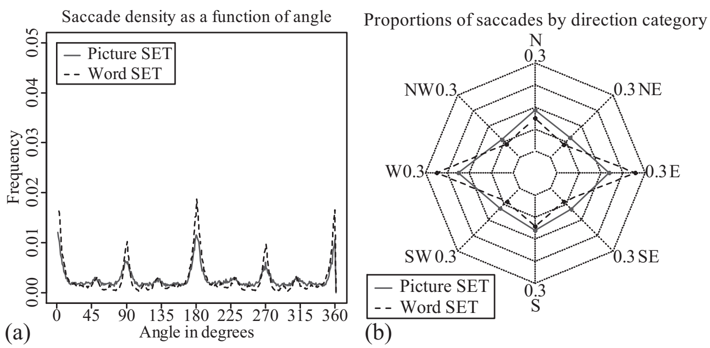


Figure 5.8: A radar chart for the proportions of saccades in each saccade category.

Ponsoda, Scott and Findlay (1995) proposed to measure the systematicity of visual search based on the proportion of diagonal saccades. The higher the proportion of diagonal saccades is, the less systematic the search. Figure 5.8b shows a radar chart with the proportions of saccades in each of the eight direction categories defined by Ponsoda et al.

Firstly, the logistic mixed-effect regression analysis indicates that there is a significant difference in proportions of diagonal saccades made in word and picture trials. The probability of a diagonal saccade in a picture trial is 0.34 (the intercept on Table 5.3). The same probability in a word trial decreases to 0.25 (the negative main effect of word trial on Table 5.3). The low probability of diagonal saccades indicates that visual search is, in general, systematic in both types of trial. The decrease in diagonal saccades in word trials indicates that subjects are less systematic in picture trials than in word trials. This difference in systematicity may account for the differences in distribution shown in Figure 5.8a.

Next, there is a small, but significant effect of trial level on the probability of a diagonal saccade. Subjects are more likely to make diagonal saccades in more difficult picture trials (the positive main effect of Trial level on Table 5.3). However, this effect is greatly reduced in the word trials (the significant negative interaction effect on Table 5.3).

Table 5.3: The results of a logistic mixed-effect regression in which the predicted value is the probability of a diagonal saccade.

	Estimate	Std. Error	*z* value	*p* value
Intercept (Picture trial)	-0.676	0.036	-18.79	< 0.001
Word trial	-0.420	0.035	-12.03	< 0.001
Trial level	0.046	0.010	4.69	< 0.001
Word trial and trial level interaction	-0.035	0.012	-2.96	0.003

The results suggest that in word set, subjects do more structured scannings with more prevalent horizontal and vertical saccades than in picture trials. The increased systematicity of the scanpaths in word set may be related to the lack of visual clues in peripheral regions to guide visual attention. As a result, subjects may be forced to do exhaustive searches in word set, as opposed to more guided searches in picture set. Such exhaustive searches can also explain the lack of evidence for dimension reduction and dissimilarity-based search in the analysis of eye movements from word set trials.

Experiment discussion

Strategy in picture trials

Evidence from the experiment indicates that there is a gradual shift from dimension reduction to dissimilarity-based search in both versions of the game. As discussed in the previous study (Nyamsuren & Taatgen, 2013b), individual steps within a strategy are the same for both dimension-reduction and dissimilarity-based search. Searching for a set is a repeated comparison of three cards. Therefore, a player's strategy ultimately boils down to finding an optimal way to decide which three cards to compare. The subject is already provided with a highlighted card, so he picks a second card and then searches for a third card that may form a set with two already selected cards. If he cannot find a suitable third card, then he picks another card as a second card and starts a new search for a third card. The choice of a second card depends on preferences toward attribute types, and on whether dimension-reduction or dissimilarity-based search is being used. For example, at the beginning of a trial, a subject is more likely to choose a second card that is similar to a highlighted card, since dimension-reduction is preferred at this point. Furthermore, it is more likely that the second card shares the same color with a highlighted card than, for example, the same shading. However, over time, the choice of a second card is geared toward less similarity to a highlighted card. The entire strategy is simple, but effective enough, and, simulated in a cognitive model (Nyamsuren & Taatgen, 2013a, 2013b) gives the same pattern of behavior as exhibited by human subjects.

Strategy in word trials

Based on an initial impression, it appears that subjects are using different strategies in picture and word trials. However, we propose that the strategies are the same. This assumption is supported by a significant positive correlation between the subjects' reaction times in word and picture trials. Furthermore, the fact that subjects need more time to find higher level sets than lower level sets in word trials as well suggests the same bias toward similarity as was found in picture trials. It also supports the assumption that the strategies are the same. Increased reaction times in word trials and other changes in behavioral data can be accounted for by a poor quality of visual information that leads to different cognitive processes being used for strategy implementation. The lack of a visual acuity of an attribute value presented as a text has several implications in terms of different cognitive processes involved in the two types of trials.

Scene gist

The lack of visual acuity in word trials hugely affects subjects' ability to leverage from peripheral vision. In picture trials, attribute values are mostly identifiable in peripheral vision, and a subject can catch the gist (Pollatsek, Rayner & Collins, 1984; Potter & Levy, 1969) of a scene almost instantaneously. Such a gist is used for guiding attention and for encoding specific objects in the scene. If one is looking for a green card, then it is almost immediately obvious where all of the green cards are. In word trials, text is not identifiable in peripheral vision. So the gist that is readily available in a picture trial is absent in word trial. One could argue that in word trials subjects can gradually build up the gist of the scene in visual short-term memory after several initial fixations. However, such a gist will be extremely complex and unpractical, since every card is encoded as a collection of four objects. In addition, visual memory has relatively short temporal persistence, usually within a few seconds (Kieras, 2010, 2011).

Card encoding

There is a difference between picture and word trials in terms of how information about the card is stored in memory once it is encoded. Previously, it was mentioned that in a word trial, a subject needs more than one fixation to encode a card (Figure 5.5). In picture set, a subject fixates on a card and encodes it into a single coherent visual object. This assumption is supported by the fact that, in 83% of instances, subjects need only fixation to encode the card. This object contains information about all four of the attributes of the card. As was mentioned earlier, individual letters can be recognized within a distance of 1º from the focal point (Kieras, 2010). In word set, the vertical distance between two neighboring words is 1.36º (42 pixels). Therefore, we assume that, in a word trial, a subject ideally needs four fixations, one fixation per attribute, to encode an entire card. Moreover, the card is encoded not as a single coherent object, but as a collection of four different visual objects, and is subsequently stored in a memory as such. This introduces an additional overhead of associating four objects with a single card.

Dimension-reduction

The fact that reaction times for lower level sets are shorter than reaction times for higher-level sets (Figure 5.3a), indicates that dimension-reduction is still being used in word trials. The slow, but steady decrease in Figure 5.7 also indicates to a certain preference for similarity-based search. However, the absence of a gist has a significant influence on how a subject does dimension-reduction. There are several studies indicating that different features are not equally identifiable in peripheral vision (Kieras, 2010, 2011; Wolfe & Horowitz, 2004). For example, it is easier to identify color than any other feature. Hence, color is present more prominently in a gist, and subjects are more likely to choose color for dimension-reduction (Nyamsuren & Taatgen, 2013b). However, absence of a gist in a word trial removes preference for any particular attribute. This is the primary reason why Figure 5.6 shows very little difference between attributes in word trials.

Scanpaths

As we have discussed earlier, in a picture trial, a subject tends to pair a highlighted card with a second card and then searches for a third card that can potentially form a set with the pair. The same strategy is applied in word set. Figure 5.9 shows a very nice example. It is a scanpath produced by one of the subjects during the trial shown in Figure 5.2a. As the scanpath indicates, the subject probably formed at least two pairs during the course of the trial. Repetitive back and forth fixations (between 50 and 70 fixations) between the highlighted card C7 and the second card C2 indicate that a pair was formed out of these two cards. Next, the subject scans for a matching third card up until 90th fixation. Since the subject was not able to find a matching card, the new pair was formed with card C3 (between 90 and 103 fixations). A new search for a matching card was done up until 122th fixation, where the subject identified card C12 as a possible match. Indeed, cards C7, C3 and C12 form a valid set, so trial finishes.

Although these are the same basic steps as in a picture trial, there is one significant difference between scanpaths. The search for a third card in picture set is supported by a scene gist. If a subject is looking for a green card, then it is immediately identifiable where all of the green cards are. However, in word set there is no gist to make such targeted attention shifts. Instead, a subject needs to fixate on every single card to check whether a card has desired attribute values. Indeed, the searches for third cards shown in Figure 5.9 are very much exhaustive.

This difference between exhaustive and targeted searches explains why scanpaths in word trials have less diagonal saccades than scanpaths in picture trials. So what appears to be a systematic visual search might rather be a search done out of necessity due to lack of proper visual features to aid the search in peripheral regions. Similarly, abundance of diagonal saccades in a picture trial is not the result of an absence of systematicity as would be suggested by Ponsoda et al. (1995). It is rather the result of subjects taking a targeted "shortcut" by using visual features that can be processed by peripheral vision during a systematic search.

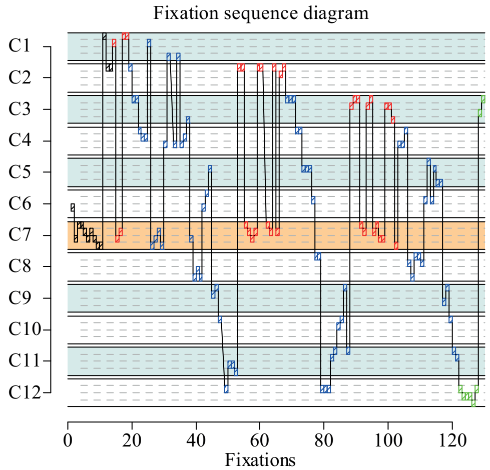


Figure 5.9: An example of an annotated raw fixation sequence produced by wst03 during the trial shown in Figure 3a. Each lane with solid boundaries represents a card, whereas each sublane with dashed boundaries represents an attribute within a card. The lane labeled as C7 is the highlighted card. The other two cards that belong to the set are C3 and C12. Each rectangular block represents a fixation on a card's attribute value. Red blocks represent fixations where a subject paired a highlighted card with another card, while blue blocks represent consecutive search for a third card. Green blocks are fixations where the subject found a set and made final verifications.

Types of visual search

There are several competing explanations of how humans accomplish visual search tasks. Early studies of visual search suggested that visual search may be sequential (only one object is encoded at a time), because of the positive slope produced by the RT×set size function. The visual search observed in the set tasks is clearly not sequential. As was discussed earlier, subjects need as little as 600ms to extract such complex visual information as identifying the largest group of cards sharing a common attribute value among 12 SET cards (Nyamsuren & Taatgen, 2013c). This result clearly refutes the possibility that visual search is a purely sequential search. There are certainly some parallel processes involved.

An alternative explanation of sequential search is a limited-capacity parallel search (Townsend, 1990). In this paradigm, several visual objects can be encoded at the same time, but the number of objects is limited by the capacity of the visual process. It is highly unlikely that limited-capacity visual search is used in set tasks. SET belongs to a group of tasks under the comparative visual search (CVS) paradigm. Previous studies showed that the eye movement patterns in CVS tasks show clear signs of well-structured sequential search (Pomplun, Sichelschmidt, Wagner, Clermont, Rickheit & Ritter, 2001). Eye movement data from the two set tasks also suggest that searches are not parallel. For example, just like other fixation sequences, the fixation sequence shown in Figure 5.9 exhibits signs of a highly structured sequential search. In another example, we found subjects often end the trial with verification fixations. Verification is characterized by repeated back and forth fixations on three cards forming a set. Such fixations were observed in both picture and word set trials. Such fixations would not have been necessary in limited-capacity parallel searches. Similar verification fixations were also observed by Pomplun et al.

As a third alternative, Wolfe proposed that all visual search tasks require the deployment of attention to the target, but such a deployment is guided by pre-attentive parallel processes (Wolfe, 1992, 1998; Wolfe & Bennett, 1997). Furthermore, Wolfe suggested that search tasks only vary with respect to the degree in which they can use parallel processes to deploy attention. It is quite likely that the visual search used in the set tasks follows Wolfe's theory. It certainly explains why subjects can quickly capture scene gists (Pollatsek, Rayner & Collins, 1984; Potter & Levy, 1969), but also exhibit sequential visual search behavior such as in Figure 5.9. Conformance to Wolfe's theory also adds additional credibility to our explanation of why visual search strategies are essentially the same in picture and word sets. Visual search is the same in the two tasks, but the use of parallel processes in word set is impaired by the poor acuity of the text.

Cognitive Models

The major question we want to answer using cognitive models is whether the differences in cognitive processes that were described in the previous section can really account for the behavioral differences subjects have shown in picture and word trials. Our previous studies (Nyamsuren & Taatgen, 2013a, 2013b) have already described the cognitive model for picture set. For this study, we have reused the same model to simulate human behavior in picture trials. We have also developed a second model that does the word trials.

The two models are nearly identical. Both models use the same set of values for adjustable parameters and follow the same strategy of playing the game. The only difference lies in the processing requirements for the two types of cards.

Cognitive architecture

We have used the ACT-R cognitive architecture (Anderson, 2007) to develop the models. ACT-R consists of several modules, such as a Vision module for handling visual processing, a Declarative module for simulating declarative memory, and a Goal module for tracking the model's state and objectives. The modules mostly communicate with each other via the Procedural module, which allows the modeler to write task specific production rules. However, in limited cases, modules can also spread activation to other modules simulating low-level cognitive processes. Figure 5.10 shows the internal working of the most important modules in detail. A description of the figure will be provided next.

We used several extra modules that are not part of ACT-R by default. The extra module most important to the task is the Pre-attentive and Attentive Vision module (Nyamsuren & Taatgen, 2013a) or PAAV for short. The PAAV module provides several functionalities that are otherwise not supported by ACT-R's default vision module. The other two extra modules are Threaded Cognition (Salvucci & Taatgen, 2008) and Base-Level Inhibition (Lebiere & Best, 2011). With Threaded Cognition, we assume that there are two separate and parallel meta-controls governing the overall top-down strategy and the bottom-up visual attention shifts, respectively. Lastly, the Base-Level Inhibition module provides a short-term activation inhibition of items in declarative memory. This module is necessary for modeling complex short-term tasks in which several alternatives need to be stored in and retrieved from memory.

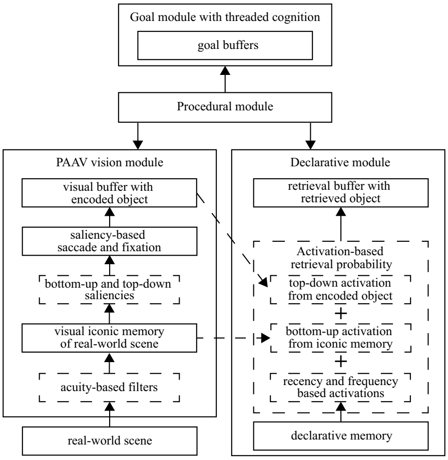


Figure 5.10: Internal workings and external connections between vision, declarative, goal and procedural modules of ACT-R architecture. These four modules provide the most of the functionalities necessary for modeling SET tasks.

Differential acuity

PAAV recognizes that not everything in a visual scene can be seen [by the model] at any given moment. Human vision is limited, especially in the extra-foveal region (Rayner, 1998). The further away an object is from a current focal point, the harder it is for the human visual system to recognize its features. Furthermore, different features, such as color or shape, have different acuities (Kieras, 2010, 2011). For example, color has a higher acuity than shape. This means that the visual system will be able to recognize the color, but not the shape, of an object that is in a certain distance from the foveal point. The PAAV module uses different acuity functions for color, shape, size and shading with color having the highest acuity. Text is also supported by PAAV in a sense that any word is treated as a shaded rectangular object of a same size as the word. However, there is a separate acuity function for recognizing the pattern of individual letters in the word. In order for individual letters to be recognized, a word should be inside the foveal region. PAAV considers this region as a circle with a radius of one degree of angular distance from the point of fixation. This estimation is provided in (Kieras, 2010).

Scene gist and visual iconic memory

It is often reported that human vision can pre-attentively capture the gist of a visual scene (Pollatsek, Rayner & Collins, 1984; Potter & Levy, 1969). This is a quick and parallel process that captures just enough details to further guide visual attention to informative parts of the visual scene for a finer grained analysis. The PAAV module also captures the gist and stores it in iconic memory. Iconic memory may contain complete information for some objects, such as an object's color, shape, shading and size. However, for most visual objects, iconic memory will contain incomplete information (e.g. color only) due to limited acuity. Furthermore, an object's features in iconic memory, despite trans-saccadic persistence, decay after a short period of time (currently 4 sec) if they are not accessible via peripheral vision anymore.

Attentional guidance

It is well known that the human visual system prioritizes parts of the visual scene for attentional capture (Mackworth & Morandi, 1967). This process is a combination of bottom-up and top-down guidance (Knudsen, 2007; Treue, 2003). Bottom-up guidance draws attention to the parts of the visual scene that are most salient due to the inherent properties of the scene. For example, a single green card among red cards will draw attention due to a pop-out effect. On the other hand, top-down guidance draws attention to the parts of a scene that are relevant to the current task at hand. For example, if a player is looking for a green set, then all green cards will be prioritized for attentional capture, while all non-green cards will be inhibited. The PAAV module mimics this process by calculating top-down and bottom-up saliency values for every object in iconic memory and choosing the one with the highest overall saliency as the next point of attention.

Spreading activation from iconic memory

In ACT-R, knowledge chunks are stored in declarative memory. Each chunk has an activation value that reflects its recency and frequency of retrieval. The chunk with the highest activation has the highest probability of retrieval. However, it has also been observed that visual stimuli can influence the result of memory retrieval (Wais, Rubens, Boccanfuso & Gazzaley, 2010). The PAAV module simulates this effect whereby each visual object in visual iconic memory spreads activation to every matching chunk in declarative memory. So, depending on the content of iconic memory, results from two identical retrieval requests can differ. ACT-R's default vision module also allows spreading activation from an encoded visual object to declarative memory, thereby simulating a more top-down influence.

Model details

Any ACT-R model is essentially a set of production rules expressing task specific instructions. A production rule consists of a left-hand side condition part and a right-hand side action part. A production rule fires when all of the conditions in the right hand side are met. Only one production rule can fire at a time. For example, if the condition part says that the current goal of the model is to attend the highlighted card, then the action part tells the PAAV module to shift attention. The right-hand side action part can also set the goal of the model to a new one. The production rules in models of set tasks implement the strategy described next.

Strategy

Although we used two separate models, both of them use exactly the same strategy. This paper describes the strategy only on a level of details necessary to understand the inner workings of the model. Please, refer to Chapters 2 and 3 (Nyamsuren & Taatgen, 2013a, 2013b) for a more complete description. The following is a description of the model's general strategy:

1. Focus attention on the highlighted card *HC*. Let $Card_{HC}$ be a set of four attribute values in the highlighted card.
2. Retrieve any attribute value V_{DM} from declarative memory with A_V being 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 by defining search space G as a group of cards that has V_{HC}. If $V_{DM} \neq V_{HC}$ then use dissimilarity strategy by defining search space G as a group of cards that does not have V_{HC}.

5. For every card *C2* in search space *G*, search for a third card *C3* that forms a set with *HC* and *C2*. If a set is found, then finish the trial.

6. If there is no more card *C2* to choose from search space *G*, then go back to step 1.

The critical step is step 2, in which a top-down influence (the highlighted card) and a bottom-up influence (the prominence of attribute values in iconic memory) determine what attribute value the model is going to pursue.

Although the two models use the same strategy, there are several essential points of difference that rise due to presentation differences.

Visually encoding a card

The model for picture set can encode all four values of a card in a single fixation, since those values are perceived as four integral features of a single object. However, the model for word set has to fixate on each individual value of a card, since each word is treated as a visual object of its own. Therefore, instead of just one fixation, four fixations are needed just to encode all four values in a word set model.

Scene gist in visual iconic memory

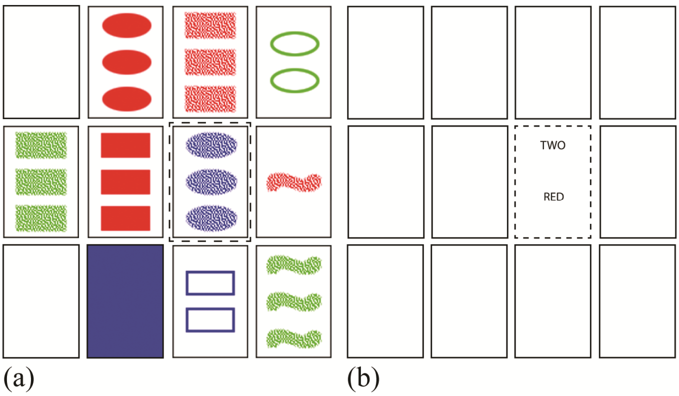


Figure 5.11: A visualization of the content of the model's iconic memory after the first fixation on the highlighted card (cards with dashed boundaries) in (a) picture and (b) word trials. Those are the same trials as shown in Figure 3.

The model for finding picture sets has a reasonably detailed representation of the trial in its iconic memory from the start. The acuity limitation of a text prevents the model for word trials from building up iconic memory with the same level of detail. Figure 5.11 contrasts the contents of iconic memories of the two models after the first fixation on the highlighted card was made. Except for three cards on the left, the model for a picture trial has near complete information about the visual scene (Figure 5.11a) in its iconic memory. This information is enough to calculate both bottom-up and top-down saliencies for cards to guide attention shifts. The model for word trials has barely any information about the visual scene (Figure 5.11b). All it has is an encoded value "TWO" for the Number attribute and a pre-attentively recognized pattern of individual letters for "RED". There is no information to guide attention shifts from either bottom-up or top-down perspectives.

Dimension-reduction

Both models have a tendency to use dimension-reduction in the early stages of a trial. Spreading activation from the encoded highlighted card biases the retrieval process. As a result, values that belong to a highlighted card have a slightly higher chance of retrieval (step 2 in the models' strategies). However, iconic memory also influences the retrieval process through spreading activation. Color values have the highest acuity, hence a higher chance of entering into iconic memory. More color values in iconic memory spread more activation to respective values in declarative memory. As a result, color values have a higher chance of being retrieved from declarative memory and used in dimension-reduction. In picture trials, this process explains why subjects often prefer color for dimension-reduction to any other attribute (Figure 5.6a). However, in a model for word trials, iconic memory has a negligible influence on the retrieval process, since it is almost empty. Hence, all attributes have a near equal chance of retrieval, thereby removing any possible preference toward a specific attribute.

Scanpaths

The model for picture set can prioritize locations for attention shifts to the parts of a scene that are both salient and relevant to the current goal reasonably well. For example, in the trial shown in Figure 5.11a (and assuming the model is looking for a set among blue cards) it can predict with high accuracy where all of the blue cards are based on the content of iconic memory. Such luxury is not available to the model for word set. Its iconic memory is almost empty, and the model has to shift attention based purely on the prior knowledge of the structure of the scene. It results in a significant difference between picture and word set models in terms of how they scan the search space *G* (steps 4 and 5 in the models' strategies). The picture set model is fairly efficient since it scans only those cards that belong to search space *G*. The word set model cannot identify pre-attentively which cards belong to search space *G*, so it scans all cards. Such scanning is done by shifting attention to the next closest card.

Model results

Both picture and word set models had to play 100 times through the same block of 32 trials that the subjects did during the experiment. The following sections discuss the results of these runs.

Models' fits

It is extremely hard to properly estimate a general fit of a model simulating a task as complex as SET. If the models of picture and word set, respectively, are valid, they should produce fixation sequences similar to sequences of human subjects. We have compared subjects' collapsed fixation sequences to collapsed sequences produced by models. The resulting comparison scores were taken as an estimation of the models' fits.

We have used *ScanMatch* (Cristino, Mathôt, Theeuwes & Gilchrist, 2010) as a method for comparing fixation sequences. *ScanMatch* provides several mechanisms that make it more suitable for comparing eye movement data than more conventional methods, such as an estimation of the Levenshtein distance (Levenshtein, 1966). *ScanMatch* is based on the Needleman–Wunsch algorithm (Needleman & Wunsch, 1970) that uses a substitution matrix to maximize the similarity score resulting from a comparison of two sequences. That substitution matrix contains scores for aligning every possible combination of two elements. Comparisons based on that substitution matrix allow for alignments based on overall similarity patterns rather than the binary equalities of individual elements in the sequence. This feature is important considering a certain degree of randomness in the pattern of fixations that arises when a scene is relatively complex.

We created substitution matrices for each trial. Each matrix contained scores for aligning a trial's cards with one another. Scores were calculated based on the similarity of two cards with respect to the highlighted card in the trial. Next, a subject's collapsed fixation sequence for each trial was compared to the corresponding 100 collapsed fixation sequences produced by the model on the same trial. Finally, the overall mean scores were taken for each subject as an estimation of the model's fit to that particular subject's data. The model's general fit to the experimental data was calculated as a grand mean of all of the subjects' scores. The scores were calculated separately for picture and word trials. We also generated random fixation sequences and compared them to the subjects' collapsed sequences the same way the models' collapsed fixation sequences were compared. This gives chance-based lower boundaries for similarity scores against which the models' scores can be compared.

Table 5.4 shows grand means of similarity scores calculated for the models' sequences and random fixation sequences. The fixation sequences produced by the two models have significantly higher similarity scores than the fixation sequences generated randomly. The significance was calculated separately for the two trial types using one-way within-subject ANOVA. The analysis result indicates that the similarity of the models' fixation sequences to the subjects' sequences is significantly above chance level. We can conclude that both models have an explanatory capability and capture the subjects' behavior at least in some degree.

Table 5.4: Grand means of similarity scores after comparing each subject's collapsed fixation sequences to randomly and model generated fixation sequences.

	Model	**Random**	***F(1, 17)***	***p* value**
Picture set	*M*=-0.347, *SE*=0.013	*M*=-0.438 *SE*=0.008	219.5	< 0.0001
Word set	*M*=-0.386 *SE*=0.007	*M*=-0.460 *SE*=0.006	359.6	< 0.0001

Finally, we did a cross comparison of fixation sequences between human subjects. The resulting grand means are *M=-0.317* (*SE=0.010*) and *M=-0.294* (*SE=0.007*) for picture and word trials respectively. Those scores are the upper boundaries of similarity against which the models' fits can be evaluated. For example, in picture set, a perfect model of a general human player will produce a fixation sequence with the similarity score of -0.317 when compared to a human fixations sequence. However, any model that has no explanatory power of human behavior in picture set will produce the similarity score of -0.438 or below. The picture set model has the similarity score of -0.347. It is not a perfect fit, but still quite better than the model with no explanatory power. Both models definitely do not produce the best possible fit. However, some deviation is expected, considering the quite complex nature of the task.

Reaction times

Figure 5.12 shows boxplots of reaction times for both picture and word set models compared to the respective reaction times from subjects. Both models' reaction times increase as a function of set level. This is to be expected, since at the beginning stages of the trial, both models prefer to search for a set among cards that are similar to a highlighted card. This is essentially a search through dimension-reduction, since the models ensure that cards share at least one attribute value with a highlighted card.

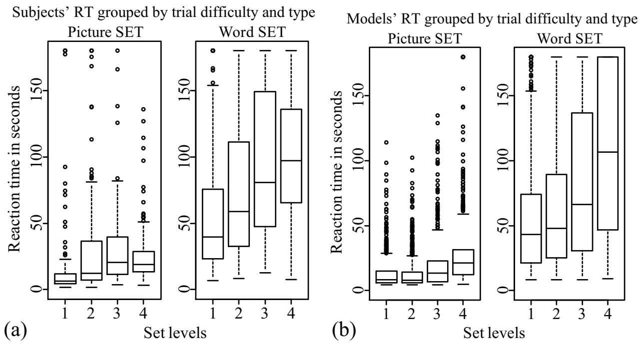


Figure 5.12: The subjects' reaction times (a) compared to reaction times from 100 runs of the experiment with picture and word set models (b).

Dimension-reduction

The model for picture set has a relatively high tendency for dimension-reduction, as is shown in Figure 5.13. Color is prioritized for dimension-reduction more than any other attribute type (Figure 5.13a), which is similar to our experimental results (Figure 5.6a). This priority is a result of color values being more readily available in iconic memory than values of any other attribute type due to their higher acuity. Dimension-reduction is preferred at the beginning of the game (Figure 5.13b) with its usage gradually decreasing as the trial progresses. At step 2 of the model's strategy, the values of the highlighted card have a higher chance of retrieval. However, those same values get inhibited on consecutive retrievals. This simple process results in an overall pattern of dimension-reduction that resembles the one shown by human subjects (Figure 5.6b). As in our previous study (Nyamsuren & Taatgen, 2013b), the picture set model exhibits a higher tendency for doing dimension-reduction than the human subjects. The most likely explanation for this difference is that not all instances of dimension-reduction were captured from the human data. Because only blocks of consecutive collapsed fixations with a chance probability of less than 5% have been included in the analysis, occasional wandering fixations produced by human subjects can significantly decrease the calculated proportions of dimension-reduction in the human data. On the other hand, the model's attention shifts are precise with no wandering fixations or other forms of noise artifacts.

Figure 5.13 and Figure 5.6 show that the model for word set is able to replicate subject behavior even better than the picture set model. The model exhibits a very slow, but steady decrease in dimension-reduction usage over the course of a trial, similar to experimental results (Figure 5.13b). We did the same mixed-effect linear regression analysis that was done on the experimental data (which includes all collapsed fixations in positions between 20 and 80). The main and interaction effects shown in Table 5.5 indicate that the decrease in dimension reduction in word trials is significant. The reader can also refer to Appendix E for an additional analysis based on Autoregressive Integrated Moving Average models applied to data on Figure 5.13b.

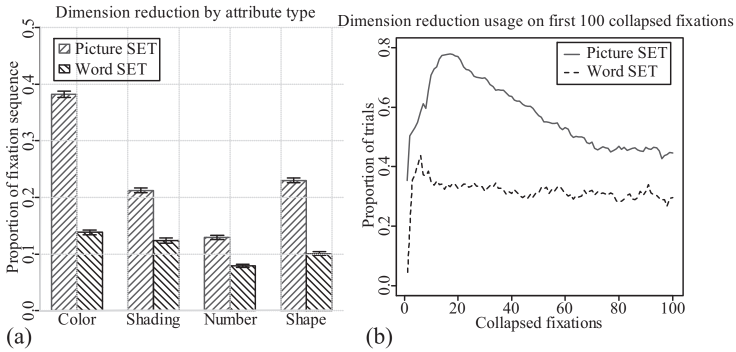


Figure 5.13: (a) The usage of attribute types in similarity-based scanning as a proportion of the trial's collapsed fixations sequence. (b) The changing proportion of trials in which dimension-reduction was used. The proportions are calculated as a function of the collapsed fixation position *x* within a trial. The proportion on position *x* is calculated by counting the trials that have a dimension-reduction block that include fixation at *x*. The lengths of blocks from word trials are also normalized to match the length scale of picture trials.

Table 5.5: The results of linear mixed-effect regression analysis of a predicted proportion of dimension reduction based on a collapsed fixation position and a trial type.

	Estimate	**Std. Error**	***t* value**	***p* value**
Intercept (Picture trial)	0.8399	0.0096	87.49	< 0.001
Fixation position	-0.0053	0.0001	-55.42	< 0.001
Word trial	-0.4898	0.0071	-68.83	< 0.001
Fixation position and word trial interaction	0.0039	0.0001	28.70	< 0.001

As Figure 5.13a shows, there is no clear preference toward a specific attribute in word trials. This is because the acuity difference among attribute types is gone. However, we know that the word set model uses dimension-reduction in a similar manner as the other model. The obvious question is why there is no clear indication of its usage in Figure 5.13b. The answer likely lies in the different scanpaths that the model for word set produces. The paths with dimension-reduction are revealed by identifying subsequences of continuous fixations on cards that share a common value with a highlighted card. It is quite easy to identify such subsequences in scanpaths produced from picture trials, since the model does targeted searches supported by the content of iconic memory. On the other hand, the word set model does exhaustive searches by attending every card. This makes it hard to identify subsequences of continuous fixations on similar cards. This results in the rather uninformative near flat line shown in Figure 5.13b. The fact that the true proportion of dimension-reduction in word set is as high as 40% becomes apparent in a verification phase. During this verification phase, both subjects and model make consecutive fixations on the same set of cards to verify whether a valid pair was made or the valid set was found. Examples of fixations belonging to the verification phase can be seen in Figure 5.9, in which such fixations are marked by red and green blocks within the fixation sequence diagram.

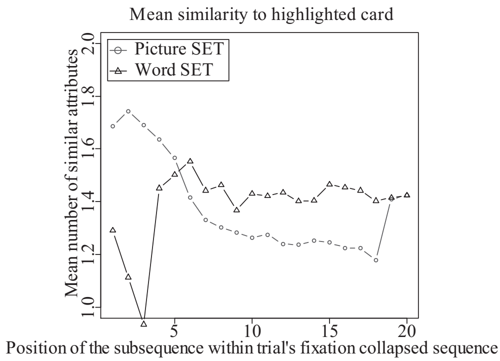


Figure 5.14: The mean overall similarity of all cards in a particular subsequence to the highlighted card. The values are calculated separately for picture and word trials.

Table 5.6: The result of a linear mixed-effect regression analysis of a predicted similarity to a highlighted card based on a subsequence's position and a trial type.

	Estimate	**Std. Error**	***t* value**	***p* value**
Intercept (Picture trial)	1.585	0.005	324.6	< 0.001
Subsequence position	-0.013	0.000	-20.2	< 0.001
Word trial	-0.197	0.005	-36.6	< 0.001
Subsequence position and word trial interaction	0.014	0.000	18.2	< 0.001

The picture set model shows a clear gradual shift from similarity to dissimilarity-based search. This gradual shift shown in Figure 5.14 resembles quite closely the one shown in Figure 5.7 obtained from experimental data. The picture is different for the word set model. The mean similarity to a highlighted card stays on more or less the same level. Results from mixed-effect regression analysis are shown on Table 5.6. The negative main effect of Subsequence shows that similarity to a highlighted card in a picture set trial decreases with each consecutive subsequence. However, the positive interaction effect of Subsequence negates the negative main effect of Subsequence in a word set trial. Combined interpretation of the main and interaction effects suggest that there is no decrease in similarity to a highlighted card in word trials. The slight downward bump between the 4th and 9th subsequences is the only visible clue that there is a preference toward similarity at the beginning. This lack of an obvious effect is explained by the same need for an exhaustive search that makes it hard to distinguish dimension-reduction scanpaths from scanpaths where dissimilarity-based search is used.

Systematic versus unsystematic scanpaths

Similar to human subjects, the two models also show a difference in scanpaths in terms of saccade directions.

The density plot in Figure 5.15a clearly shows these differences. Like human subjects, the model for word set shows a higher preference for vertical and horizontal saccades. However, the distributions are narrower and have higher peaks. This is to be expected, since the model is much more precise than human eye movement data.

The radar chart shown in Figure 5.15b fails to show considerable differences between the two models that is present in Figure 5.15a. It most likely due to the combined effect of discrete categorization and averaging that is required for calculating proportions for each saccade category. However, a logistic mixed-effect regression analysis applied to the model data, shown in Table 5.7, reveals similar main and interaction effects of trial type and trial level as found in subject data. Probabilities of diagonal saccades in picture and word trials are 0.45 (the intercept) and 0.39 (the main effect of word trial), respectively. Trial level again has a positive effect on the probability of a diagonal saccade in picture trials (the main effect of Trial level). However, such an effect is absent in word trials, as shown by the interaction effect on Table 5.7.

Table 5.7: The results of logistic mixed-effect regression in which the predicted value is the probability of a diagonal saccade.

	Estimate	Std. Error	*z* value	*p* value
Intercept (Picture trial)	-0.197	0.012	-16.22	< 0.001
Word trial	-0.229	0.013	-17.04	< 0.001
Trial level	0.037	0.004	9.49	< 0.001
Word trial and trial level interaction	-0.041	0.004	-9.33	< 0.001

This decrease in diagonal saccades is mostly the result of an absence of iconic memory content, which would direct attention shifts straight to the cards relevant to the search. Instead, the word set model resorts to shifting attention to the closest card. In most cases, the closest card is a card that is in either in the same column or in the same row.

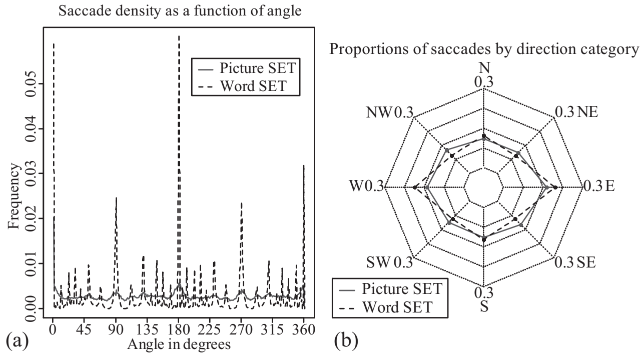


Figure 5.15: A radar chart for proportions of saccades in each saccade category.

Figure 5.15 shows that both models have a tendency toward horizontal saccades with East and West having the highest proportions. Another interesting aspect is the fact that both models produce a higher proportion of diagonal saccades than subjects. This should be considered together with the fact that neither of the models do unsystematic searches. At any time, the models' attention shifts are always guided by some top-down goal. This result indicates that an estimation of diagonal/non-diagonal saccades is not the reflection of the systematicity of the search, but rather an indication of the structure and presentation style of the scene. In other words, subjects may not have explicit preference toward diagonal or non-diagonal saccades during a systematic search. Proportions of two saccade types are likely to be defined by a function with two variables: 1) how structured the scene is (e.g. visual stimuli are organized in grids or positioned randomly) and acuity properties of the scene (high versus low acuity stimuli). The fact that model shows considerable amount of diagonal saccades while doing a systematic search (Figure 5.15b) supports the dependence of diagonal/non-diagonal saccades on the scene structure, while dependence on acuity properties is supported by subject data (Figure 5.8b) where there is a difference in saccade proportions between picture and word sets.

Discussion and Conclusion

In the previous section, we have described two nearly identical models. Both models use exactly the same strategy and the same set of values for adjustable parameters. However, the two models produce behavioral data that on the surface look very different. The entire difference in behavior can be explained by a simple change in presentation style of the task. Furthermore, both models show a good fit to the experimental data, suggesting that a similar change in presentation style affected human subjects in much the same way: the behavioral data may change significantly without changes in overall strategy.

The contrast between the picture and word versions of SET shows that the style of presentation alone can have a drastic effect on performance in a problem-solving task. With no changes in isomorphic structure, a simple replacement of an iconic representation with a textual representation resulted in more than a twofold increase in reaction times. However, as experimental results and model simulations show, the overall strategy, the way the problem-solving task is approached, did not change. Our original model for picture set was adapted to play word set with the minimum changes necessary to compensate for the absence of the perceptual components of the game. Yet, the model for word set was able to closely replicate subjects' behavior with respect to reaction times and eye movements. Furthermore, the model provides a perspective from a level of individual cognitive processes. We explored how task performance may be affected by visual acuity, differences in visual encoding and dynamics of iconic memory. Exploring how these processes change based on the nature of a task, helps us to understand how subjects manifest different behaviors in two versions of SET, while still following the same strategy.

It is interesting how a simple change in visual presentation style can result in what can be called a cascading domino effect in cognitive processes. Change in presentation style triggers change in a cognitive process that itself triggers change in one or more other processes. The changes propagate like a chain reaction. In SET, replacing a high acuity stimulus with a low acuity stimulus removed the advantage of peripheral vision. This lack of peripheral vision resulted in a lack of content in iconic memory and imposed changes on how a visual stimulus, such as a card, was encoded. Changes in iconic memory and encoding further affected the prioritization of attention shifts that manifested itself in different scanpaths. All those changes added up, resulting in increased reaction times and a different pattern of fixation sequences. Furthermore, the data initially appear to give an overall false impression that there are fundamental changes in the strategy subjects use to find a set. However, results of this study do in fact show that changes in presentation style do not necessarily trigger changes in how a subject approaches a problem-solving task. Instead, there are more subtle changes on the level of cognitive processes. The strategy remains the same, but the cognitive processes that are used to implement the overall strategy can change. Such a change can be either beneficial or damaging to performance. For example, in picture trials, peripheral vision is extremely useful in locating cards relevant to the search. In word trials, peripheral vision does not provide any leverage, given that the only option is that of deliberate top-down scanning. This transition from faster low-level processes to more top-down cognition has a rather significant negative effect on reaction times.

Jacob and Hochstein (2008), who originally proposed dimension-reduction, assumed that the bias toward similarity in SET is a result of the highly perceptual nature of the game. They argued that players prefer to search for lower level sets, because it is easier to identify similar cards using bottom-up visual processes. Our experiment with word set showed that this is not the case. Even in absence of bottom-up encouragement, subjects needed less time to find lower level sets, indicating that bias toward similarity still exists. This bias is definitely part of a deliberate strategy, rather than an artifact of mechanisms based on perceptual similarity. However, we are yet to identify what exactly causes players to look for similar cards first, rather than for dissimilar cards.

Exploring beyond SET

It is completely possible that the changes in underlying cognitive processes are responsible for better performances in the Marble Drop game. Meijering at al. (2010) also acknowledge the importance of context, although from a perspective of higher-order reasoning. The advantage of Marble Drop is that it provides a bottom-up visual context using colors, trapdoors and bins of decreasing heights. This context is more intuitive and easier to process using bottom-up cognition. One obvious example is the clear advantage peripheral vision provides in Marble Drop. It is much easier to detect difference in color and color-grades using peripheral vision, than to deliberately compare numeric values. It is the possibility to use visual processes that are bottom-up, pre-attentive and parallel that makes Marble Drop an easier game.

Smarter than expected?

One can argue that bottom-up processes should be able to extract at least some semantic information in order to provide a necessary performance boost in a problem-solving task.

For example, in Marble Drop, it is easier to visually differentiate and compare payoffs due to a distinct color-grade associated with each payoff. However, this also implies the presence of some form of a semantic association between darker color and a higher payoff at the pre-attentive level. If there is no such association, deliberate comparison will still be necessary. However, there is a mounting amount of research suggesting that pre-attentive visual processes are not as dumb as they were considered to be before (Proverbio, Zotto & Zani, 2007). It is often ignored how much information is processed subconsciously. For example, richer information in iconic memory that is processed subconsciously seem to give participants a significant performance boost in picture set over word set. It is further likely that even more complicated processing at the semantic level is done by our visual system. Rensink (2007) proposed an architecture where a certain amount of semantic information is processed pre-attentively by the human vision system. Perhaps it is exactly that kind of visual information that is readily available in picture trials that makes the original version of the game so much easier than the word version of the game. In word trials, the semantic information that otherwise would have been extracted more efficiently by visual bottom-up processes needs to be processed by deliberate top-down reasoning.

Exploring through models

Models are useful tools for exploring differences that are otherwise difficult to reveal by means of statistical analysis. Computer modeling is the only objective way currently available to explore the behavior of a complex modular system in which changes in one module can propagate throughout the entire system. The human cognitive system is definitely a good representative of such. For example, it is hard to statistically calculate the outcomes of the domino effect described in the previous subsection. Instead, we used a computational model based on a cognitive architecture to directly simulate these outcomes. The model for word set worked quite well, especially considering the fact that it was directly adapted from the existing model of picture set with minimal changes to suit the new presentation style.

Data and source code

All of the data related to this study, including the model source code and the experiment data can be downloaded via the following link: http://www.ai.rug.nl/~n_egii/models/. The source code for the PAAV module can be downloaded at http://www.ai.rug.nl/~n_egii/models/codes/paav-module-no-vstm.lisp

.

CHAPTER 6

Human Reasoning Module

This chapter will be published as:

Nyamsuren, E. & Taatgen, N. A. (accepted).
Human Reasoning Module.
Biologically Inspired Cognitive Architectures.

Abstract

This paper introduces a framework of human reasoning and its ACT-R based implementation called the Human Reasoning Module (HRM). Inspired by the human mind, the framework seeks to explain how a single system can exhibit different forms of reasoning ranging from deduction to induction, from deterministic to probabilistic inference, from rules to mental-models. The HRM attempts to unify previously mentioned forms of reasoning into a single coherent system rather than treating them as loosely connected separate subsystems. The validity of the HRM is tested with cognitive models of three tasks involving simple casual deduction, reasoning on spatial relations and Bayesian-like inference of cause/effect. The first model explains why people use an inductive, probabilistic reasoning process even when using ostensibly deductive arguments such as modus ponens and modus tollens. The second model argues that visual bottom-up processes can do fast and efficient semantic processing. Based on this argument, the model explains why people perform worse in a spatial relation problem with ambiguous solutions than in a problem with a single solution. The third model demonstrates that statistics of Bayesian-like reasoning can be reproduced using a combination of a rule-based reasoning and probabilistic declarative retrievals. All three models were validated successfully against human data. The HRM demonstrates that a single system can express different facets of reasoning exhibited by the human mind. As a part of a cognitive architecture, the HRM is promising to be a useful and accessible tool for exploring deeps of human mind and modeling biologically inspired agents.

Introduction

In this paper, we introduce a framework that attempts to unify various approaches to human reasoning. The Human Reasoning Module, or HRM, is an implementation of this framework developed as a part of the ACT-R cognitive architecture (Anderson, 2007). As opposed to ACT-R's core modules that represent specific types of cognitive resources such as vision or memory, the HRM does not add a new type of cognitive resource. The HRM extends the theoretical frameworks and corresponding computational functionalities of the existing modules of ACT-R. Therefore, the HRM is both a theory and a tool for modeling. As a theory, it advocates for a specific structure of knowledge organization in our declarative memory. The structure is still based on knowledge chunks, but adds specific requirements on chunk types and its slots. Furthermore, the HRM advocates the existence of task-general procedural knowledge that gives us the ability to reason and solve problems based on real-time information and previous experience. The proposed structures of declarative and procedural knowledge define grammar, axiom schemata and inference rules of human logic. As a tool, the HRM both extends and constrains the functionality of ACT-R's declarative module and also adds a set of task-general production rules to ACT-R's procedural module. Ideally, if the HRM is a valid model of human reasoning it should be able to tackle any form of reasoning process. However, the HRM's current unification attempt is limited to two dimensions depicted in Figure 6.1. The next subsection discusses in details these dimensions.

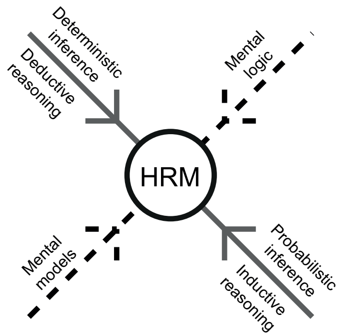


Figure 6.1: Two dimensions of human reasoning that the HRM attempts to unify.

Inductive and deductive reasoning

At the core of the HRM, there is an assumption that the human general reasoning skill is inherently probabilistic or inductive. Any true form of classical deductive reasoning requires a *closed world assumption* stating that what is not currently known to be true is false. This is an extremely unpractical assumption in the real world full of uncertainties (Rajasekar, Lobo, Minker, 1989), and we subconsciously or consciously recognize this fact. Cummins (1995) demonstrated that even when someone is reasoning with ostensibly deductive arguments one still uses an inductive, probabilistic reasoning process. Further uncertainty arises due to limitations of our cognitive resources: our perception of the world can be noisy or limited and our memory may be forgetful. With such uncertainties, any deductive system will fail the tests of *validity* and *soundness*, necessary requirements for any formal deductive inference (Jeffrey, 1981). Furthermore, we do not often try to satisfy both of these requirements in our reasoning process (Thompson, 1996). Therefore, the HRM operates under the *open world assumption*, what is not proven is not necessarily false, and tries to prove truthfulness rather than falsity of knowledge.

However, the HRM does not exclude a possibility that deductive reasoning occurs within the context of specific tasks. Let us assume a specific problem that eliminates environmental uncertainties by clearly and unambiguously specifying contextual boundaries, constraints and rules. We can further assume that the problem is tractable within capacities and limitations of our cognitive resources, and there is no interference to the solution from our past knowledge outside of the problem's context. Such context will follow the closed world assumption, and, hence, deductive reasoning may be used. Therefore, in the HRM, there are no two separate processes for deductive or inductive reasoning. Instead, the HRM assumes that deductive reasoning is an instance of inductive reasoning over a specific domain of discourse with a near-zero uncertainty. A degree of uncertainty is the common dimension that implicitly unifies inductive and deductive reasoning in the HRM.

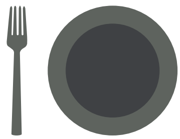
Mental logic, mental models and bottom-up reasoning

Next, the HRM further argues that general human reasoning does not necessarily rely on formal propositional forms and is not strictly top-down (conscious). There is a long history of debate over the theories of mental models and mental logic. The mental logic theory argues that a set of inference rules is applied to logical forms abstracted from stimuli (Rips, 1983). A commonly agreed interpretation of mental models theory dictates that stimuli are abstracted into a form of mental diagram where configuration information reflects the relationship between entities (Banks & Millward, 2009; Johnson-Laird, 1983). In the HRM, the two theories are part of the same reasoning process. It is based on the assumption that these two are not mutually exclusive strategies. Roberts (1993) rightfully pointed to the fact that there are no obvious reasons why the two types of theories should be incompatible. Coney (1988) argued for individual differences based on a study showing that some people are better at spatial reasoning while others prefer reasoning based on formal propositions. Johnson-Laird (2004), a chief proponent of the mental models theory, admitted that the model theory does not imply that reasoners never rely on rules of inference.

The HRM consolidates the two theories by assuming that a mental model is a form of working memory that allows convenient representation and storage of knowledge required for reasoning. New premises, including ones not explicitly stated by the problem context, are assumed to be extracted on demand from the mental model during a rule-based inference similar to the mental logic. The mental model as a working memory simplifies a manipulation and retrieval of knowledge that otherwise has to be stored in a less efficient long-term memory. For example, items in the existing model can be easily reconfigured to produce an alternative model. The smaller amount of cognitive effort required by the mental model can explain why people prefer it over direct inference on given propositional forms. This interpretation of the mental model implies that it is not the main tool of reasoning by itself. This is a major distinction from Johnson-Laird's (2004) interpretation arguing that the probability of a conclusion is estimated based on the proportion of equipossible models in which it holds. Certainly, our interpretation of the mental model is more parsimonious.

At this point, we need to map a metal model onto a specific cognitive resource. Johnson-Laird (2004) provided three functional requirements for the mental model: 1. A mental model should have an imagery capability to abstract meaning of premises into a mental diagram; 2. A mental model should be iconic; 3. Mental models should represent what is true, but not what is false. The cognitive resource that matches all above requirements is visual short-term memory (VSTM). It is specialized visuo-spatial mechanism in working memory for storing visual iconic information for a short duration (Logie, Zucco, Baddeley, 1990). VSTM stores a factual representation of the current state of affairs and, therefore, implies that information in it is assumed true. Arguably, one of the most important roles of VSTM is to retain and combine information gathered across successive fixations to construct dynamically a high-level internal representation of the outside world (Henderson & Hollingworth, 2003; Rensink 2000a, 2000b). The same process of retaining and combining information is likely to be necessary for building a mental model. Furthermore, VSTM is likely to have at least some imagery capability (Phillips, 1983; Wintermute, 2012). Phillips (1983), one of the first to introduce the concept of VSTM, emphasized that VSTM facilitates our ability to visualize problem space and is not just a sensory store. Jiang, Olson, and Chun (2000) reported that spatial information stored in VSTM includes not only object's location but also its relationship to other objects in VSTM. Based on these studies, we can conclude that VSTM is a suitable candidate for storing a mental model.

Now, we should discuss whether VSTM is distinct from long-term declarative memory. Unlike declarative memory, VSTM needs to provide a fast and reliable access to information to allow the scene representation to be constructed dynamically across rapid fixations. Thus, VSTM is functionally different from declarative memory. Furthermore, Phillips (1983) made a clear distinction between VSTM and long-term visual memory noting that head injuries affecting long-term memory do not affect visualization. Baddeley (2003) argued for distinction between long-term memory and the multi-component working memory that includes visuospatial sketchpad with imagery capability, a functional analogue to VSTM. Neuroimaging studies suggest that the short-term memory responsible for storing visuo-spatial information is located in parietal lobe (Baddeley, 2003; Lum, Conti-Ramsden, Page, & Ullman, 2012; Xu & Chun, 2005) and not in the hippocampus that is commonly associated with declarative memory. Finally, Formisano, Linden, Di Salle, Trojano, Esposito, Sack, Grossi, Zanella, & Goebel (2002) showed that parietal lobe also performs distinct functions of mental imagery. All these factors together support our assumption that VSTM is a distinct memory suitable for building a mental model.



(r-left-of fork plate)

Figure 6.2: The image on the left contains an implicit knowledge that the fork is on the left side of the plate. Such knowledge can be extracted to form explicit proposition on the right.

The HRM treats the content of VSTM as a mental model unless it is irrelevant to the task. When available, the HRM extracts premises necessary for inference from the iconic content of VSTM. This process assumes that implicit semantic information is converted into explicit information. As an example, imagine that the VSTM contains visual objects as shown on left side of Figure 6.2. Each object has set of features describing it such as color, shape, spatial position, etc. There is also relative spatial information, such as, the fork being on the left side of the plate. This information was not encoded as part of any object. However, it implicitly exists inside VSTM even though we may not be consciously aware of it until it is parsed. The relative spatial position can be quickly extracted on demand and converted into explicit propositional form shown on the right side of Figure 6.2. Rensink (2007) indicated bottom-up visual processes may be able to process information at a semantic level subconsciously and even pre-attentively. It is feasible to assume that the same bottom-up processes are responsible for extracting explicit knowledge from implicit knowledge. Within the HRM, we refer to such process as *visual bottom-up reasoning* mechanism (not to be confused with inductive reasoning). We will further explore the mental logic and the mental model using an example task and a cognitive model based on the HRM.

Deterministic and probabilistic inferences

In the previous section, we have mentioned that the HRM uses rule-based inference that is inherently deterministic. This determinism relies on the assumption that the knowledge source is consistent and reliable. We also discussed that visual short-term memory is a source of knowledge for reasoning. As a form of working memory, VSTM provides a reliable access to reasonably consistent knowledge and does not violate above-mentioned assumption. Therefore, when the reasoning process relies on VSTM only it can be deterministic and deductive.

However, there is a second source of knowledge, a long-term declarative memory. The HRM uses ACT-R's declarative memory (DM). As a proper model of human long-term memory, DM has inherited its peculiarities as well. DM can contain inconsistent and often competing knowledge. Knowledge chunk retrieval is governed by probabilities based on activation values. As a result, retrieved knowledge may not match completely what is requested, or retrieval may even fail. It has been already suggested that DM plays a central role in casual reasoning (Drewitz & Brandenburg, 2012). The uncertainty over retrieved knowledge from DM transforms the HRM's rule-based inference into probabilistic inference. Based on example models, this paper describes how the HRM is used to simulate casual deduction, pragmatic reasoning and even inductive Bayesian inference.

Finally, little is known about the form of cognitive processes that provide meta-control over reasoning strategies. For example, how do we decide whether to use as a source of knowledge the mental model in a form of visual short-term memory or declarative memory? Not every problem context can be converted into an iconic form, and in such cases, there is no other choice but to use knowledge in declarative memory. However, what if both VSTM and DM contain relevant or even conflicting knowledge? The HRM introduces a simple, but effective cognitive construct referred to as a *reasoning pipeline* that addresses these issues. A reasoning pipeline assumes a sequential process where alternative strategies are used one by one in increasing order of cognitive effort required until a conclusion is reached. For example, access to VSTM requires less time than a declarative retrieval. Thus, the HRM prefers reasoning based on VSTM knowledge to reasoning on declarative knowledge.

Architecture of the HRM

Knowledge representation in declarative memory

Chunk types and chunks, instances of chunk types, represent factual knowledge in ACT-R. A chunk type defines a set of slots its instance chunks can inherit. Those slots can contain values describing chunk's properties. Those values can be either other chunks or atomic values such as strings of characters or numeric values. ACT-R provides no restrictions on chunk types and chunks that can be defined by a modeler. The HRM restricts a modeler to a predefined set of chunk types thereby encouraging a commitment to a common knowledge structure that is not model specific. The core set of chunk types in the HRM are ones describing concepts, triples and inference rules.

Concepts and triples

The atomic unit of knowledge in the HRM is a *concept*. Any unit of knowledge that has distinct semantic meaning can be a concept. There are two types of concepts in the HRM: property instance and class instance. Property instance is any concept that is used to relate two other concepts semantically. As such, the knowledge organization inside the HRM revolves around a predicate construct referred to as a *triple*: (*property subject object*). Inside a triple, *property* establishes a semantic connection between *subject* and *object*. The following is an example of a triple: (*r-left-of fork plate*). In the HRM, *r-left-of* is a property instance that is used to represent a spatial relation between two class concepts. In example above, the meaning of the triple is equivalent to "*a fork is on the left side of a plate*".

A property instance can also be used as triple's subject or object. For example, the HRM has two different property instances, *r-left-of* and *r-dir-left-of*, for expressing a similar spatial relation between two class instances. *r-dir-left-of* expresses semantically more restrictive spatial relation implying that subject is to the left of an object, and both subject and object are aligned vertically. Therefore, triple (*r-dir-left-of fork plate*) entails triple (*r-left-of fork plate*). One way to express such one-way relation is to have another triple (*entails r-dir-left-of r-left-of*). Here, property instance *entails* semantically connects two other property instances instead of class instances. Otherwise, *entails* is no more special from other property instances such as *r-left-of* or *r-dir-left-of*.

Most of the studies of human mental logic advocate for some form of predicate construct as a way of knowledge organization. We have chosen the triple form because it closely resembles a linguistic predicate typology consisting of subject, verb and object. It is the most common sentence structure found across different languages. Such commonality strongly indicates that underlying knowledge from which a sentence is constructed may also be organized in the same form consisting of subject, object and verb (Crystal, 1997).

The HRM has a limited notion of time. A triple can be assigned a specific timestamp. For example, the sentence "*John ate sandwiches yesterday and today*" can be expressed with two triples with the same structure but different timestamps:

(*eat John sandwich* (*ts "yesterday"*))
(*eat John sandwich* (*ts "today"*))

A special slot named *ts* is used to assign a timestamp. When necessary, the above two triples can be differentiated by timestamps, otherwise they are semantically similar. In current implementation of the HRM any value can be used as a timestamp. This implementation required the least amount of effort, but it is not a realistic representation of human temporal cognition. Ideally, there should be restrictions on what kind of values can be used to represent time. On the one hand, it can be an explicit class instance to represent our high level understanding of time and data. On the other hand, timestamp value can be more implicit estimations of time intervals done by our internal biological clock. ACT-R already provides a temporal module (Taatgen, Van Rijn & Anderson, 2007) that provides such time interval estimations. Future updates of the HRM should include more restrictions on time values as well as integration with the temporal module.

Statements

In the HRM, *statement* is a type of triple that represents factual knowledge. It is a statement of a fact that is true or was true. The example triples from the preceding subsection are all valid *statements*. The HRM provides several ways to create a *statement*. Firstly, a modeler can explicitly define custom *statements*, as model's background knowledge. Secondly, the model itself can create *statements* in real-time via production rule calls to a special *reasoner* buffer. This option simulates the ability to obtain new explicit knowledge through external input, such as stimuli from the outside world. Finally, a model can generate a new statement by inferring it from existing statements using top-down reasoning, or by deriving it from an implicit connection between concepts using bottom-up reasoning.

Implicit and explicit knowledge

The HRM makes a distinction between explicit and implicit knowledge. *Statements* are explicit knowledge, a form of a knowledge that is known consciously. Implicit knowledge is knowledge that is represented by slot values of concept chunks. Such knowledge is implicit because it is assumed that ACT-R is not consciously aware of its presence, but subconsciously can extract it to form explicit *statements* using bottom-up processes. Following the previous example, there may not be any *statement* such as (*r-left-of fork plate*). However, concepts chunks for *fork* and *plate* may have slot values with *x* and *y* coordinates implicitly indicating relative spatial positions of two concepts. Those values then can be converted into explicit concepts such as *r-left-of* when necessary.

Inference rules

In the HRM, rules describe how a new *statement* can be inferred from existing *statements*. The HRM assumes that rules reflect our past experience and are formed as a result of our observations of relations among real-world entities such as cause/effect, pre-condition/action, action/post-condition observations, etc. Rules use special triples called rule-statements. Semantically, a rule-statement is not a fact, but either a condition or an implication of a possibility. Any rule consists of left- and right-hand sides. A left-hand side must have one or more rule-statements (antecedent), and the right hand-side should have exactly one rule-statement (consequent). In order for a consequent to be true, all antecedent rule-statements should also be true. For example, the rule below states "*if the fork is on the left of the plate then the plate is on the right of the fork*":

(*r-left-of fork plate*) ==> (*r-right-of plate fork*)

Unlike ordinary *statements*, rule-statements can use variables as one of the entities in the triple. The previous example rule can be rewritten as:

(*r-left-of "@item" plate*) ==> (*r-right-of plate "@item"*)

Above rule states "*if any item is on the left of the plate then plate is on the right of that item*". In this rule, "*@item*" is a variable, not a *concept*. The HRM recognizes as a variable any string value that starts with "@". It can be replaced by any valid concept that is factually on the left side of the plate. Variables provide a possibility to generalize rules beyond a scope of a particular concept or even an entire model. It also introduces a possibility to reuse the same rules across different ACT-R models, at least partially, addressing one of the major reusability challenges in ACT-R.

Assertion

Assertion is another type of triple used by the HRM. Assertion represents a query questioning the HRM whether a triple is true. For example, the assertion (*r-right-of plate fork*) represents the query: "*Is the plate on the right side of the fork?*" Similar to rule-statements, assertions can have variables. The assertion (*r-right-of plate "@item"*) asks the HRM to find any class instance that is on the right side of the *plate*. In ACT-R, the HRM can be queried with an assertion via *reasoner* buffer. Upon receiving an assertion, the HRM starts a reasoning process called a *backward reasoning pipeline*. The task of reasoning pipeline is to check if assertion can be proven to be true or to find/prove any *statement* that matches the assertion if assertion contains variables. If assertion is true then it is converted into a *statement* and placed inside *reasoner* buffer. If a matching *statement* is found then that *statement* is put inside *reasoner* buffer.

Schema and inference types

Conditional proof schema

The HRM uses the same conditional proof schema defined by Braine & O'Brien (1991): to derive or evaluate *if p then q*, first suppose *p*; when *q* follows from the supposition of *p* together with other information assumed, one may assert *if p then q*. This schema together with the open world assumption has several implications that make the HRM's inference different from an inference based on material conditionals of a classical logic:

1. The HRM does not follow the closed world assumption unless it is explicitly required. Therefore, what the HRM cannot prove is not necessarily false.
2. There can be two or more competing or conflicting inference rules that can be true at different instances: e.g. *if p then q*; *if p then k*. For example, the agent may build following two inference rules through observations of rolling dice: *If throw dice then get 6*; *If throw dice then get 3*.
3. The *sufficiency* requirement will not necessarily hold: the antecedent *p* is not necessarily a sufficient condition for a consequent *q* because other information may be assumed to assert *if p then q*. Consider following common sense rule: *If brakes are pressed then car stops*. Most of the times, the rule is true. However, there it is assumed that, for example, the brakes are not broken.
4. The *validity* requirement of deductive reasoning will not necessarily hold: the conclusion may not be true even if the premises are true. For example, the HRM may fail to assert *if p then k* because it already asserted *if p then q*. Consider the dice example from the implication 2. If a dice is thrown then the HRM may assume that result is 6. The second possible conclusion of 3 remains untrue even though its premise

of dice being thrown is true. Furthermore, the validity requirement cannot hold if the sufficiency requirement is not met.

5. The law of *contrapositive*, or *Modus Tollens (if p then q, therefore if ¬q then ¬p)*, also does not necessarily hold. Consider the contrapositive of the example from the implication 3: *If a car hasn't stopped then the brakes were not pressed*. Because of violation of the sufficiency requirement, the contrapositive argument may not be true: *The brakes were pressed, but the car hasn't stopped because the brakes were broken*. In this case, the assumed information that the brakes are not broken is not true. Therefore, the HRM does not automatically generate contrapositives from inference rules. The HRM assumes that a contrapositive should be observed and memorized as an inference rule of its own right.

The law of *syllogism* (*if p then q, if q then k, therefore if p then k*) is at the center of the HRM's capability for complex reasoning. Consider following example: *If the sun sets then a night comes. If a night comes then a temperature drops. Therefore, if the sun sets then a temperature drops*. There is no explicit relation between the sun setting and the temperature dropping in two rules. However, it can be inferred using of law of syllogism. The ability to chain the inference rules together allows the HRM to explore different reasoning strategies with the same inference process.

Reasoning types

The inference rules can be used for two types of reasoning in the HRM: backward and forward. Backward reasoning is used to determine whether a specific conclusion can be reached. Forward reasoning is used to determine what kind of conclusion can be reached given set of evidences. Backward reasoning retrieves an inference rule by matching its consequent, while forward reasoning retrieves the inference rule by matching its antecedent. For further explanation, let us assume that there is the following *Rule 1*:

Rule 1:	Interpretation:
(have-state brake pressed)	*If a brake is pressed, and it is not broken*
(NOT-have-state brake broken)	*then car speed decreases.*
==>	
(decrease car speed)	

With *Rule 1*, the HRM can answer two types of questions. The first question is "*Is car speed decreasing?*". It is a question answerable by backward reasoning. The HRM's equivalent of this question will be an assertion *(decrease car speed)* sent to a *reasoner* buffer with an expected conclusion that it is true or not true. The assertion will be true if there is a rule that (1) has a consequent matching the assertion and (2) has an antecedent where all rule-statements are true or inferred to be true via the law of syllogism. In this case, the HRM will use the *Rule 1* because its consequent matches the assertion. However, to infer that the assertion is true the HRM will also have to infer that *Rule 1*'s antecedent is also true. We will discuss later various strategies used for such inference.

The second question is "*What happens if the brake is pressed, and it is not broken?*". It is a question answerable by forward reasoning. The HRM's equivalent of this question will be supplying two facts, *(have-state brake pressed) and (NOT-have-state brake broken),* to the *reasoner* buffer and expecting some or no conclusion. The conclusion will be reached if there is a rule that (1) has an antecedent matching the given facts in the *reasoner* buffer and (2) has an antecedent where all rule-statements are true or inferred to be true due the law of syllogism. The facts in the *reasoner* buffer can be used to assert truth-values of the antecedent. In this case, the HRM concludes that the car speed should decrease (*decrease car speed*) because of the *Rule 1*. It is possible to ask another question such as "*What happens if a brake is pressed?*". The HRM's equivalent of this question will be supplying only single fact *(have-state brake pressed)* to a *reasoner* buffer. However, according to the *Rule 1*, the second fact, *(NOT-have-state brake broken),* is required to reach a conclusion. In such case, the HRM will try to prove the second fact using backward reasoning.

A *reasoning pipeline* provides a meta-cognitive control over reasoning processes. The HRM uses two reasoning pipelines for backward and forward reasoning respectively. In ACT-R, reasoning pipelines are implemented as a series of automated calls to production rules built into the HRM. These production rules are task-general reasoning rules and are part of the cognitive architecture. This approach differs from traditional ACT-R modeling practices that treat all production rules as part of a model. On the other hand, the declarative inference rules are often treated (but not necessarily always) as being task-specific. The inference rules together with statements of facts provide a problem context within which the task-general production rules can reason and derive conclusions.

Following the threaded cognition theory (Salvucci & Taatgen, 2008, 2011), reasoning pipelines are contained within the HRM's own cognitive thread that runs in parallel with other (model-specific) cognitive threads. This means that model-specific production rules irrelevant to reasoning pipelines can fire in-between production rules belonging to the HRM. It opens the possibility that declarative retrievals requested by other threads can interfere with the HRM's reasoning that relies heavily on declarative memory. Such interference is possible despite the fact that ACT-R locks access to declarative memory during individual retrieval instances (it is not possible to recognize a thread that initiated retrieval). Therefore, the HRM uses a stricter control that locks declarative memory through entirety of the reasoning pipeline.

Backward reasoning pipeline

As it was discussed earlier, new knowledge can be generated from existing knowledge using one of several different strategies. The backward reasoning pipeline establishes priority among those strategies and organizes them into series of consecutive steps. The highest priority strategy receives an assertion first and tries to prove it. If it fails then the assertion is passed to the next highest priority strategy. The HRM triggers calls to backward reasoning pipeline as soon as it receives an assertion request inside *reasoner* buffer. The backward reasoning pipeline recursively calls itself (the law of *syllogism*) until either the assertion is proven or it is decided that the assertion cannot be proven.

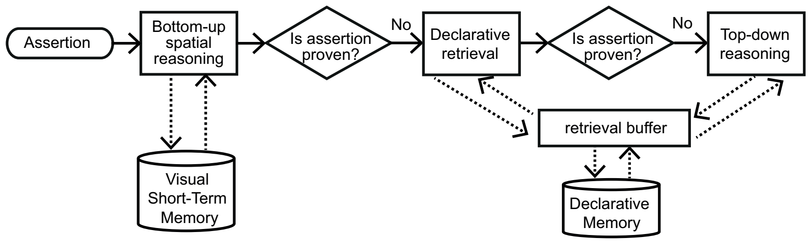


Figure 6.3: A simplified workflow of an HRM reasoning pipeline in ACT-R.

Currently, backward reasoning pipeline supports three different strategies: bottom-up reasoning, declarative retrieval and top-down reasoning. Figure 6.3 shows the prioritization of those strategies. Bottom-up reasoning is preferred requiring the least amount of cognitive effort. Bottom-up reasoning is followed by declarative retrieval and top-down reasoning in decreasing order of priority.

Bottom-up reasoning

The current implementation of the HRM's visual bottom-up reasoning supports only spatial reasoning. As with other forms of reasoning, spatial reasoning requires a source of knowledge based on which it can derive a new knowledge. In the HRM, such knowledge source is a visual short-term memory (VSTM). VSTM was introduced by the newer version of the Pre-Attentive and Attentive Vision module (Nyamsuren & Taatgen, 2013a), an extension to ACT-R's default vision module. VSTM is a high resolution, but low capacity visual memory. Every visual object encoded from the external world is temporarily stored inside VSTM until it decays out or is deleted due to capacity limitations. Unlike declarative memory, VSTM is considered as a visual analog of a working memory. Hence, objects inside VSTM can be accessed by the HRM with no cognitive cost, and explicit knowledge can be derived with little effort.

The HRM can take advantage of VSTM whenever it receives an assertion about spatial relation between two concepts such as (*r-right-of plate fork*). VSTM contains detailed information about each visual object currently in its store, including the object's original position in real world. In ACT-R, those are two-dimensional spatial coordinates. The HRM can use such implicit knowledge to quickly derive explicit *statements* about spatial relations between concepts inside VSTM. If one of those derived statements supports the assertion then the assertion is proven.

Declarative retrieval

If bottom-up reasoning fails then the HRM tries to retrieve from declarative memory any *statement* that can directly confirm the assertion. In ACT-R, a declarative retrieval can be a time-costly process. Furthermore, there is a chance that retrieval will fail even if a matching *statement* exists. Those are the reasons why bottom-up reasoning takes priority over declarative retrieval as a more reliable and faster process.

Top-down reasoning

Top-down reasoning is only invoked if declarative retrieval fails. It involves rule-based reasoning where a chain of inference rules is used to prove an assertion.

The current implementation of the HRM supports a fully functional backward-chaining algorithm implemented as a set of ACT-R production rules. The first production retrieves from declarative memory any consequent rule-statement that matches the assertion. If the retrieval of a rule's consequent is successful then the next production retrieves the first antecedent rule-statement of the same rule. The retrieved antecedent rule-statement is converted into an assertion and fed back to the HRM. This starts a new recursive call with a new reasoning pipeline. If recursive call was able to prove that current antecedent rule-statement is true then the next antecedent rule-statement is retrieved, converted into assertion and fed back to the HRM. This process continues until all antecedent rule-statements are proven. In such a case, the consequent rule-statement is also true, and, hence, the original assertion is true as well. If any of the antecedent rule-statements cannot be proven then the HRM stops the reasoning process and sets the *reasoner* buffer to an error state.

The top-down reasoning consists of a series of production calls coupled with frequent declarative retrievals. Not only it is a hugely time-consuming process, but also it is very costly in terms of cognitive resources. Since ACT-R allows only one production call at the time, it creates a bottleneck for other task-specific productions. Furthermore, declarative memory is locked through entirety of the time the HRM uses it to prove an assertion. Hence, other cognitive processes cannot access declarative memory. The overall high cost puts top-down reasoning in the lowest priority position.

Forward reasoning pipeline

The simplified workflow of a forward reasoning pipeline is shown in Figure 6.4.Given statements of facts as a query, the HRM retrieves from declarative memory any rule that has antecedent rule-statements matching the statements in the query. A rule selection is governed by several criteria. Firstly, a rule must have rule-statements matching all query statements. Secondly, the order of rule-statements must be the same as the order of corresponding query statements. Thirdly, irrelevant rules that may not lead to a desired conclusion can be ignored. One of the unique aspects of human reasoning is that we can do it with an intention of achieving a particular conclusion. It is also possible to do the same in the HRM. If a target concept is specified in a query then the HRM ignores all rules that do not mention that concept in its consequent rule-statements. All three criteria applied to rule retrieval are based on principles of memory retrieval during decision-making under uncertainty. It was suggested that memory chunks are evaluated during retrieval with respect to relevancy, availability and accessibility (Kahnemann, 2003) as well as cross compared with alternative retrieval candidates (Schooler & Hertwig, 2005).

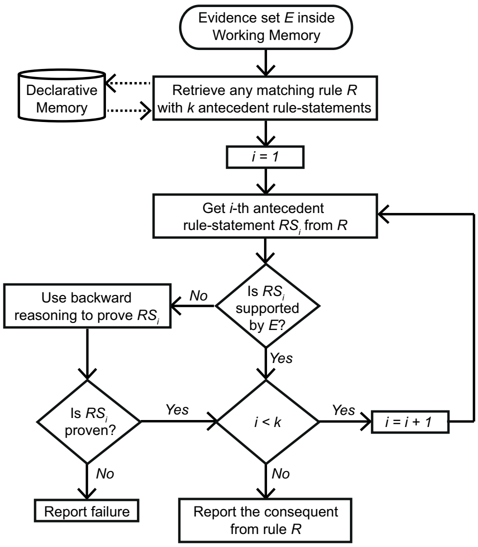


Figure 6.4: A simplified workflow of a forward reasoning pipeline.

If forward reasoning uses a rule that has antecedent rule-statements that were not specified in the query then those rule-statements are verified for truthfulness by invoking a backward reasoning. As such, forward reasoning is not purely forward and can have series of nested backward reasoning calls. This is different from traditional view where backward and forward reasoning are considered two distinct processes. The heterogeneous nature of the reasoning pipeline significantly increases a range of inference problems that the HRM can solve. The power of mixed forward and backward reasoning will be explored via example model of blicket categorization task (Griffiths, Sobel, Tenenbaum & Gopnik, 2011).

Validation Models

This section introduces three models of different experimental tasks. Each model is used to replicate human behavioral and validated against human performance data.

The model of a casual deduction task is used to demonstrate the HRM's basic reasoning abilities based on inference rules in declarative memory. It is the simplest of the validation models described in this study. It uses only declarative knowledge and does not require other modules such as vision. The reasoning strategy used by this model is limited to declarative retrieval of rules. The model demonstrates how competing and conflicting declarative knowledge can affect outcomes of even simple reasoning. It shows the importance of considering uncertainty in declarative retrieval results during any logical reasoning task.

The model of a spatial reasoning task demonstrates the full potential of the HRM's backward reasoning ability. It uses all three backward reasoning strategies: bottom-up reasoning, declarative retrievals and top-down reasoning with recursive calls to the backward reasoning pipeline. The reasoning in this model uses knowledge in declarative memory as well as in visual short-term memory.

This final model based on a blicket task is a demonstration of the HRM's ability to use both inductive and deductive reasoning approaches to solve problem of inferring cause and effect relationship from series of observations. The model mainly uses forward reasoning with series of nested calls to backward reasoning. In addition to declarative knowledge, the model is presented with new knowledge during the progress of the trial. As such, it is a good demonstration of how reasoning outcomes can change based on dynamic events even if the underlying set of inference rules remains the same.

Model of Casual Deduction Task

Cummins, Lubart, Alksnis and Rist (1991) and Cummins (1995) extensively studied the process of casual deduction. Subjects are provided with a sentence describing a cause/effect in a form of "*If <cause>, then <effect>*". The sentence is followed by four different forms of arguments: Modus Ponens (MP), Affirming the Consequent (AC), Modus Tollens (MT) and Denying the Antecedent (DA). Each argument consists of a fact and an implication. Subjects are asked to evaluate how likely it is that the implication is true given a cause/effect sentence and the argument's fact. Here is an original example from Cummins et al. (1991) of a cause/effect sentence: "*If the brake was depressed, then the car slowed down.*" The four arguments with respect to this sentence are: "*The brake was depressed. Therefore the car slowed down.*" for MP; "*The car slowed down. Therefore the brake was pressed.*" for AC; "*The car did not slow down. Therefore, the brake was not depressed.*" for MT; and "*The brake was not depressed. Therefore, the car did not slow down.*" for DA.

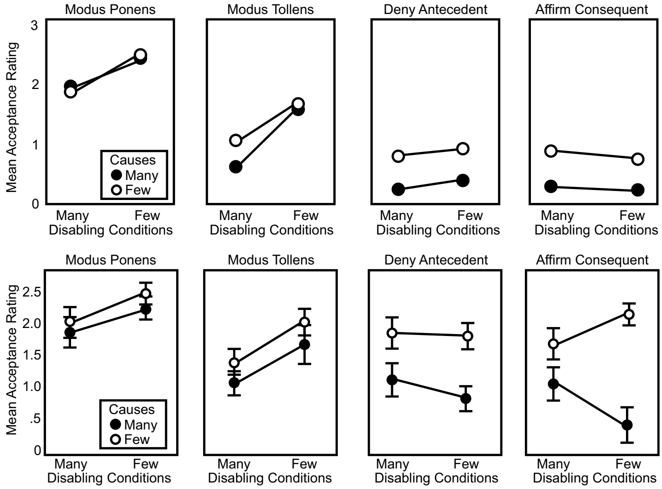


Figure 6.5: Mean acceptance ratings of four argument forms in casual deduction experiments conducted in (a) Cummins et al. (1991) and (b) Cummins (1995).

The study revealed that acceptance of arguments is influenced significantly by subjects' previous experience. The casual deduction was sensitive to two factors: alternative causes and disabling conditions (Cummins et al., 1991). An alternative cause is a cause that is different from one given in a sentence but still can result in the same effect. A disabling condition is a condition that prevents the effect from occurring despite the presence of a cause. Figure 6.5 shows the acceptance ratings of the four conditions gathered from two separate studies. Firstly, there is a robust effect of disabling conditions on acceptance of MP and MT arguments. When there are many possible disabling conditions, subjects are less likely to accept truthfulness of these two types of arguments. Secondly, there is a persistent effect of alternative causes on acceptance of DA and AC arguments. When there are many possible alternative causes of the effect, subjects are less likely to accept DA and AC arguments. Thirdly, it is not surprising that the acceptance rating varies a lot between two studies. The nature of the task is extremely subjective and participants' previous experiences vary a lot. It is likely that the rating further depend on the specific materials used in two experiments.

Using an ACT-R model that uses the HRM's knowledge structure, we explore the nature of effects invoked by alternative causes and disabling conditions on our ability of casual deduction.

Model's knowledge structure

In this experiment, the model used the same 16 cause/effect sentences described in Cummins (1995). The model stored both affirmative and negatives versions of all 16 sentences in its declarative memory in form of rules. For example, the previously mentioned example cause/effect sentence was converted to the following two rules:

Rule 1:	*Rule 2:*
(*have-state brake pressed*)	(*NOT-decrease car speed*)
==>	==>
(*decrease car speed*)	(*NOT-have-state brake pressed*)

Inside declarative memory, the model also had alternative causes and disabling conditions for each sentence. They were also stored in form of rules. Here is an example of affirmative and negative rules for an alternative cause:

Rule 3:	*Rule 4:*
(have-state car go-uphill)	*(NOT-decrease car speed t)*
==>	==>
(decrease car speed)	*(NOT-have-state car go-uphill)*

An affirmative version of the same disabling condition can be written as two following rule forms:

Rule 5:	*Rule 6:*
(have-state brake pressed)	*(have-state brake pressed)*
(have-state brake broken)	*(NOT-decrease car speed)*
==>	==>
(NOT-decrease car speed)	*(have-state brake broken)*

Both forms were stored in declarative memory. Finally, an example of a negative version of a disabling condition would be as following:

Rule 7:
(have-state brake pressed)
(NOT-have-state brake broken)
==>
(decrease car speed)

Sentences were divided into four groups. In Many/Many group, a sentence had three disabling conditions and three alternative causes. In Many/Few group, there were three disabling conditions and one alternative cause. Similarly, the other two groups were Few/Many and Few/Few.

Model's reasoning strategy

With each sentence, the model had to do four trials, one for each argument form. The model's general strategy was simple: given an argument, retrieve any matching rule from declarative memory and verify if the rule supports the argument. The workflow of the strategy is shown in Figure 6.6. Depending on the argument form, the model used different forms of reasoning. For MP arguments, the model did forward reasoning with fact. It retrieved any rule that had antecedent rule-statement matching the fact and checked if retrieved rule's consequent matched the implication. If a match was found, then the argument was accepted. For AC arguments, the model did backward reasoning with fact: it retrieved any rule that had consequent matching the fact and checked if any of the antecedent rule-statements matched the implication. If match was found then argument was accepted. In a similar manner, the model did forward reasoning with fact for MT arguments and forward reasoning with implication for DA arguments.

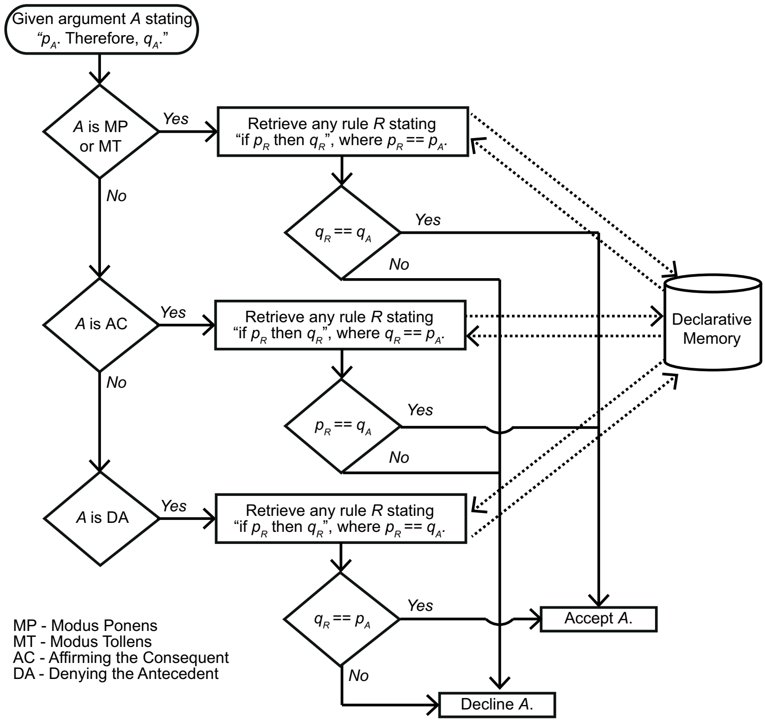


Figure 6.6: A workflow of the strategy used by the model of the Casual Deduction task.

Results

The model repeated the same experiment 50 times, accounting to a total of 3200 trials. Figure 6.7 shows proportions of trials where arguments were accepted. The proportions were calculated separately for each combination of four argument forms and sentence groups. The model shows the same behavior as human subjects. The model is more likely to accept MP and MT arguments for cause/effect rules that have few disabling conditions. Next, the model is more likely to accept DA and AC arguments for cause/effect rules that have few alternative causes.

The effects are explained by a mutual interference among rules during the step when the model tries to retrieve a proper rule that can support an argument. For example, let us assume that the model received following MP argument:

Fact: (have-state brake pressed)
Implication: (decrease car speed)

In this scenario, the model will use the fact (*have-state brake pressed*) to retrieve any rule with matching antecedent rule-statement. These include not only the original cause/effect rule 1, but also the affirmative and negative disabling condition rules 5, 6 and 7. In presence of several matching chunks during a declarative retrieval, ACT-R randomly picks one. The rules 5 and 6 have consequents that are different from the argument's implication. Therefore, if either rule 5 or 6 is retrieved then the model will not accept the argument's implication. It is quite easy to see that as the number of disabling conditions increases, the model will be less likely to retrieve a rule that supports the argument and, hence, more likely to reject it. This rule interference mechanism is also responsible for the effects observed in other three argument forms.

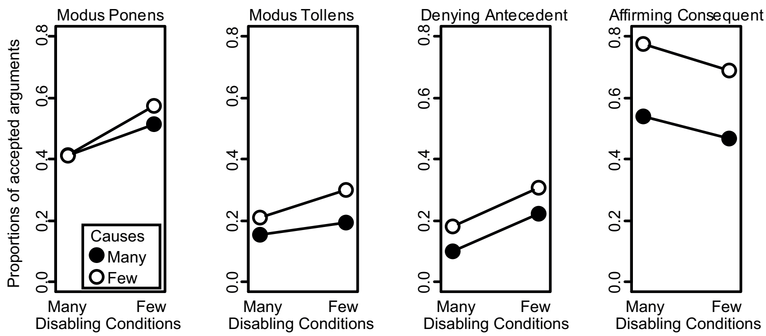


Figure 6.7: Proportions of arguments accepted by the model in four forms of arguments.

One aspect that should be considered is that the rules have the same activation values in the model. Hence, the rules have the same probability of retrieval. This is an unlikely scenario with human subjects. Firstly, an activation value for the same rule may differ between subjects. Secondly, different rules may have different activation values within a subject. For example, despite leading to the same conclusion, rule 7 is likely to have less activation than rule 1 because people do not worry often about state of the brakes. Use of equal activation values for all rules may have affected model fit. The model's acceptance rate of AC arguments is a bit higher than subjects' rate. Assigning lower activation values to negative versions of disabling conditions, such as rule 7, can decrease the acceptance rate of AC arguments and result in better fit. However, the current simpler version of the model serves better for the purpose of demonstrating the HRM's basic reasoning capabilities.

It is certainly possible that other computational models can explain the same effects. However, in case of our model the main explanatory power comes not from the model built for this specific task, but rather from the aspects of the cognitive architecture: a combination of ACT-R's activation-based declarative memory and the HRM's conditional proof schema (Braine & O'Brien, 1991).

Model of Spatial Relations Task

This task is used to study people's fundamental ability to derive a spatial relation from a set of premises. Three problems below are examples of such task. In each problem, subjects are given four premises and then queried about the spatial relation between two items that were not explicitly connected in any of the premises.

The studies showed that people prefer to use strategy of mental states rather than formal representations (Byrne & Johnson-Laird, 1989). In such strategy, people build mental states or imagery using abstract objects representing items in the premises. Such mental state is built iteratively as premises are processed one by one (Carreiras & Santamaria, 1997). With such mental states, the spatial relation between two query items can be derived directly. Examples of mental states are shown below. Problem 1 results in one mental state. Problems 2 and 3 result in two possible mental states. Furthermore, the same studies have shown that one-state problems are easier than two-state problems.

Problem 1:
1. A is on the right of B
2. C is on the left of B
3. D is below C
4. E is below B
What is the relation between D and E?

Possible mental state:

```
C B A
D E
```

Problem 2:
1. B is on the right of A
2. C is on the left of B
3. D is below C
4. E is below B
What is the relation between D and E?

Possible mental states:

```
C A B      A C B
D   E        D E
```

Problem 3:
1. B is on the right of A
2. C is on the left of B
3. D is below C
4. E is below A
What is the relation between D and E?

Possible mental states:

```
C A B      A C B
D E        E D
```

Byrne and Johnson-Laird (1989) reported 61% and 50% correct responses in one- and two-state problems respectively. Similarly, Carreiras and Santamaria (1997) reported 99% and 89% correct responses in one- and two-state problems. There are also two-state problems that have no valid conclusion. In those problems, mental states resulted in contradicting relations between two query items, and subjects were required to report that there is no single solution. For example, Problem 3 results in two possible mental states contradicting each other. Problems with no valid conclusion result in the lowest proportion of correct responses. Two separate experiments by Byrne and Johnson-Laird resulted in 18% and 15% of correct responses in problems with no valid conclusions.

It is assumed that a two-state problem is more difficult because it requires higher working memory load than a one-state problem. However, it does not explain why accuracy drops even lower in a two-state problem with no valid conclusion. Both types of two-state problems have equal numbers of mental states, premises and items. Furthermore, both types of problems require two swaps to derive the second mental state from the first one. Therefore, the working memory load should be the same in both types of problems. As result, an explanation based on a working memory load is not sufficient to explain subjects' performance. Our ACT-R model that uses the HRM module provides a possible explanation for this effect.

Model's design

The model's strategy can be divided into five steps:

1. The model constructs a mental state of the problem inside VSTM. The mental state is built iteratively by processing premises one at a time and updating VSTM on each iteration. Items from a premise are converted into abstract visual objects and given (*x, y*) coordinates based on positions relative to the items already existing inside VSTM. A premise is also converted into a logical *statement* stored inside declarative memory, but it is done only after VSTM is updated (Figure 6.8a). The model can handle two-state problems. For example, while processing the second premise in Problem 2, the model uses *assertion* (*r-dir-left-of "@item" B*) to check if there is already another item present to left of *B*. This assertion triggers bottom-up spatial reasoning and the HRM returns any visual object that is to the left of *B*. In case of Problem 2, *A* is returned. Then the model stores both *C* and *A* inside its working memory as items to be swapped positions in a second mental state (Figure 6.8b). In Problem 2, the mental state inside VSTM will be as following at the end of step 1:

C A B
D   E

2. After all premises are processed and a mental state is built inside VSTM, the model sends an assertion to the HRM to try to answer a query. The assertion is in form of (*"@property" D E*). To answer the assertion, the HRM uses bottom-up spatial reasoning to evaluate relative positions of *D* and *E* inside VSTM. In case of Problem 2, the model's answer will be either (*r-left-of D E*) or (*r-dir-left-of D E*).

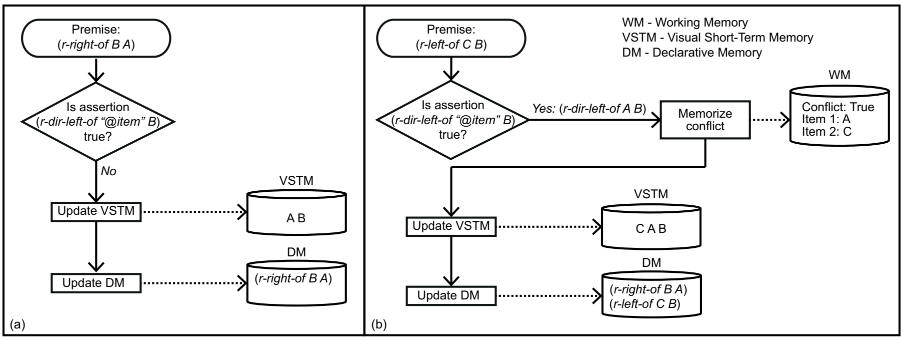


Figure 6.8: Two diagrams are describing how the model processes the first (a) and the second (b) premises of the Problem 2 during step 1 of the strategy.

3. If it is a one-state problem then the model does nothing else. However, if there are two possible mental states then, after answering the query and storing it in declarative memory, the model creates the second possible mental state inside VSTM. This is done by swapping positions of the two objects previously stored inside working memory. In case of Problem 2, *C* is placed at the position of *A*, and *A* is placed at the former position of *C* changing the mental state inside VSTM into following:

A C B
D   E

4. At this step, the model checks if any visual object was positioned relative to the swapped objects. If that is the case then the model verifies if relations still hold, if not then positions of those objects are corrected as well.

5. After creating the second mental state, the model sends to the HRM the same assertion as in step 2. The answer for this assertion is compared to the answer from step 2 that is retrieved from declarative memory. If answers are not the same then the model assumes that problem does not have valid conclusion and reports the inconsistency.

Results

Model's proportions of correct responses in one-state problems, two-state problems with valid conclusion and two-state problems with no valid conclusion are 100%, 74% and 31% respectively. The model always gives correct answers in one-state problems. However, it starts making mistakes in two-state problems. Furthermore, the model shows lowest accuracy in two-state problems with no valid conclusion. The cause of mistakes is model's confusion between similar spatial properties such as *r-below* and *r-dir-below*.

The first mistake can be made during step 4. Consider following example from Problem 3 where the model just finished step 3 by swapping positions of *A* and *C*:

```
C A B        A C B
       ==>
D E          D E
```

During step 4, the model has to verify whether the spatial relation between *D* and *C* still holds. One of two possible assertions can be used for such verification: *(r-below D C)* or *(r-dir-below D C)*. The model's choice is random in this case. However, if *r-below* is used then the assertion will be evaluated to be true since bottom-up reasoning with *r-below* does not check for vertical alignment. This leads the model to a wrong conclusion that *D*'s position does not need to be corrected. Such mistake can lead to a situation where, for example, in Problem 3, the relation between *D* and *E* is still the same in both mental states. The second mistake can be made during comparison in step 5. Let us consider the case where, in Problem 2, the answers to the assertions in step 2 and 5 were *(r-left-of D E)* and *(r-dir-left-of D E)* respectively. These two statements, although similar, are not the same. Hence, if no explicit top-down reasoning is used to prove that one entails the other, the two answers are considered different. The model decides randomly whether to invoke top-down verification since it is not always necessary.

The model makes more mistakes in two-state problems with no valid conclusion because it is vulnerable to both types of mistakes in those problems. However, only second mistake is possible in two-state problems with valid conclusion. In one-state problems, the model is not susceptible to any of those mistakes.

Model of Bayesian-like Inference in Blicket Task

We focused on a simulation of the first experiment conducted by Griffiths et al. (2011). The task context consists of ordinary pencils (blocks) and super pencils (blickets). We further refer to ordinary and super pencils as blocks and blickets. Subjects were asked to rate on a scale of 1-7 how likely a block was to be a blicket. Subjects' ratings were based on observations that consisted of one or two blocks placed on a special detector. The detector activated when at least one blicket was placed on it. Griffiths et al. used this task to study people's ability to infer casual relations based on number of observations and prior knowledge.

The experiment consisted of three consecutive phases: a training phase, AB-event phase and A-event phase. Subjects were divided into five groups that received different trainings during the first phase. During the training, each subject was shown ten blocks placed individually on the detector one after another. Some blocks activated the detector (Figure 6.9a) others did not (Figure 6.9b). A subject's group determined the frequency of blickets among the ten blocks. In group 1/6, only one of ten blocks was a blicket. In group 1/3, three of ten were blickets. Similarly, subjects in groups 1/2, 2/3 and 5/6 observed five, seven and nine blickets respectively.

After the training phase, the subjects were shown two new blocks, A and B. At this point, subjects were asked to provide initial ratings of how likely each was to be a blicket. Following the initial ratings, both A and B blocks were simultaneously placed on the detector causing it to detect a blicket (Figure 6.9c). This phase is referred to as AB-event. After AB-event, subjects were asked to rate both blocks again. Finally, block A was placed alone on the detector activating it (Figure 6.9d). This phase is referred to as A-event. Subjects were asked to rate A and B blocks after A-event as well.

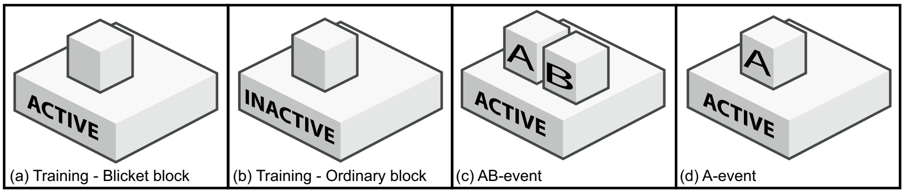


Figure 6.9: A blicket activates the detector during the training phase. (b) The detector remains inactive when ordinary block is placed on it during the training phase. (c) Two blocks, A and B, are placed on the detector activating it during AB-event. (a) During A-event, only block A is placed on the detector activating it.

Before conducting the experiment, Griffiths at al. (2011) created a Bayesian model predicting the probabilities of objects A and B being rated as blickets. Figure 6.10a shows those predictions for all five groups. According to the model predictions, the initial ratings reflect prior probabilities of encountering a blicket established by a training phase. Those ratings are higher in groups that observe a higher number of blickets during the training phase. After AB-event, the mean ratings increase above baseline level. However, such increase gets smaller as baseline prior probability gets higher. After A-event, the object A is given a maximum rating. However, the rating of object B goes down. As shown in Figure 6.10b, subjects' mean ratings closely follow predicted Bayesian probabilities.

Model's knowledge structure

In addition to the basic set of concepts describing elements of the task, the model started with a core set of inference rules that are used to reason based on both previous experience and real-time evidence. Those rules are described on Table 6.1.

Rules 0 and 1 reflect the commonly reported simple inductive strategy of solving a problem by analogy (Gentner, Holyoak, & Kokinov, 2001; Winston, 1980). Analogies are the basis for any integrated cognitive systems (Gust, Krumnack, Kühnberger, & Schwering, 2008). Therefore, it is reasonable to assume that subjects have rules to classify blocks by analogy given uncertainty.

Rules 2-7 reflect the task structure and instructions subjects receive during the introduction to the experiment. Subjects were given demonstrations of blocks and blickets and their interactions with the detector. Subjects were shown cases with one and two blocks blocks placed on the detector simultaneously. The demonstrations were given to ensure that subjects clearly understood the activation laws. Rules 2 and 3 reflect laws of activation when only one block is placed on the detector. Rules 4-7 reflect laws of activation when two blocks are placed on the detector at the same time.

Rule 8 is based on a *backward blocking* paradigm (Chapman, 1991; Miller & Matute, 1996; Shanks, 1985). When subjects are shown two cues (A and B) occurring with an outcome, subjects associate both cues with the outcome. Next, if subjects are shown only one of those cues (A) occurring with the outcome then subjects associate only the latter cue (A) with the outcome. The diminished association between the second cue (B) and the outcome in light of latter evidence is backward blocking effect. Furthermore, Sobel, Tenenbaum and Gopnik (2004) found that the degree of cue-outcome association in backward blocking is affected by the base rates of blickets. Similarly, Rule 8 considers the order of evidence and base rate of blickets to re-evaluate chances of a block being a blicket.

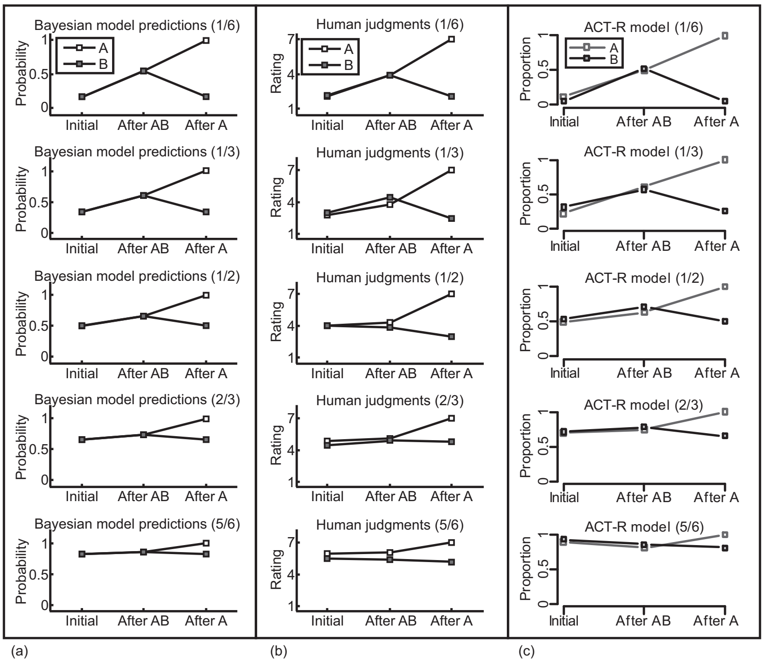


Figure 6.10: (a) Probability predictions of the Bayesian model created by Griffiths et al. (2011). (b) Human mean ratings of A and B pencils at initial stage, after AB event and after A events. (c) Probabilities produced by our ACT-R model.

Table 6.1: Core set of rules used by the model to categorize A and B blocks.

Rules	Descriptions
Rule 0: *(have-role"@block1" "@role" (ts "base"))* *==>* *(have-role "@block2" "@role" (ts "init"))*	If a block on the antecedent has some category then assign the same category to the block in the consequent.
Rule 1: *(have-role "@block" "@role" (ts "@t1"))* *==>* *(have-role "@block" "@role" (ts "@t2"))*	If a block had some category at some time *t1* then it has the same category at some time *t2*.
Rule 2: *(alone-on "@block" Detector (ts "@t1"))* *(have-state Detector Active (ts "@t1"))* *==>* *(have-role "@block" Blicket)*	If a block is alone on the active Detector then it is a blicket.
Rule 3: *(alone-on "@block" Detector (ts "@t1"))* *(have-state Detector Inactive (ts "@t1"))* *==>* *(have-role "@block" NON-Blicket)*	If a block is alone on the inactive Detector then it is not a blicket.
Rule 4: *(on "@block1" Detector (ts "@t1"))* *(on "@block2" Detector (ts "@t1"))* *(have-state Detector Active (ts "@t1"))* *==>* *(have-role "@block1" Blicket)*	If, at the same time, two blocks are on the active Detector then the first block is a blicket. *Rule 5* is similar to *Rule 4*, but concludes that the second block is a blicket.
Rule 6: *(on "@block1" Detector (ts "@t1"))* *(on "@block2" Detector (ts "@t1"))* *(have-state Detector Active (ts "@t1"))* *(have-role "@block1" Blicket (ts "@t1"))* *==>* *(have-role "@block2" NON-Blicket)*	If, at the same time, two blocks are on the active Detector, and one of the blocks is a blicket then the other block is not a blicket. The *Rule 7* is similar to *Rule 6,* but concludes that the first block is not a blicket.
Rule 8: *(alone-on "@block1" Detector (ts "@t1"))* *(have-state Detector Active (ts "@t1"))* *(on "@block1" Detector (ts "@t2"))* *(on "@block2" Detector (ts "@t2"))* *(have-state Detector Active (ts "@t2"))* *(have-role "@block2" NON-Blicket)* *==>* *(have-role "@block2" NON-Blicket)*	If there are two possible blocks that can activate Detector, and one was observed to activate the Detector alone, and the other one is likely not to be a blicket then the latter is not a blicket.

Model's overall strategy

The overall strategy consists of two major steps. The first step is evaluating presented evidence. This step is done every time the model is presented with one or more blocks placed on the detector. The model categorizes each block based on the detector's state and prior knowledge. Such evidence evaluation is done via forward reasoning using rules on Table 6.1. The resulting categorizations of blocks are stored in model's declarative memory.

The second step is a query response. When a query asking to categorize a block is received, the model tries to retrieve from declarative memory the recent categorization of the queried block. If the retrieval is successful then the retrieved categorization is reported. Otherwise, the model uses an analogy-based induction to decide block's category. The model retrieves another block that was already categorized and assigns its category to the queried block. The second step was implemented as two sequential backward reasoning calls for retrieval and analogy-based induction respectively.

Model's strategy for training phase

During the training phase, ten blocks are sequentially presented to the model. For example, the evidence presented to the model for the first block is:

(alone-on Block1 Detector (ts "base"))
(have-state Detector Active (ts "base"))

The model uses forward reasoning to evaluate evidence and categorize ten blocks. Only Rules 2 and 3 are used because those rules always provided the best match to the presented evidence. The resulting categorization stored in the model's declarative memory can be as following: *(have-role Block1 Blicket (ts "base"))*. The two rules represent typical instructions human subjects would receive during the task.

Next, the model receives an initial request to categorize A and B blocks. Since the model has no existing categorization of the two blocks in its declarative memory, it has to use analogy-based induction to categorize each block. A backward reasoning with an example assertion *(have-role BlockA "@role" (ts "init"))* invokes Rule 0 from Table 6.1. Antecedent from Rule 0 triggers retrieval of any category statement belonging to one of ten blocks categorized during the training phase. Because all ten blocks have equal probabilities of retrieval, the probability of block A being categorized as blicket is equal to a prior probability of blickets established during the training phase. For example, if the model retrieves *(have-role Block2 NON-Blicket (ts "base"))* then block A will be also categorized as non blicket: *(have-role BlockA NON-Blicket (ts "init"))*.

Model's strategy for AB-event

The evidence for AB-event is presented to the model as:

(on BlockA Detector (ts "AB"))
(on BlockB Detector (ts "AB"))
(have-state Detector Active (ts "AB"))

The order of the first two statements in the evidence is random. Given such evidence, the model uses a forward reasoning to categorize both A and B blocks during AB-event. Rules 4-7 have equal match to provided evidence. Rules 4 and 5 result in a block being categorized as blicket, while Rules 6 and 7 can result in a negative categorization. Four rules allow the model to guess based on the notion that at least one of the blocks should be a blicket without excluding the probability that the other one may not be a blicket. Because of Rules 6 and 7, positive categorization for one block can result in negative categorization of another block.

After evidence evaluation, the model is queried about A and B. The models reports with categories it has inferred during AB-event. The model may fail to categorize a block if either Rule 6 or 7 is used and the model does not have any fact supporting the last antecedent rule-statement *(have-role "@block1" Blicket (ts "@t1"))*. In such cases the model reports category inferred during the training phase. This analogy-based induction is governed by Rule 1 and allows the model to fall back to prior decision if it is confused by ambiguous evidence such as in AB-event.

Model's strategy for A-event

The evidence for A-event is presented to the model as:

(alone-on BlockA Detector (ts "A"))
(have-state Detector Active (ts "A"))

Given this evidence, the model again has to infer the categories for A and B. Inferring A's category is straightforward since Rule 2 is always the best match to infer A's category given the evidence above. Correspondingly, block A is always categorized as blicket.

Inferring B's category is trickier since above evidence does not provide any information about B. The model uses Rule 8 to infer B's updated category. This rule was introduced to the model based on the effect of *backward blocking*. Backward blocking is observed in a task with two potential causes (A and B). It was found that subjects who observe that A alone can cause the outcome are less likely to accept B as a second cause than subjects that only observed A and B causing the outcome together (Shanks, 1985). Rule 8 allows the model to backward block B if it was previously observed together with A. According to the rule, if at any time B was categorized as non-blicket then that decision will be reinforced given the positive evidence about A.

Model results

The model repeated the experiment 50 times. Proportions of times the model reported A and B as blickets are shown in Figure 6.10c. The model's good fit supports the hypothesis that the casual learning in blicket tasks is not simply associative (Griffiths et al., 2011). Furthermore, our model provides a detailed account of underlying cognitive processes happening in human brain. The original Bayesian model by Griffiths et al. lacks such explanatory power. In addition to reflecting a knowledge structure required for the task, rules also govern how the knowledge should be evaluated and updated.

The most intriguing aspect of our model is its ability to simulate Bayesian-like inference despite using an inherently deterministic rule-based inference. Just like the Bayesian model, our model is able to incorporate not only the immediate knowledge, but also prior knowledge that is being constantly updated throughout the task. Such behavior is facilitated by the fact that outcome of new inference is dependent on outcome of the previous inference. Furthermore, there are multiple competing rules that can be used for the same inference, and probabilistic nature of DM's retrieval is the defining factor over which rule is chosen.

Discussion and Conclusion

In this study we have proposed a computational module of human reasoning system called the HRM. We have also described three models of different reasoning tasks. These models tested and validated individual cognitive functionalities of the HRM based on a fit to human data.

The model of spatial relations task shows an in depth view of how rule- and mental model-based reasoning strategies are used together in the same task. It is imperative for success that both strategies complement each other. The core of model-based reasoning is bottom-up reasoning, an ability to derive explicit knowledge from an implicit knowledge. Although fast and efficient, bottom-up reasoning has limitations on the complexity of semantics it can operate. Those limitations make the model-based reasoning prone to mistakes if not corrected by rule-based reasoning. On the other hand, top-down rule-based reasoning is a slow and costly process not feasible for real-time interactive tasks. It has to rely on a model-based reasoning to speed up the reasoning process. When a reasoning pipeline recursively calls itself, it blurs the boundary between rule- and model-based strategies since both of them may be used during the same reasoning process.

The first casual deduction model is a demonstration of how a triple-based knowledge structure can help to explain how complex background knowledge can influence an outcome of even simple deductive reasoning. As such, it is no longer a deductive reasoning, but rather a pragmatic reasoning, a reasoning based on both a given propositional form and its content, previous knowledge (Braine & O'Brien, 1991). It is interesting to see a rise of the pragmatic reasoning in the HRM since it does not incorporate any dedicated controls for it. The very dependency of the HRM's deductive reasoning on ACT-R's declarative mechanisms gives rise to a quite natural pragmatic reasoning. As such, there is a possibility that a pragmatic reasoning is not a different logical process, but a deductive reasoning bound by properties and limitations of our long-term declarative memory.

The model of blicket task further extends the notion of pragmatic reasoning and steps into a territory of Bayesian probabilistic inference. The model's good fit challenges the traditional view of vertical division between deterministic and probabilistic inferences of human reasoning. Instead, the model shows that given an inconsistent nature of human memory and uncertainty of its recall the deterministic inference can become probabilistic.

One of the unexpected outcomes of this study is a seamless unification of similarity- and rule-based reasoning within the HRM. Earlier studies suggested that both rule- and similarity-based processing may emerge from application of a single learning rule (Pothos, 2005; Verguts & Fias, 2009). In the blicket model, Rules 0 and 1 are used for similarity-based reasoning while others are rules defined by the task. Both types of rules are handled by the HRM's reasoning pipelines, and transition from one form of reasoning to another is seamless and on-demand.

The eventual goal of developing Human Reasoning Module is to create a unified theory of human reasoning and a practical tool for simulating it. As such, the HRM was designed to be general and task-independent. It is not constrained to specific formal system of logic. These properties make the HRM potentially suitable for modeling wide variety of tasks. However, the same properties raise concerns whether the module can reliably simulate human behavior in specific reasoning tasks. We tried to mitigate those concerns by modeling three different tasks that address human reasoning from very different perspectives.

We are still in the process of elaborating what the unified model of human reasoning should be. However, the HRM is promising to be a step in the right direction. Figure 6.11 shows the six forms of reasoning used by the three models. The first dimension unifies two popular theories of mental logic and mental models. The HRM assumes that a mental model is a form of working memory, Visual Short Term Memory, which has the capability to extract basic semantic relations from its content using fast and efficient bottom-up cognitive processes. Then, these semantic relations can be used by mental logic to perform more complex semantic processing. Therefore, the HRM argues that human reasoning is not strictly top-down and can rely on subconscious bottom-up processes to evaluate semantic relations. The second dimension unifies probabilistic inductive reasoning and deterministic deductive reasoning. The HRM suggests that the human general reasoning skill is likely to be inherently probabilistic and inductive due to stochastic nature of knowledge access and retrieval. However, deterministic deductive reasoning is still possible when knowledge-related uncertainty is minimized. Ideally, deductive reasoning is an instance of inductive reasoning with zero uncertainty. Therefore, the amount of uncertainty is the common dimension that unifies inductive and deductive reasoning. Furthermore, a degree of uncertainty may be one of the main factors defining reasoning strategy. Inductive reasoning can be viewed as an instance of probabilistic reasoning with a strong prior toward a particular conclusion. Probabilistic reasoning is inference based on significant past experiences defined by strengths of cause/effect, pre-condition/action, action/post-condition observations. Inference without prior knowledge about the given instance is either reasoning by analogy or simply guessing. In the HRM, reasoning by analogy is still done via rule-based reasoning. This unification of similarity-based and rule-based reasoning is the final dimension depicted in Figure 6.11.

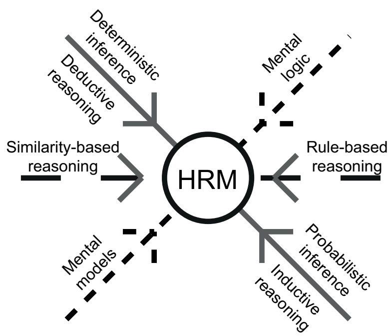


Figure 6.11: Current forms of reasoning that were used by three models based on HRM.

Many open questions still remain. One of them is still how inference rules are constructed. For example, verbal instructions given to subjects in blicket task should be somehow translated into set of rules shown in Table 6.1. On the one hand, it is possible that we have set of general rules that serve as templates and are translated into task specific forms. On the other hand, there might be set of meta-rules similar to schema that govern how inference rules should be constructed based on the perceived information.

The source code and related data for the HRM module and validation models can be downloaded from here: http://www.ai.rug.nl/~n_egii/models/. The current implementation of the HRM is in the prototype phase, and its features may change with future revisions.

CHAPTER 7

Qwirkle: From fluid reasoning to visual search

Abstract

This study focuses on subjects' performance and strategies in the board game called Qwirkle. Earlier scientific studies mostly used Qwirkle as a tool for improving mathematical reasoning skills in children. However, Qwirkle may additionally require significant visuo-spatial processing. Results from our study indicate that subjects use a simple local maximum strategy in which scores at local decision points of the game are maximized. As was expected, the strategy requires mathematical reasoning skills. However, results also suggest that subjects' performance is significantly affected by visual search skills. We conclude that the visual and reasoning systems are deeply intertwined. On the one hand, the reasoning outcome is highly dependent on both attentive and pre-attentive visual knowledge. On the other hand, visual processing requires capability of reasoning on concepts more higher level than visual features.

Introduction

In the previous chapters, we have tackled two major issues that commonly occur in complex problem solving tasks. The first is the role of our visual system as a major source of real-time information (Nyamsuren & Taatgen, 2013a, 2013b, 2013c). The result of this research was the Pre-Attentive and Attentive Vision module (Nyamsuren & Taatgen, 2013a). The second is task-general declarative and procedural knowledge that enables us to reason and understand how specific problems should be approached and solved. The result of this study was the Human Reasoning Module (Nyamsuren & Taatgen, 2013d).

As was mentioned in Chapter 1, isolated understanding of human reasoning and human vision is not enough to understand the cognitive underpinning of human problem solving. For a complete picture, we should also understand how reasoning and visual systems interplay in a single coherent cognitive architecture. We have yet to draw a bridge between PAAV and the HRM that allows full interaction between the two modules. The model of spatial reasoning task (Nyamsuren & Taatgen, 2013a; Byrne & Johnson-Laird, 1989) described in Chapter 6 only scratched the surface of necessary cognitive functionalities connecting the two systems. Therefore, additional investigation based on a more complex task is required.

Qwirkle

We have chosen the board game of Qwirkle[4] as a representative of a problem solving task that requires both complex visual processing and reasoning. Qwirkle is a competitive game that requires at least two and at most four players. It is a game based on tiles. Each tile has a shape of a certain color. There are six unique colors and six unique shapes resulting in 36 unique tiles shown in Figure 7.1. In total, there are 108 tiles with three copies of each unique tile. Tiles are usually kept in a bag so they are not visible to players.

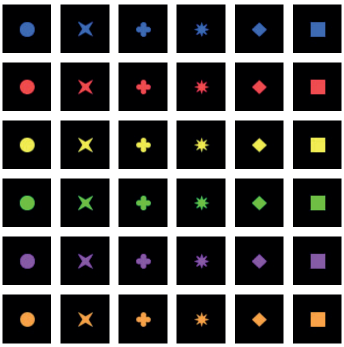


Figure 7.1: 36 unique tiles of Qwirkle.

The game starts with each player drawing six random tiles from the bag. Next, another three tiles are drawn randomly and put on the center of the board face up next to each other. None of the players can see the other players' tiles. Players make moves in turns. During her turn, a player can perform either one of two actions: put one or more tiles with the same color or shape on the board, or replace one or more tiles from her stack with random tiles

[4] Qwirkle is a game by MindWare (www.mindware.com).

from the bag. The replaced tiles are put back into a bag. After putting tiles on a board, a player replenishes her six-tile stack with new tiles randomly picked from a bag. It is not necessary for a player to put all tiles in the same row or column. Instead, individual tiles can be put in places where they fit best. There are two main rules governing where a tile can be put. First, a tile should be put next to another tile that is already on the board. Second, any sequence of tiles on the board should have either the same color and different shapes or different colors and the same shape. The longest possible sequence consists of six tiles and is referred to as a *qwirkle*.

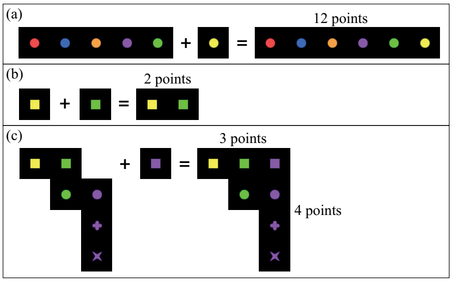


Figure 7.2: Scoring rules in Qwirkle.

A player receives points for each tile put on a board. A player who puts the final tile of a qwirkle receives 12 points (Figure 7.2a), the maximum amount of points possible to get from a single tile sequence. Otherwise, scoring is based on the length of the sequences a new tile forms. For example, forming a sequence with two tiles results in two points (Figure 7.2b). If a newly put tile forms a horizontal sequence with three tiles and a vertical sequence with four tiles then the move results in seven points (Figure 7.2c).

The game ends if a player puts her last tile on a board and there are no more tiles in the bag to replenish from. A player who finishes the game first receives bonus six points. The player with the highest amount of points is the winner of the game. A single-player version[5] of the game can be found at www.ai.rug.nl/~n_egii/qwirkle/.

What makes Qwirkle interesting?

Qwirkle requires a significant degree of reasoning skills including mathematical problem-solving skills and the ability to consider alternative options. In fact, Qwirkle has been used to improve mathematical problem solving skills in schoolchildren (Klanderman, Moore, Maxwell, & Robbert, 2013; Maloy, Edwards, & Anderson, 2010). Furthermore, Mackey, Hill, Stone and Bunge (2011) argued that computerized and non-computerized reasoning games, with Qwirkle among them, improve children's fluid reasoning, the capacity to think logically and solve problems in novel situations (Cattell, 1987; Horn & Cattell, 1967). Although not widely recognized in previously mentioned studies, Qwirkle has a significant visual component in it. On the one hand, it requires visuo-spatial reasoning. In fact, Mackey et al. also found that subjects playing Qwirkle along with other games significantly improved their spatial working memory. On the other hand, Qwirkle requires basic processes of visual feature-based search. This dual nature of Qwirkle that involves both the

[5] The online and experiment versions of the Qwirkle game were developed and used under the terms of fair use for non-profit educational purpose only.

reasoning and visual systems makes the game an ideal candidate for investigating how both processes work together in single task.

Objectives

The purpose of this study is to investigate the form of a strategy the players use in Qwirkle. Revealing the strategy is crucial for detailed understanding of the types of cognitive processes involved in the game. The overall strategy can be divided into individual steps, and each step can be assigned into specific cognitive resources. In this way, it is possible to investigate the specific roles the reasoning and visual systems play in Qwirkle. Furthermore, it should give an insight into the interplay between these two systems.

Experiment

Subjects

In total, 17 subjects participated in the experiment. Results from three subjects were excluded from the analysis due to technical errors and the high amount of noise in the eye tracking data. The average age of the subjects was 22 ($SD = 3.29$). There were six female and eight male subjects.

Design and procedure

Each subject was required to play ten games against a single computer opponent. The computer opponent had the simple strategy of maximizing its score for each turn. In each turn, the computer opponent would consider all possible unique combinations out of six (or less) tiles it has in its stack. For each combination of tiles, the computer opponent found a combination of positions on the board that resulted in maximum amount of points. Finally, the combination of tiles and corresponding combination of positions on the board that gave the maximum possible number of points for the turn were chosen as the computer's move for the turn. Computer opponent did not plan ahead or consider subject's moves. Hence, it is not the optimal strategy for the game. Subjects were not informed about the strategy used by the computer opponent.

The experiment was divided into two blocks of five games each: a hint block and a no-hint block. In the hint block, the subject received hints at the start of each of her turns. A hint consisted of one of the six tiles in the subject's stack being highlighted with a red frame. The hint indicated that the tile belongs to a combination of one or more tiles that results in the highest possible score for the turn. Subjects were given instructions about the meaning of the hint. Subjects were also explicitly told that they were free to ignore the hint and pursue their own strategy. Half of the subjects started the experiment with a hint block while the other half started with a no-hint block. Hints served as a good reference point to deduce the strategy if subjects chose to use it.

A single game can last for quite a long time, especially if there are only two players. Therefore, any single game used only 54 tiles instead of 108 tiles. All 54 tiles were chosen randomly for each game. In addition, subjects were limited to 90 seconds to make moves in each turn. Figure 7.3 shows the example screen capture of the game during an experiment. The game board had a size of 15×15 cells. All games started with subjects' making the first moves.

All subjects were requested to read the game instructions and play an online version of it prior coming to the experiment. The online version of Qwirkle was similar to the version used in the experiment except having no computer opponent. At the beginning of the experiment, an experimenter again explained the instructions to the subject. Subjects also had an opportunity to play two practice games: one with and one without hint. Results from practice trials were not included in the analysis.

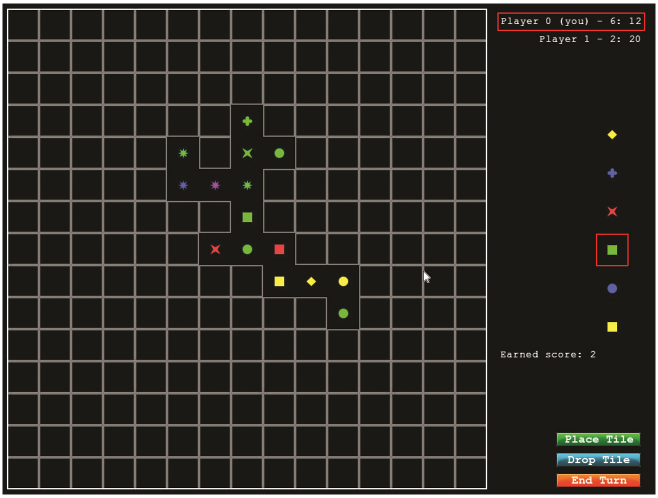


Figure 7.3: Example screen capture of Qwirkle game during an experiment.

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 500 Hz and a spatial resolution of <0.01° RMS. The card images were shown on a 20-inch LCD monitor with screen size of 1,024×768 pixels and screen resolution of 64 pixels/inch. Subjects were asked to use a chin-rest to fix the head position during a recording. The tile image had a size of 50×50 pixels, or 1.62º×1.62º in angular size. The image of a shape within a tile fitted inside a square of 15×15 pixels, or 0.49º×0.49º in angular size. Angular sizes were calculated based on a viewing distance of 70 cm. The gaze position was calculated using the eye's corneal reflection captured using 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 game, subjects were asked to do a drift correction as an additional corrective measure.

Action log and questionnaire

The progress of each game was recorded in a log file. The log file contained information about every action (placing a tile on the board or replacing a tile) performed by both the subject and the computer opponent. The log file contained sufficient information to restore any player's state or board state at any time during the game. At the end of the experiment, subjects were asked to fill in short questionnaire. Subjects were requested to provide information about their expertise level and previous experience with Qwirkle. They were also asked a few specific questions regarding the strategy such as preference toward any attribute, predicting opponent's moves or planning several turns ahead.

Experiment results

According to the questionnaires, none of the subjects had previous experience of playing Qwirkle prior to registering for the experiment. However, all subjects played the online version of the game prior coming to the experiment.

Attribute preference

In our previous studies with the card game of SET (Nyamsuren & Taatgen, 2013b), we have found that players had a preference for the color attribute over any other attribute such as shape. Surprisingly, we were not able to find any evidence of attribute preference in Qwirkle. We have tested whether subjects used combinations of tiles with the same color more than combinations of tiles with the same shape and vice versa. The usage did not significantly differ from one another. The statistical results also match with answers provided in the questionnaires. Eight subjects reported absence of preference toward either color or shape. Three subjects reported preference toward color, and another three subjects reported preference toward shape. In overall, there is no overall preference either toward color or shape.

Subjects' scores

Figure 7.4 shows subjects' performance in terms of proportions of games won and total amount of points gathered during an experiment. Subjects were sorted in increasing order of their total scores. Three subjects who won 50% of games each showed the highest performance. Two subjects who won only 10% of games each showed the lowest performance. There is a strong correlation between number of games won and total points gathered, $r(12) = 0.64$, $p = 0.014$. Increasing a score is not the only possible strategy in Qwirkle and does not necessarily guarantee a victory. For example, 43% of subjects reported in the questionnaire that they tried to block the opponent from completing the qwirkle. Similarly, 57% of subjects reported that they would be hesitant to put a fifth tile in a sequence without having a sixth tile because the opponent may put the final tile in the next turn. These strategies are highly situational, but still may affect the outcome of a game. Furthermore, a high score does not guarantee a victory, since it is always relative to the opponent's score. Nevertheless, the significant correlation suggests that gathering more points improves the chance of winning the game. Further in this work, we will treat the total score as a main indicator of subjects' performance.

If players were at least as good as the computer opponent then the success rate should be around chance probability of 50%. However, subjects had a relatively low success rate with on average three wins out of ten games. This result already indicates that subjects either used a strategy that is inferior to the one used by the computer opponent or used the same strategy but failed on some of the steps during the implementation. The second option is more likely considering that the computer opponent used a very simple strategy. It is not feasible to simplify the strategy even further unless subjects were making completely random moves.

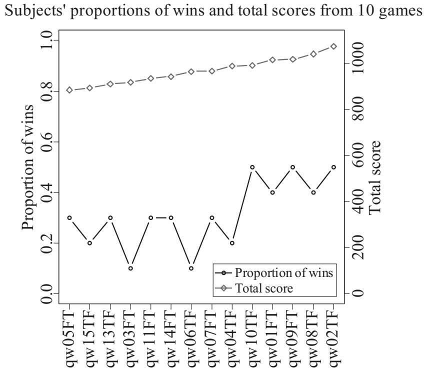


Figure 7.4: Proportions of wins and total number of points gathered by each subject during ten games.

One way repeated measures ANOVA applied to subjects' final scores shows no significant performance differences between the Hint and No-hint conditions $F(1, 13) < 1$. It is not surprising considering that scores gathered in individual games can differ significantly for the same subject. Analyses based on higher-granularity data described in following sessions show that there is indeed a difference between two conditions.

Figure 7.5 shows the mean scores gathered by subjects during individual turns in the Hint and No-hint conditions. In the figure, subjects were again sorted in increasing order of their total scores. The figure shows that there is a significant difference between subjects with low- and high- performance in No-hint condition. High performance subjects were able to gather at least two more points per turn than low performance subjects. Subjects' mean turn scores in No-hint condition significantly correlate with subjects' total scores, $r(12) = 0.87$, $p < 0.001$.

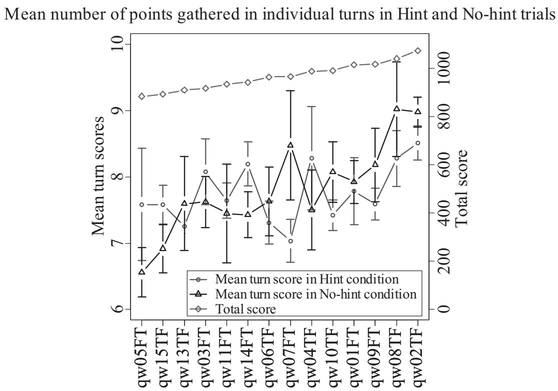


Figure 7.5: Mean number of points gathered by each subject in an individual turn in Hint and No-hint conditions. Subjects were sorted in ascending order of their total points.

Next, Figure 7.5 shows that providing a hint helped the subjects to increase turn scores. As a result, the difference between low and high performance subjects is less prominent in Hint condition than in No-hint condition. It is supported by insignificant correlation between subjects' mean turn scores in Hint condition and total scores, $r(12) = 0.47$, $p = 0.1$. This result suggests that both low- and high performance subjects may have been using the same strategy. Low performance subjects may have been more prone to making mistakes while implementing the strategy. However, providing hints may have helped them to decreases chances of mistakes. There is no significant effect of Hint and No-hint order on mean turn scores. Neither is there a significant effect of the order on subjects' average trial score.

Finally, given opportunities such as hints, subjects chose to maximize their turn scores. This fact suggests that subjects may have been using the same strategy as the computer opponent: maximize points gathered in each turn. The next section investigates further to confirm this assumption.

Subjects' moves

Previously, we have suggested that subjects used the same strategy as the computer opponent. The core of the strategy is to find an optimal combination of moves that results in the highest possible score for the turn. It is essentially a local maximum strategy because it tries to maximize local reward at individual turns rather than a global reward from a sequence of turns. If subjects indeed used the local maximum strategy then proportions of turns with the highest possible scores obtained should increase as subjects' total scores increase. Secondly, the same proportions should be higher in the Hint condition than in the No-hint condition. Those proportions were calculated for each subject and separately for Hint and No-hint conditions. The results are shown in Figure 7.6a. The data on the figure confirms that both previous assumptions are true.

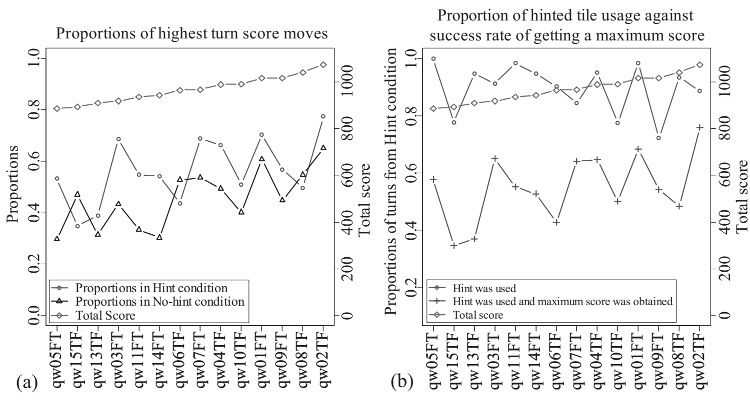


Figure 7.6: (a) Proportions of turns where subjects got highest possible scores. (b) Proportions of turns where subjects used tiles provided as hints in their moves and proportions of turns where subjects were able to get maximum scores while using the hinted tile.

Firstly, subjects with higher total scores were more successful at getting highest possible turn scores. However, correlation between subjects' proportions in Hint condition and total scores is not significant, $r(12) = 0.53$, $p = 0.053$. On the other hand, there is a significant correlation between proportions in No-hint condition and total scores, $r(12) = 0.74$, $p = 0.003$. Results of correlation tests again suggest that low and high performance subjects differ mostly in No-hint condition while presence of the hint helps to negate skill differences among subjects.

Secondly, proportions are generally higher in the Hint condition than in the No-hint condition. On average, subjects were able to find the optimal combinations in 56% (SE = 3%) of the turns in Hint condition compared to 45% (SE = 3%) in No-hint condition. This difference is significant according to one-way repeated measures ANOVA, $F(1, 13) = 11.28$, $p = 0.005$.

Because subjects were explicitly told that they can ignore hints, it is possible that subjects chose to do so most of the times. This could explain why subjects have relatively low success rate even in Hint condition. Figure 7.6b shows how often subjects used the hinted tile. The figure shows that subjects have chosen to use the hinted tile, on average, in 90% (SE = 2%) of the turns where hint was provided. However, the figure also shows that the success rate of obtaining the maximum turn score is still much lower (M = 55%, SE = 3%) even when the hint was used. It is likely that subjects often failed during one of the two steps described previously. The test of correlation between total scores and proportions of hint usage is not significant, $r(12) = -0.18$, $p = 0.54$. The insignificant correlation test indicates that usage of a hint by itself does not guarantee success in the getting the maximum turn score.

There is again no effect of the order of blocks on subjects' performance (Figure 7.7). Firstly, there is an overall learning effect from the first block to the second block. The learning effect is independent of order of two conditions. Secondly, subjects perform better at getting the highest score when a hint is given. The positive effect of the hint is quite significant. For example, the group of subjects who started with a Hint condition show lower performance in the second block with No-hint condition. Even the learning effect is not enough to compensate for the absence of a hint.

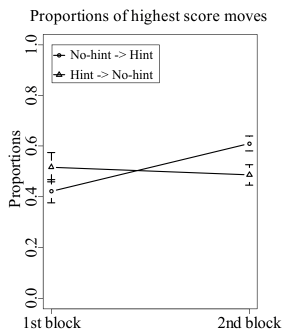


Figure 7.7: Effects of block order and trial type on proportions of turns where subjects got highest possible scores.

The local maximum strategy

The analyses from the last two sections support our hypothesis that subjects used the strategy of maximizing their turn scores. It is a simple strategy that can be described in two steps:

1. Identify an optimal combination of tiles that is likely to result in the highest possible score.
2. Identify an optimal combination of board positions for a chosen combination of tiles that is likely to result in the highest possible score.

Despite the simplicity, subjects were prone to making mistakes that prevented them from getting the maximum score as suggested by the low success rate in Figure 7.6a. Figure 7.6b leads us to conclude that, even given a hint, subjects still may fail to find a proper combination of tiles or a proper combination of board positions to put those tiles.

Optimal combination of tiles

To find out how often subjects failed during the first step, we have calculated the proportions of turns where subjects used a combination of tiles that could have resulted in the maximum possible points for that turn (it does not necessarily mean that subjects actually got maximum points).

Figure 7.8a shows that subjects were extremely good at finding a combination of tiles that could have resulted in the highest possible score for the turn. Whenever subjects used the hinted tiles, they were able to find the proper combination of tiles in 90% (*SE* = 2%) of the turns on average (the blue line in Figure 7.8a). Furthermore, even if no hint was provided or subjects chose to ignore the hint, subjects were able to find alternative combo that could have resulted in the highest possible score in 75% (*SE* = 2%) of all turns (the black line in Figure 7.8a).

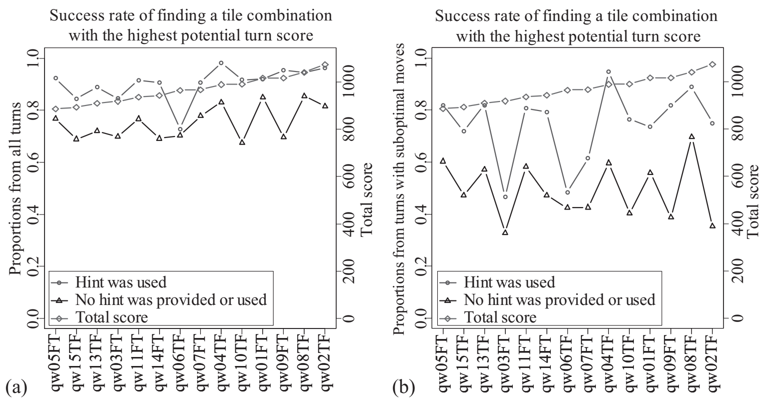


Figure 7.8: Proportions of turns where subjects were able to find the combinations of tiles that could have resulted in the highest possible turn scores. The proportions were calculated from (a) all turns and (b) only from those turns where subjects failed to get the maximum possible score.

If the same proportions are calculated using only those turns where subjects failed to get maximum possible score then the proportions are still quite high (Figure 7.8b). The average proportions are 74% (*SE* = 4%) and 49% (*SE* = 3%) in turns where hint was used and turns where hint was ignored or not provided at all.

None of the proportions shown in Figure 7.8 significantly correlated with subjects' total scores. It suggests that failures in the step 1 of the strategy cannot fully account for the performance differences among subjects. Furthermore, Figure 7.8b suggests that at least 49% - 74% of all failures to get maximum possible turn score should be due to the failure in the second steps of the strategy. Subjects do pick the right combinations of tiles, but not necessarily put them on the optimal spots on the board.

Optimal combination of board positions

Previous analyses suggest that an important process defining a subject's performance is how well she can find an optimal combination of board position that maximizes the number of gathered points. If this proposition is true then a subject with a lower total score should fail more than a subject with a higher total score during step 2 of the strategy. This can easily be tested by calculating the proportions of turns where subjects were able to find optimal board positions for the combination of tiles they have chosen in each turn. Figure 7.9 shows those proportions calculated for each subject. On average, subjects succeeded in finding an optimal combination of board positions in 67% (*SE* = 3%) of turns. The proportions are strongly correlated with subjects' total scores, $r(12) = 0.65$, $p = 0.01$. The significant correlation suggests that the ability to find an optimal combination of board positions is a strong indicative of subjects' performances.

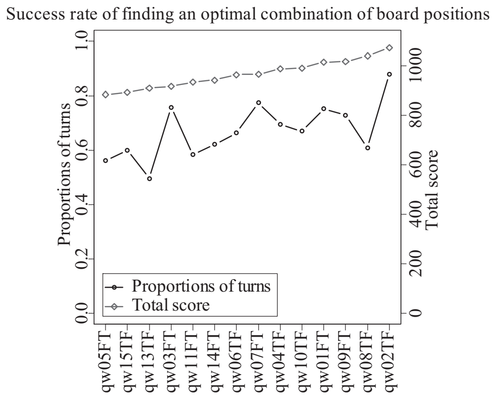


Figure 7.9: Proportions of turns where subjects were able to find optimal combinations of board positions that lead to the highest possible scores for a chosen combination of tiles.

Discussion

The earlier study involving Qwirkle (Mackey et al., 2011) focused on children of ages 5 to 9. Children in Mackey's study still exhibited significant improvements in general reasoning skills and spatial working memory. The result suggests that the strategy should be simple and intuitive enough to be used by children and yet sophisticated enough to involve cognitive resources ranging from vision to general problem solving. The local maximum strategy suggested by results of our study matches this profile. It is simple but reasonably effective strategy. More importantly, the successful use of this strategy is equally dependent on reasoning skills and visuo-spatial processing skills. This dependency explains why Mackey et al. observed improvement in spatial working memory of subjects.

Our results suggest that the reasoning skills play an important role during the first step of the strategy: identifying the combination of tiles with the highest potential score. However, it remains unclear how subjects decided which combination of tiles to choose. Ideally, it is possible to exhaustively search through all possible combinations of tiles and board positions, the same way the computer opponent does. However, it is highly unlikely that human subjects use exhaustive search due to time and cost inefficiency. It is more likely that subjects employ some form of probabilistic mechanism of making a near-optimal decision under uncertainty (Doya, 2008). Such stochastic mechanism could involve calculating a likelihood of obtaining the highest score given combination of tiles and current board state. Correspondingly, distribution and frequency of color and shape features on the board may affect the likelihood estimation. It was already shown in chapters 2 and 4 that frequency of both attentively and pre-attentively processed visual features can affect decision making (Nyamsuren & Taatgen, 2013b, 2013c). If it is indeed the case then it will be a direct evidence of visual system directly interfering with reasoning processes. It is also likely that the size of the tile combination plays an important role. More tiles are associated with a higher score. However, bigger tile combination also increases the effort required to find the optimal combination of cells. These and other factors (Busemeyer & Townsend, 1993), such as time pressure, are likely to be considered by subjects in calculating the likelihood of getting highest possible score.

The visuo-spatial processing skills are important during the second step of the strategy. Finding an optimal combination of positions is basically a problem of visual search with multiple targets and multiple partially matching distracters (Anderson, Fincham, Schneider, & Yang, 2012; Hong & Drury, 2002; Horowitz & Wofle, 2001). Targets are the board positions with the highest scores, and distracters are the positions with lower scores. Multiple-target visual search is a demanding process that requires combination of visual feature matching, spatial memory and higher-level mathematical reasoning. In addition to matching visual features, subjects need mathematical reasoning to compare gains across alternative board positions. As such, it seems that visual search is not purely visual and higher-level reasoning is invoked within its context. It is an opposite form of interaction between visual and reasoning systems than the one may be used during the choice of tile combination. Finally, given the complexity of such visual search it is not surprising that subjects fail often during this step. It also explains why subjects are not as good as the computer opponent despite the simplicity of the common strategy. The computer opponent does the perfect visual search.

If we take an overall view of the subjects' strategy in light of the earlier discussion then it is not as simple as having two steps, one with reasoning and one with visual search. Instead, it seems that the visual and reasoning systems are deeply intertwined. On the one hand, reasoning outcome is highly dependent on both attentive and pre-attentive visual knowledge. On the other hand, visual processing requires capability of reasoning on concepts more higher level than visual features.

Conclusion

In this chapter, we have described the preliminary study of human behavior and strategy in Qwirkle. The next step is to create a cognitive model that can (1) provide empirical validations of the assumptions and hypothesis proposed here, and (2) test whether our explanation is compatible with wider theory of human cognition through the use of ACT-R cognitive architecture (Anderson, 2007).

This study suggests that three types of cognitive resources are most important for this modeling effort. Firstly, declarative memory is necessary to store task specific and general knowledge and rules. Next, the study emphasizes the importance of a visual system as a medium of gathering and processing real-time knowledge. Considering the complexity of the required visual processing the use of the Pre-Attentive and Attentive Vision module (Nyamsuren & Taatgen, 2013a) is required to gain access to such cognitive resources as iconic memory and short-term visual memory. The final type of cognitive resources is fluid reasoning that is capable of integrating declarative and visual knowledge to solve the problem of playing the game. Catell (1987) proposed that fluid reasoning serves as a scaffold that allows us to form and acquire new cognitive skills and knowledge. Halford, Wilson and Phillips (1998) proposed relational integration of fluid reasoning, the ability to jointly consider distinct relationships between stimuli. The Human Reasoning Module (Nyamsuren & Taatgen, 2013d) was developed with the same principle in mind as fluid reasoning. The HRM can serve as a scaffold for deriving new knowledge by combining existing knowledge in the declarative and visual systems.

We have not really touched upon subjects' eye-movement data in this chapter. The future plans definitely include paying more attention to eye-movement data, especially as a means of estimating the cognitive model's fit to subjects' behavior.

CHAPTER 8

Summary & Concluding Remarks

Summary

The Human Reasoning Module and the Pre-Attentive and Attentive Vision module are both theory and tool.

PAAV significantly extends the ACT-R's canonical vision module. PAAV's implementation is based on a theory of sequential attention shifts guided by low-level information pre-attentively captured via parallel visual processes. Parallel processes do not have infinite capacity and are limited by acuity properties of different visual feature dimensions. PAAV supports a more sophisticated attentional guidance that includes both bottom-up and top-down components. Bottom-up guidance is governed by inherent salient properties of a scene rather than explicit criteria provided in an attention shift request. Similarly, top-down guidance is integrated into an overall saliency map rather than being treated as a separate attentional guidance system. PAAV further integrates top-down and bottom-up processes by proposing an implicit activation network between visual and declarative memories. These improvements enable ACT-R to model tasks ranging from canonical visual search to complex problem solving tasks where real-time visual information is essential part of decision-making.

The motivations for developing the HRM module are more complicated than those behind PAAV. On the one hand, the HRM was written to extend theory and functionality of existing modules. Firstly, it seeks to explain how single architecture can exhibit several different forms of reasoning ranging from strictly classical logics to pragmatic reasoning, from induction to deduction, and from deterministic reasoning to scholastic Bayesian reasoning. It is an ambitious goal that requires further research beyond ones described in this dissertation. On the other hand, the HRM was written to add constraints to current modeling practices. Unrestricted definition of chunk types and production rules are the two major issues currently plaguing the ACT-R modeling community. Among the negative consequences of this practice is a hindered reusability of models that makes it unnecessarily difficult to check consistency among related cognitive theories or falsify incorrect ones among competing theories. The HRM proposes a very specific construct of knowledge organization on top of chunk-based format and set of production rules that can be used in a wide range of tasks. By proposing standardized task-general production rules, the HRM makes a good trade-off between a constraint-free architecture and an architecture that is functionally crippled by too restrictive constraints.

Myths of Cognitive Modeling

In the following sections, I would like to address, firstly, some of the criticisms given toward cognitive modeling and, secondly, some of the open questions and issues that should be resolved in the roadmap toward the unified theory of human mind.

86 bn parameters for 86 bn neurons?

A common criticism applied toward cognitive architectures is that they do not provide sufficient restrictions on the possible parameter space. The criticism applies to the HRM (Chapter 6) and PAAV (Chapter 4) as these two modules add additional sets of parameters to the architecture.

Definitely, a model of a specific task can have one too many parameters. However, can we apply the same logic when it comes to a cognitive architecture that is designed to simulate something as complex as human brain? According to the recent estimate, a human brain consists of 86 billion neurons (Azevedo, Carvalho, Grinberg, Farfel, Ferretti, Leite, Jacob-Filho, Lent, Herculano-Houzel, 2009). Does that mean that we need 86 billion parameters in the cognitive architecture? It is a rhetorical question given to emphasize the vast complexity of a human brain. Furthermore, the estimation above does not include non-neuronal cells that account for another 84 billion cells. According to embodied cognition (Ballard, Hayhoe, Pook, & Rao, 1997; Clark, 1997), the human mind is not only a product of the brain, but of a physical body as well. Correspondingly, how many additional parameters are required to account for the motor system, the human eye, etc.? The human mind is by far the most complex computational system we are familiar with. One can fit several different models to explain human behavior in a distinct task. However, I challenge anyone to propose a complete cognitive theory of human mind that is too general and has too many parameters.

Finding a unified theory of the human mind is not the same problem as finding an optimal solution with a minimal set of parameters. The two problems are totally different in both scale and purpose. The latter one is an optimization problem in a system with clearly defined properties and constraints (Boyd & Vandenberghe, 2004). The former one is an exploratory research where we are still trying to define properties and constraints of the human cognitive system. Therefore, having a cognitive architecture with a wide parameter space and many alternative models of the same task is not an issue in itself. Competing models can be falsified and the parameter space can be narrowed down, as more evidence is uncovered. Such an approach is certainly easier than trying to revise an entire cognitive architecture upon realization that it was too simplified or restricted to be able to handle tasks more complex than the ones the architecture was originally designed for. The real issue arises when the cognitive architecture remains static even in light of new evidence. The next section elaborates more on this problem.

So how many parameters is enough? I do not think anyone can put forward even an approximate number, but ACT-R certainly does not have enough parameters to be able to fully simulate a human mind.

Breaking the seal on ACT-R

My argument in the preceding section is based on the very important assumption that the cognitive architecture is revised and refined over time as more knowledge about human mind is uncovered. Unfortunately, the assumption does not hold. The Figure 8.1 shows the timeline of ACT-R releases. The last one was in 2005 with the release of ACT-R 6.0. Since then there have been small changes to various components of ACT-R. However, it has been more than eight year without any new major release.

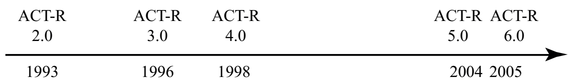


Figure 8.1: Timeline of ACT-R releases.

There are several issues directly connected with the lack of advancements in development of cognitive architecture. I have mentioned one of them in Chapter 1. In today's studies, the cognitive modeling efforts are limited to models of individual tasks or groups of similar tasks (Busemeyer & Diederich, 2010; Farkaš, 2012). The cognitive architectures are treated as a modeling tools rather than theories of human cognition. As such, there is little incentive for researchers to contribute to the architecture itself. ACT-R is both a theory and a tool, and as any theory it should be refined, expanded and corrected when necessary.

Next, there is a certain resistance to making changes to cognitive architectures. It is caused by two issues. Firstly, it is an assumption that architecture should remain static and fixed (Newell, 1990). It is indeed a very important criterion for creating plausible and consistent models. However, the assumption applies only when the architecture provides all necessary and sufficient constituent components (Goertz & Starr, 2003; Mackie, 1965) of the system it represent. It is not even remotely true for cognitive architectures: the requirement of sufficiency is not fulfilled. There is no rational basis for claiming plausibility and consistency of models without meeting architectural requirements. Secondly, there is a certain misunderstanding of a structure of cognitive architecture, at least in case of ACT-R. One of the peer-reviewers of the article about Human Reasoning Module (Chapter 6; Nyamsuren & Taatgen, 2013) was concerned about proliferation of modules in ACT-R architecture. Modules in ACT-R serve two purposes. Firstly, modules represent specific regions of human brain that are responsible for particular types of cognitive processes such as declarative memory, visual processing, motor control, etc. (Anderson, 2007; Borst & Anderson, 2013; Borst, Taatgen, & Van Rijn, 2011). The core modules of ACT-R, such as Declarative and Vision, are examples of modules that represent distinct parts of a brain. On the other hand, new modules are convenient tools for expanding the functionalities of already existing modules. The Pre-Attentive and Attentive Vision module (Chapter 3) expands ACT-R's default vision module. The Human Reasoning Module expands (Chapter 6) ACT-R's procedural and declarative modules. Proliferation of modules that claim new types of cognitive processes should be perceived with caution. However, we very much need proliferation of modules that expand and improve existing ones.

Exploratory experiments, confirmatory models and explanatory architectures

This dissertation is an attempt to have a balanced approach to the study of human cognition. On the one end of the scale, we need exploratory experiments to propose new psychological theories or to draw a connection between existing ones. On the other end of the scale, we need cognitive modeling for two major purposes: empirically confirm proposed theories and advance the understanding of the human mind as a single unified architecture.

Exploratory experiments

A study of SET described in Chapter 2 is a perfect example of an exploratory experiment that not only tries to explain human behavior in a particular task, but also seeks to connect separate theories of cognition. Firstly, it emphasized importance of bottom-up visual processes in decision-making. What we see influences our decision at a subconscious level. As a result, we have proposed a theory of implicit connection between declarative and visual memories. It is a connection based on activations spread among similar items in two memories. Later in Chapter 4, the theory helped to explain an interesting phenomenon of scene perception in extremely short durations of time. The study in Chapter 5 further emphases the influence of visual processes in problem-solving tasks to the extent that such influence can affect subjects' behavior and performance.

Secondly, the SET study reveals that the line between top-down and bottom-up cognitive processes is not as clear-cut as it is perceived to be. Novice players prefer perceptual similarity while skilled players prefer top-down rules. There is reasoning at two levels: visual and declarative. Reasoning at the visual level is fast, but ultimately limited to similarity-based comparisons. Reasoning at the declarative level is slow, but powerful enough to consider distinct relationships between different elements via inference rules in declarative memory. These finding later become part of initial set of theories upon which PAAV (Chapter 4) and the HRM (Chapter 6) are built. The experiment based on Qwirkle (Chapter 7) provides further insight into the interplay between our reasoning and visual systems. It is peculiar that visual search skill is a factor that defines subject's performance in the task traditionally viewed as a mathematical reasoning task. Is it possible that the visual system is much smarter than we suspect it to be? Can we attribute to the visual system the ability to process information on a semantic level, such as basic numerical reasoning? It is a "*rebellious thought*" from a perspective of classical logical reasoning. However, if it is true then there is no such thing as top-down reasoning. There may not be two separate systems such as System 1 and System 2. Instead, the single reasoning system may operate directly on knowledge forms at both top-down and bottom-up layers.

Confirmatory models

In addition to box-and-arrow models, a special emphasis was put on the necessity of empirical validations of the cognitive theories using computational models. Computational models are necessary because of several reasons. Firstly, box-and-arrow models may fail to consider obscure but important aspects of the experimental tasks may have significant effect on long run. Secondly, subjects in tasks close to real-world problems often exhibit complex non-linear behaviors. Solutions for such problems cannot be found via mathematical or analytical means.

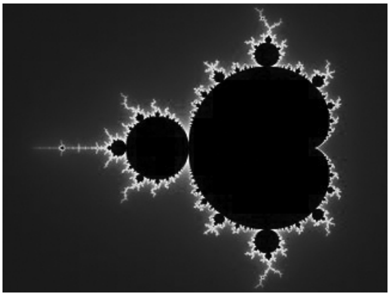


Figure 8.2: Visualization of the Mandelbrot set as a fractal shape with its amazing self-similarity: if magnified at the boundaries then smaller versions of the shape can be found at the arbitrarily small scales.

The necessity to consider the above two points can be visually demonstrated by taking the example of a quadratic polynomial system such as $z_{n+1} = z_n^2 + c$. It is the equation for the Mandelbrot set, the first fractal extensively studied in mathematics using computer simulations. On the surface, the equation is very simple. However, it has a non-linear and recursive nature that makes exploring the dynamics of the system over many recursions an extremely challenging task. The variables in the equation are complex numbers, and the Mandelbrot set is a plot in the plane of the complex parameter *c*. As a result, it generates an extremely complex geometrical structure show in Figure 8.2. No box-and-arrow model will be able to describe the dynamics of the Mandelbrot set. In fact, mathematics itself was not an optimal tool for exploring it, and only advent of computer visualization allowed mathematicians to comprehend the complexity of the Mandelbrot set. The human mind shares many properties with fractals and other complex systems[6]. The mind is non-linear, recursive and stochastic. Similar to the case of the Mandelbrot set, it is highly unlikely that we will be able to understand true dynamics of human mind even in simplest tasks without computer modeling.

The competitive parallelism of bottom-up similarity-based comparison and top-down rule-based comparison described in Chapter 2 is a complex dynamic process. It is difficult to account for the evolution of declarative activations that are responsible for the transition from one form of comparison to another using only a box-and-arrow model. Confirmation that such cognitive process indeed can explain differences between novice and expert players requires empirical evidence from a computational model. The same argument applies to models of Most Abundant Value task in Chapter 4 and effects of visual presentation styles on SET performance described in Chapter 5. Perhaps, it is possible to construct a box-and-arrow model of the domino effect described in Chapter 5 where changes in one cognitive process start a chain of changes in other processes. However, no box-and-arrow model or mathematical analytical solution can predict the eye movements.

Roberts (1993) suggested that two theories of human reasoning, the mental models and mental logics, are not mutually exclusive. However plausible his theory is, there is a lack of explanation how exactly the two theories should merge and interplay. The spatial reasoning model described in Chapter 6 provided an empirical confirmation that the two theories can coexist in a single reasoning system. The model is not concrete evidence that two forms of reasoning indeed coexist in human reasoning system. However, the model provides the confirmation that coexistence is possible. Confirmation of possibility or impossibility of proposed cognitive theories within constraints of cognitive architectures is the value of cognitive models that is often overseen.

Explanatory architectures

Ideally, any computational model should rely solely on the cognitive architecture to provide sufficient functionality to simulate cognitive processes. In practice however, the modelers more often than not have to write custom pieces of code as an attachment to the model to mitigate limitations of the architecture. This practice has significant drawbacks. The custom code is neither part of the model or architecture. At least in ACT-R, the important constraints of the architecture do not apply to the custom code. Absence of constraints raises the concern of plausibility of the custom code as well as the cognitive theory behind it. As a result, the more custom code there is the less feasible the model becomes. It is an import issue that should be addressed.

For the reason described above, we have adopted an approach separating the experimental task into task-specific and task-general components. The task-general components were

---

[6] Although never mentioned by neuroscientists, neurons in a brain are organized into a fractal structure (Fernandez, Bolea, Ortega, & Louis, 1999; Kiselev, Hahn, & Auer, 2003).

modeled as validated extensions to the architecture, fully applying any of its constraints. Task-specific components were addressed by cognitive models. Such an approach minimizes the amount of custom code that needs to be written and has a positive contribution to the development of the architecture. Furthermore, the major explanatory power of task-general elements comes from the cognitive architecture rather than custom code or task-specific model.

Perhaps the best example of this approach is the models of picture and word SET described in Chapter 5. The two models are essentially the same. They differ only in the set of meta-cognitive production rules that are responsible for handling different presentations styles in the two types of SET. However, the two models show radically different behavior. This difference is accounted for by the PAAV module (Chapter 3), more specifically by its visual acuity mechanisms and iconic memory. This kind of cross validation add additional layer of feasibility to the models and underlying psychological theories. Such modeling would have been impossible with ACT-R default vision module. PAAV's explanatory power was supported by other studies as well: the task of decision-making based on scene gist perception (Chapter 4) and three types of visual search (Chapter 3).

Although not as extensively tested as the PAAV module, the Human Reasoning Module is also promising to be in right direction toward explaining task-general reasoning capabilities. The module provides its own set of task-general production rules for reasoning. All three models described in Chapter 6 use this same set of production rules. It saves the modeler time that would have been spent writing reasoning production rules for individual models. More importantly, it also provides a single vocabulary among different models again shifting explanatory power from models to the architecture.

Are we there yet?

This research started with an optimistic question of how we can model a believable computer opponent in games. This is the last section of this dissertation, and, naturally, a question arises: Are we done yet? Can we finally build a believable computer opponent? In the simplest case, a believable model should produce a data that is literally indistinguishable from human data even for an expert eye. We are not there yet. We have only scratched the surface of human cognition necessary to play even the simpler computer games. The model of SET initially introduced in Chapter 2 was revised in Chapter 4. An entire vision module was developed to support the model. However, it is still a largely simplification of human behavior in SET. For example, the current SET model treats working memory (WM) as a problem state. Is it a realistic implementation of WM? What if there is no single cohesive WM, but rather a collection of memories highly specialized for specific categories of tasks. How various proposed forms of working memories, such as visual short-term memory, spatial memory, mental models and problems state, fit together in a single system? On the other hand, I haven't even begun to explore complex low-level visual processing such as contour recognition or Gestalt principles that allow grouping of shapes into a single image of a card.

These and other issues are difficult research questions that need to be answered through a long-term methodological research. The key for eventually achieving the original goal of this research is continuous refinement and improvement of cognitive theories as well as cognitive modeling tools. Chapter 7 described theoretical findings from Qwirkle experiment. Modeling efforts are on the agenda of future work. During such efforts, necessary changes should and will be made to the HRM and the PAAV modules. Such refinements and improvements are necessary in light of new experimental evidence. Such modeling approach is slow and challenging, but necessary in the end.

Samenvatting

Het oorspronkelijke doel van mijn promotieonderzoek was het bestuderen van technieken en methoden om geloofwaardige kunstmatige tegenstanders voor computerspellen te ontwerpen met behulp van de cognitieve architectuur ACT-R. Daarvoor heb ik een eerste onderzoek uitgevoerd naar de manieren waarop mensen beslissingen nemen en redeneren wanneer ze een computerversie van het bordspel SET spelen (hoofdstuk 2). Uit een uitvoerige analyse van de gegevens over oogbewegingen en andere gedragsmaten van de deelnemers bleek dat de visuele waarneming rechtstreeks kan beïnvloeden hoe beslissingen worden genomen. Het onderzoek liet echter ook zien dat de bestaande cognitieve architecturen, waaronder ACT-R, niet geavanceerd genoeg zijn om complexe taken zoals SET te modelleren. Het onderzoek gaf duidelijk aan dat er in ACT-R twee belangrijke componenten ontbraken. Naar aanleiding van deze resultaten is het oorspronkelijke doel opnieuw geformuleerd als het onderzoeken van de minimale componenten van een cognitieve architectuur die nodig zijn om aannemelijke cognitieve modellen voor complexe taken als computerspellen te ontwikkelen.

De specifiekere doelen waren (1) het opstellen van een samenhangend theoretisch kader voor top-down controle (het beslissen en plannen) en perceptuele bottom-up processen (controle van de visuele aandacht), en (2) het ontwikkelen van software-instrumenten die de bovenstaande processen onafhankelijk van specifieke taakparadigma's kunnen simuleren. Ik heb gedragsexperimenten uitgevoerd om de processen te onderzoeken die een rol spelen bij de visuele aandachtscontrole tijdens een probleemtaak. Ik heb twee modelleringsinstrumenten ontwikkeld waarin de theoretische bevindingen uit mijn onderzoek zijn opgenomen: de Pre-Attentive and Attentive Vision (PAAV)-module en de Human Reasoning Module (HRM). De PAAV-module is een implementatie van theorieën over de menselijke visuele waarneming (waaronder visuele geheugens en op contrasten gebaseerde saliency maps), waarin wordt gesteld dat naast parallelle preattentieve processen ook sequentiële attentieve visuele processen optreden (hoofdstuk 3). De HRM biedt een samenhangend kader voor het menselijk redeneren. Deze module stelt dat verschillende aspecten van het menselijk redeneren kunnen worden verklaard door deterministische inferentie van stochastische kennis (hoofdstuk 6). De PAAV-module en de HRM zijn onderdeel van de cognitieve architectuur ACT-R en kunnen worden gebruikt om het menselijke gedrag te modelleren voor een breed scala aan taken. Beide instrumenten zijn gevalideerd met experimentele gegevens. De broncodes zijn voor iedereen toegankelijk via deze link: http://www.ai.rug.nl/~n_egii/models/.

Het menselijke zien

Allereerst is voor elke complexe taak waarover op basis van realtime kennis van de omgeving een beslissing moet worden genomen, een betrouwbaar middel nodig om informatie te verzamelen. Van de vijf zintuigen is het visuele systeem ongetwijfeld het belangrijkste middel om die informatie te verzamelen. De meeste moderne technologieën zijn ontworpen rondom visuele input als belangrijkste informatiebron voor de gebruiker. Het onderzoek naar efficiënte communicatie van informatie draait grotendeels om de visualisatie van informatie. Een van de kernmodules van ACT-R is de visuele module. Helaas biedt deze module slechts een basale implementatie van het menselijke visuele systeem.

De controle van de visuele aandacht is een zeer complex proces dat uit zowel top-down als bottom-up componenten bestaat (Orban, Fiser, Aslin & Lengyel, 2008). Aan de ene kant wordt de visuele aandacht gecontroleerd door inherente bottom-up kenmerken van een visuele scène, zoals de op contrasten gebaseerde opvallendheid (saliency) van de samenstellende delen. Aan de andere kant wordt de visuele aandacht ook gecontroleerd door top-down componenten, zoals het directe doel en een context die door eerdere ervaringen is bepaald. De visuele standaardmodule biedt geen ondersteuning voor de aandachtscontrole via bottom-up processen. Er ontbreken ook diverse andere essentiële functionaliteiten, zoals visuele lange- en kortetermijngeheugens, de omschrijving van visuele objecten aan de hand van verschillende soorten kenmerken, en een voorstellingsvermogen.

De Pre-Attentive and Attentive Vision (PAAV)-module (hoofdstuk 3) is een uitbreiding van de visuele standaardmodule van ACT-R. Zoals de naam al aangeeft, voegt de PAAV-module een aanzienlijke hoeveelheid preattentieve functionaliteit toe aan het visuele systeem van ACT-R. In feite is de PAAV-module een implementatie van een aantal goed onderzochte theorieën over de menselijke visuele waarneming, over onderwerpen die uiteenlopen van het visuele geheugen tot op contrasten gebaseerde saliency maps die de visuele aandacht controleren (Itti, Koch & Niebur, 1998). De implementatie van de PAAV-module berust op een theorie over sequentiële aandachtsverschuivingen die worden gecontroleerd door informatie van een lagere orde, die preattentief via parallelle visuele processen wordt waargenomen. De capaciteit van deze parallelle processen is niet oneindig. Ze worden beperkt door de scherpte waarmee de verschillende soorten visuele kenmerken worden waargenomen. De PAAV-module ondersteunt een geavanceerdere aandachtscontrole die zowel bottom-up als top-down componenten omvat. De bottom-up controle wordt door inherente opvallende kenmerken van een scène aangestuurd, in plaats van door expliciete criteria die in een verzoek om een aandachtsverschuiving worden verstrekt. Op dezelfde wijze is de top-down controle geïntegreerd in een algemene saliency map, en wordt deze niet als een afzonderlijk systeem voor de aandachtscontrole behandeld. De PAAV-module integreert de top-down en bottom-up processen nog verder met een impliciet activeringsnetwerk tussen het visuele en declaratieve geheugen. Door deze verbeteringen kan ACT-R taken modelleren die uiteenlopen van standaard visuele zoektaken tot complexe probleemtaken waarbij realtime visuele informatie essentieel is voor het nemen van beslissingen.

De PAAV-module is met succes gebruikt om menselijk gedrag te modelleren voor een visuele zoektaak voor kenmerken en conjuncties, waarbij een doelstimulus moest worden gevonden tussen een aantal distractors. Ook werd de module succesvol toegepast voor een vergelijkende visuele zoektaak, waarbij een verschil moest worden gevonden tussen twee sets stimuli die verder hetzelfde waren. Dit waren eenvoudige taken die werden gebruikt om de basale componenten van de PAAV-module te valideren. Om het modellerende en verklarende vermogen van de PAAV-module verder te bestuderen, heb ik diverse eye-trackingonderzoeken uitgevoerd waarin complexe besluitvorming en redenering op basis van realtime visuele gegevens een rol speelden. Uit deze onderzoeken bleek dat modellen die op de PAAV-module waren gebaseerd, menselijk gedrag met succes konden nabootsen. Zo heb ik in mijn onderzoek van hoofdstuk 5 een op de PAAV-module gebaseerd model gebruikt om te verklaren hoe het beslissingsproces verandert als dezelfde taak op verschillende manieren visueel wordt gepresenteerd. In hoofdstuk 4 liet ik effecten zien die aanvankelijk in tegenspraak leken te zijn met de algemeen aanvaarde theorie over aandachtscontrole op basis van op contrasten gebaseerde saliency maps. De connectionistische wijze waarop de PAAV-module het visuele en declaratieve geheugen benadert, helpt echter te verklaren hoe de top-down controle de bottom-up aandachtscontrole op basis van opvallendheid rechtstreeks kan opheffen.

Het menselijke redeneren

Iedere probleemtaak vereist een zekere mate van redeneren. Dat kan elke vorm van redeneren zijn: redeneren door analogie, redeneren op basis van regels of eenvoudigweg op basis van associaties. Hoewel de afzonderlijke stappen in het redeneerproces aan de specifieke context van een taak kunnen worden gekoppeld, is ons algemene vermogen om te redeneren een fundamenteel proces dat niet afhankelijk van een specifieke taak is. Redeneren op basis van analogie wordt bijvoorbeeld als een fundamenteel onderdeel van de menselijke cognitie beschouwd (Gust, Krumnack, Kühnberger & Schwering, 2008). Cummins (1996a, 1996b) stelde dat deontisch redeneren een aangeboren eigenschap is. Daarentegen zijn er ook bewijzen dat de kennisverwerving door jonge kinderen grotendeels afhangt van aangeboren concepten en principes (Baillargeon, 2008). In het specifieke geval van ACT-R zou er een algemene, taakonafhankelijke set productieregels moeten zijn die op basis van een gegeven context schematische regels voor het redeneren biedt. De regels in SET (hoofdstuk 2) schrijven bijvoorbeeld voor dat als twee kaarten van een mogelijke set groen en blauw zijn, de derde kaart rood moet zijn om een geldige set te vormen. Deze regel kan worden geschreven als (Blauw, Groen) => (Rood). Dit is een taakspecifieke regel. Het vermogen om op basis van deze regel te redeneren moet echter taakonafhankelijk zijn. ACT-R zou moeten beschikken over taakonafhankelijke kennis over de conjunctie van concepten. ACT-R zou ook over de kennis moeten beschikken dat de conjunctie van bepaalde concepten kan duiden op een ander concept.

De Human Reasoning Module (HRM) uit hoofdstuk 6 voegt essentiële declaratieve en procedurele kennis aan ACT-R toe, waarmee deze op basis van taakspecifieke instructies kan redeneren. De HRM introduceert concept als expliciet begrip. De module weet ook dat afzonderlijke concepten gecombineerd kunnen worden tot complexere statements. Deze statements kunnen op hun beurt weer samengevoegd worden tot declaratieve regels die gebruikt kunnen worden om taakspecifieke instructies te coderen. De concepten, statements en regels op basis van statements vormen het declaratieve deel van de kennis van de HRM. De procedurele kennis van HRM bestaat uit een set taakonafhankelijke productieregels die beschrijven hoe declaratieve en andere vormen van kennis kunnen worden gebruikt om te redeneren. Als de HRM wordt toegepast en in het declaratieve geheugen de juiste set instructies voor een taak wordt gegeven, dan hoeft degene die het model maakt in het ideale geval slechts enkele taakspecifieke productieregels te schrijven die vooral verantwoordelijk zijn voor de metacontrole.

De HRM streeft ernaar te verklaren hoe één enkele architectuur verschillende vormen van redeneren kan vertonen, van strikt klassieke logica tot pragmatisch redeneren, van inductie tot deductie en van deterministisch redeneren tot scholastisch Bayesiaans redeneren. Dat is een ambitieus doel, waarvoor meer onderzoek is vereist dan binnen het kader van mijn promotieonderzoek mogelijk is. De voorlopige resultaten die in hoofdstuk 6 worden beschreven, zijn echter veelbelovend. Het op de HRM gebaseerde model voor causale redeneertaken helpt bijvoorbeeld verklaren waarom iemand een pragmatische inductieve redenering blijft gebruiken, zelfs wanneer hij schijnbaar deductieve argumenten krijgt aangereikt, een effect dat Cummins in experimenten heeft waargenomen (1991, 1995). Een ander op de HRM gebaseerd model voor blicketdetectortaken (waarin wordt bepaald welke stimuli een speciaal apparaat in werking zetten) laat zien hoe een verzameling deterministische symbolische regels kan leiden tot Bayesachtige probabilistische inferentie als gevolg van de onzekerheid waarmee informatie uit het declaratieve geheugen wordt opgehaald.

Een laatste kenmerk waarmee de HRM zich onderscheidt van de traditionele opvattingen over de menselijke mentale logica (Rips, 1983), is dat het redeneren in de HRM niet uitsluitend een top-down proces is. De feiten en bewijzen die nodig zijn voor het redeneren, kunnen tussentijds uit andere informatiebronnen dan het declaratieve geheugen worden gehaald. Een voorbeeld van zo'n alternatieve bron is het visuele geheugen in de PAAV-module. Als iemand mij vraagt welke positie een vork ten opzichte van een bord heeft en ik ze beide voor me zie liggen, dan hoef ik geen propositioneel statement uit mijn declaratieve geheugen op te halen. Ik haal dan gewoon onmiddellijk de locatie-‘cue’ van de vork uit mijn ruimtelijk geheugen. Op dezelfde wijze kan de PAAV-module tijdens het redeneren gebruikmaken van bottom-up informatie in het visuele geheugen. De module kan bijvoorbeeld ruwe ruimtelijke informatie uit het visuele geheugen halen en deze direct vertalen naar een declaratief statement.

Het menselijke zien en redeneren verbonden

De combinatie van de PAAV-module en de HRM binnen de ACT-R-architectuur levert een krachtig instrument op waarmee het menselijke gedrag voor de meest eenvoudige taken tot aan complexe probleemoplossing kan worden gemodelleerd. Hoofdstuk 6 beschrijft een model voor een ruimtelijke redeneertaak dat uiterst afhankelijk is van zowel de HRM als de PAAV-module. Met behulp van dit model probeer ik de verbanden te verklaren tussen twee concurrerende theorieën over deductief redeneren: mentale modellen (Johnson-Laird, 1983) en mentale logica (Rips, 1983).

Hoofdstuk 7 beschrijft een onvolledig maar veelbelovend en interessant onderzoek naar de strategie van spelers bij het spel Qwirkle. Bij dit spel moeten de spelers uitvoerige vergelijkingen maken en alternatieve zetten tegen elkaar afwegen om een zo hoog mogelijke score te behalen. Tegelijkertijd bevat het spel sterk perceptuele elementen, aangezien het redeneren voor een groot deel berust op het al dan niet combineren van bepaalde kleuren en vormen. Hoewel we nog geen cognitief model hebben, wijst de analyse van de gegevens over mensen er al wel op dat de prestaties sterk afhangen van zowel het visuele systeem als het redeneervermogen dat zorgt voor het nemen van een optimale beslissing. Een aannemelijk model voor een Qwirklespeler zal daarom zowel de PAAV-module als de HRM moeten bevatten.

Ik verwacht dat de combinatie van de HRM en de PAAV-module een krachtig instrument zal bieden voor het opzetten van aannemelijke en praktische cognitieve modellen. Het laatste criterium ten aanzien van het praktische aspect is net zo belangrijk als het eerste. Naast de gebruikelijke toepassing om modellen op te zetten die enkel experimentele bevindingen bevestigen, biedt ACT-R veel mogelijkheden voor praktische toepassingen. ACT-R kan als softwareplatform worden gebruikt voor cognitief plausibele intelligente systemen, zoals cognitieve robots of intelligente onderwijssystemen. Ik doel dan niet op eigendomsmatige versies van ACT-R, zoals de versie achter de Cognitive Tutor van John Anderson, of ACT-R/E van het Amerikaanse Naval Research Laboratory. Ik zie juist een gemeenschappelijke opensourceversie van ACT-R voor me, die net zo geavanceerd is als de eigendomsmatige cognitieve architecturen. Zowel de HRM als de PAAV-module zijn op mijn homepagina vrij beschikbaar onder de GNU General Public License. Ik moedig iedereen aan om de broncodes te downloaden, ermee te experimenteren, en zelfs eigen versies van de PAAV-module of HRM te ontwikkelen. Ook doe ik een oproep aan iedereen die geïnteresseerd is in de gezamenlijke ontwikkeling van geavanceerde opensourceversies van cognitieve architecturen voor praktische doeleinden.

Appendix A

Card game of SET

The SET[7] 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 A.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 A.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 the number of attributes that are shared between the two. For example, the cards in a level 1 set have a perceptual similarity of three among each other as they have three attributes with identical values. Cards in a level 4 set have a perceptual similarity of 0 because all attribute values are different.

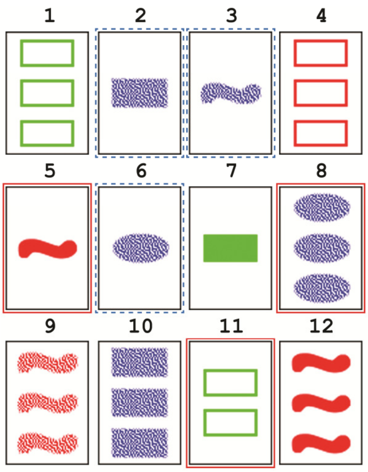


Figure A.1: An example array of 12 cards. The cards with the solid highlight form a level 4 set (all attributes are different), and cards with dashed highlight form a level 1 set (shape is different, and all other attributes are the same).

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.

[7] SET is a game by Set Enterprises (www.setgame.com).

Appendix B

ACT-R cognitive architecture

ACT-R[8] has a modular organization where each module is dedicated to a distinct type of cognitive resources. ACT-R consists of several modules, such as Vision module for handling visual processing, Declarative module for simulating declarative memory, and Goal module for tracking a model's state and objectives. Figure B.1 shows six core modules of ACT-R architecture as described in Anderson (2007). The modules mostly communicate with each other via Procedural module that allows a modeler to write task specific production rules. However, in limited cases, modules can also spread activation to other modules simulating low-level sub-symbolic cognitive processes. ACT-R procedural system is completely serial, and, therefore, only one production rule can be called at the time.

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 most important for this study are described next.

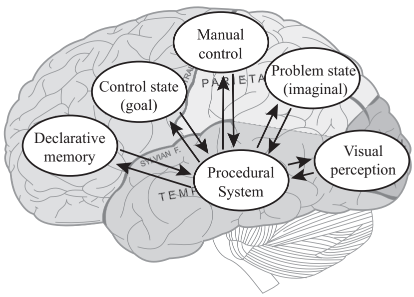


Figure B.1: ACT-R architecture from the perspective of six core modules.

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

Declarative module is an implementation of human long-term declarative memory. Every chunk that has been cleared from any buffer is stored in declarative module and can be retrieved again. The declarative module can retrieve only one chunk at a time, which is stored in the module's buffer. Each chunk in declarative memory has a base-level activation value, which represents frequency and recency of use (e.g., Anderson & Schooler, 1991). A chunk's activation in declarative memory can also be influenced by chunks contained in buffers at the time of retrieval via a spreading activation mechanism. Based on activation, the module computes the probability and time cost of retrieving a chunk from memory.

Lastly, there is an imaginal or problem state module that serves as a working memory. This module is unique as it can create new chunks that are neither perceived in the

[8] ACT-R stands for Adaptive Control of Thought-Rational

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 other two extra modules are Threaded Cognition (Salvucci & Taatgen, 2008) and Base-Level Inhibition (Lebiere & Best, 2011). With Threaded Cognition, we assume that there are two separate and parallel meta-controls governing the overall top-down strategy and the bottom-up visual attention shifts respectively. Lastly, the Base-Level Inhibition module provides a short-term activation inhibition of items in declarative memory. This module is necessary for modeling complex short-term tasks in which several alternatives need to be stored in and retrieved from memory.

1

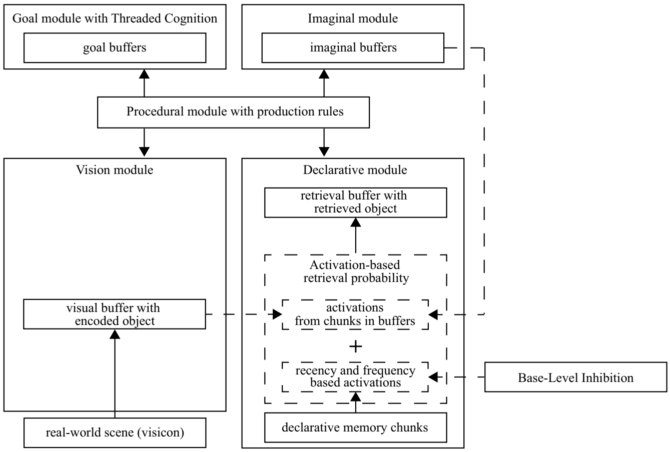


Figure B.2: Internal workings and external connections between vision, declarative, goal, imaginal and procedural modules of ACT-R architecture. These four modules provide the most of the functionalities necessary for modeling SET tasks.

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 a 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/).

Appendix C

Calculation of significant dimension reduction blocks

The probability of *k* subsequent fixations falling on cards that have 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.

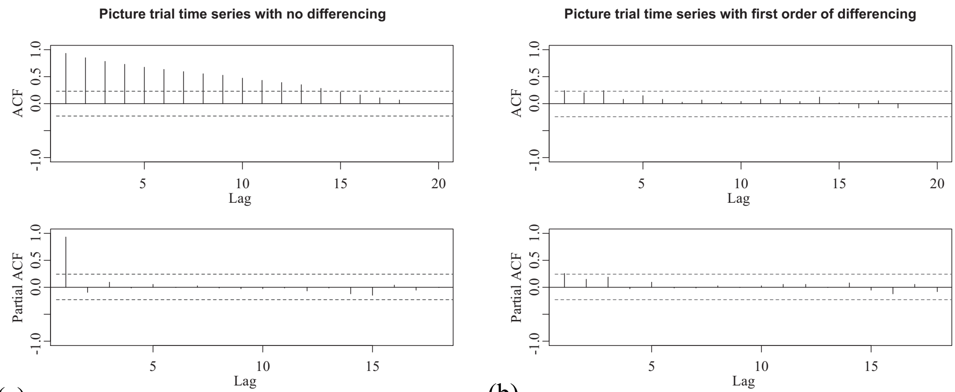
Let's assume that there are five green cards among 12 cards on the desk. If we assume that subjects is always fixating on one of the cards before fixating on another card then the number of possible cards on which subject can fixate is 11. Probability of randomly fixating on one of those 11 cards is 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 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.

Appendix D

Analysis of residuals of ARIMA models applied to proportions of dimension reduction usage by human subjects

We used ARIMA models to reveal the trends in dimension reduction data shown in Figure 5.6b. In both word and picture set, we used points from first 70 fixations as time series for building the ARIMA models. Both ACF and PACF were non-significant for both time series after first order differencing (figures D.1b and D.2b). Therefore, we used ARIMA of order (1, 0, 1) with no seasonal component and no constant terms included in the models. In ARIMA model for picture set, the AR(1) coefficient is 0.9953 (SE = 0.0053, $|z|$ = 187.7925), and MA(1) coefficient is 0.2047 (SE = 0.1106, $|z|$ = 1.851). In ARIMA model for word set ARIMA model, AR(1) coefficient is 0.9978 (SE = 0.0028, $|z|$ = 356.3571) and MA(1) coefficient is 0.0354 (SE = 0.1099, $|z|$ = 0.322111). All coefficients except MA(1) coefficient in word set model are significant. We did a residual diagnostics as goodness-of-fit tests. In both models, the residuals show no trend with no significant correlations present among ACF values of residuals. The Ljung-Box-Pierce statistics done for each lag up to 20 resulted in all non-significant p-values for both ARIMA models (figures D.3 and D.4). All in all, fit for both models were good. We used both models to forecast future trends of using dimension reduction for 60 fixations ahead. The forecast are shown in Figure D.5. Forecasts show downward trends in usage of dimension reduction usage that is in conformance with results from previous analysis.



(a) (b)

Figure D.1: ACF and PACF graphs of proportions of dimension reduction usage for a picture trial (a) with no differencing applied and (b) with first order differencing.

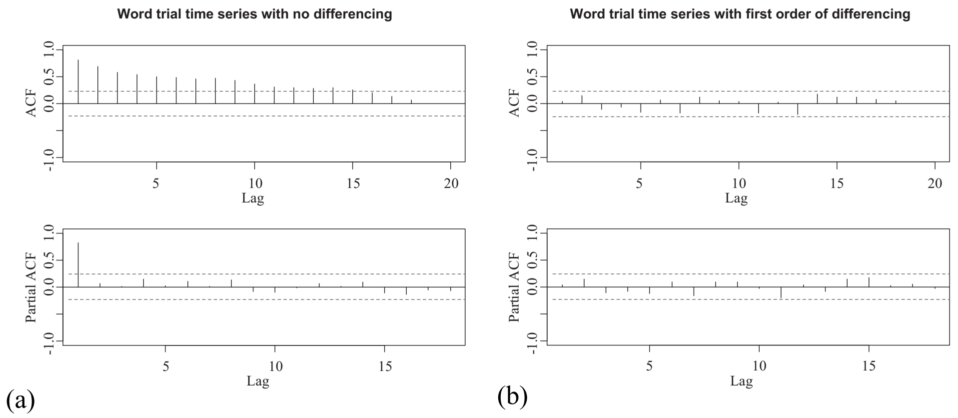


Figure D.2: ACF and PACF graphs of proportions of dimension reduction usage for a word trial (a) with no differencing applied and (b) with first order differencing.

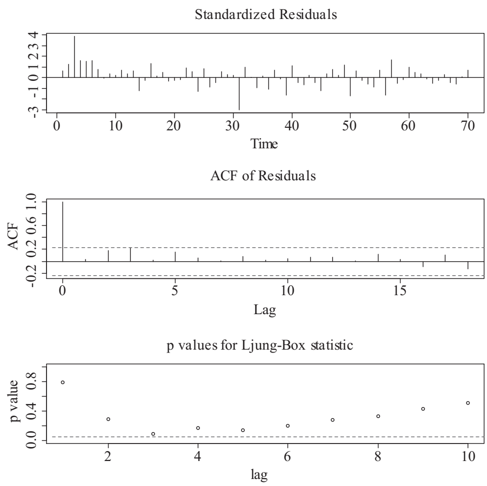


Figure D.3: Analysis of residuals for goodness-of-fit for ARIMA(1, 0, 1) model of picture trial.

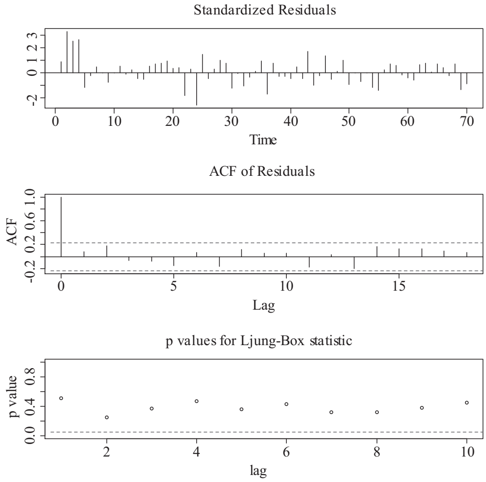


Figure D.4: Analysis of residuals for goodness-of-fit for ARIMA(1, 0, 1) model of word trial.

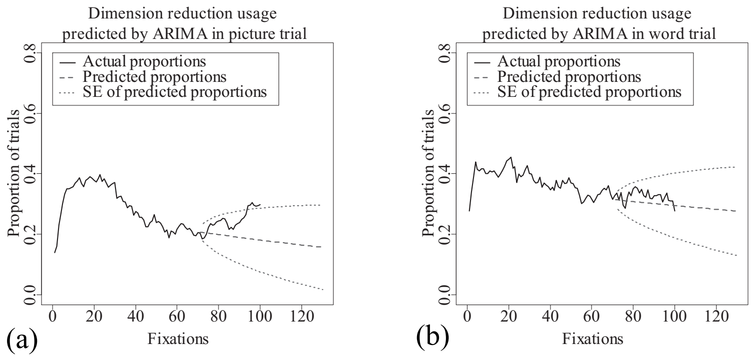


Figure D.5: The proportions of dimension reduction usage in (a) picture and (b) word trials as predicted by ARIMA models.

Appendix E

Analysis of residuals of ARIMA models applied to proportions of dimension reduction usage by ACT-R models

We also used ARIMA models to reveal the trends in models' dimension reduction data. Similarly to analysis of human data, we used points from first 70 fixations as time series. Both ACF and PACF were non-significant for picture set time series after first order differencing (Figure E.1b). Both ACF and PACF decayed quickly for word set time series without any differencing (Figure E.2a). Therefore, we used no differencing in both ARIMA models. No seasonal component and no constant terms were included in the models. A square root transformation was done on the picture set data for better fit of ARIMA model. We used models with (2, 0 1) and (1, 0, 1) orders for picture set and word set respectively. Table E.1 shows the resulting fitted coefficients. Absolute z values indicate that all coefficients are significant.

Table E.1: Fitted coefficients for ARIMA(2, 0, 1) and ARIMA(1, 0, 1) models for picture and word set respectively.

	Picture set			**Word set**	
	AR(1)	**AR(2)**	**MA(1)**	**AR(1)**	**MA(1)**
Coefficients	1.9938	-0.9946	-0.7787	0.9876	0.5452
S.E.	0.0095	0.0096	0.0990	0.0119	0.0937
|z|	209.87	103.60	7.87	82.99	5.82

We did a residual diagnostics as goodness-of-fit tests. In both models, the residuals show no trend with no significant correlations present among ACF values of residuals. The Ljung-Box-Pierce statistics done for each lag up to 20 resulted in all non-significant p-values for both ARIMA models (figures E.3 and E.4). Overall, models fit well the data and can be used for predictions. We used both models to forecast future trends of using dimension reduction for 60 fixations ahead. The forecast are shown in E.5. Similar to human data, forecasts show downward trends in usage of dimension reduction usage.

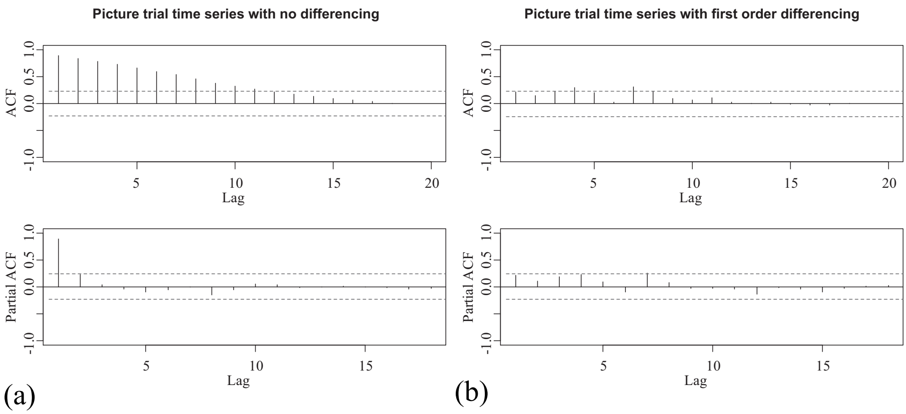


Figure E.1: ACF and PACF graphs of proportions of dimension reduction usage for a picture trial (a) with no differencing applied and (b) with first order differencing.

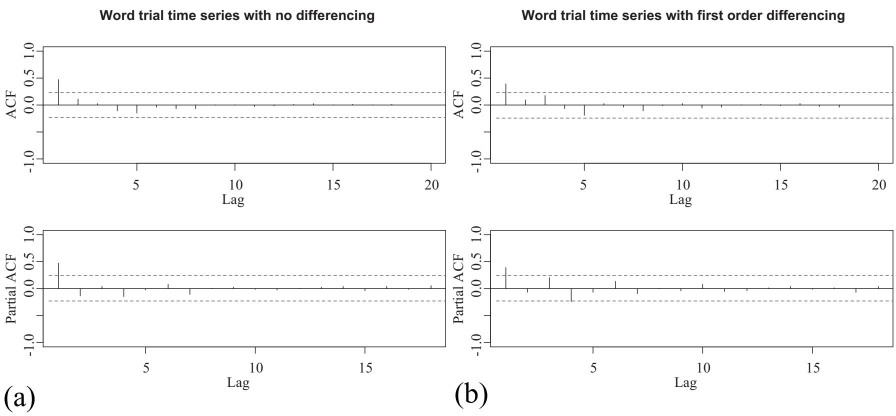


Figure E.2: ACF and PACF graphs of proportions of dimension reduction usage for a word trial (a) with no differencing applied and (b) with first order differencing.

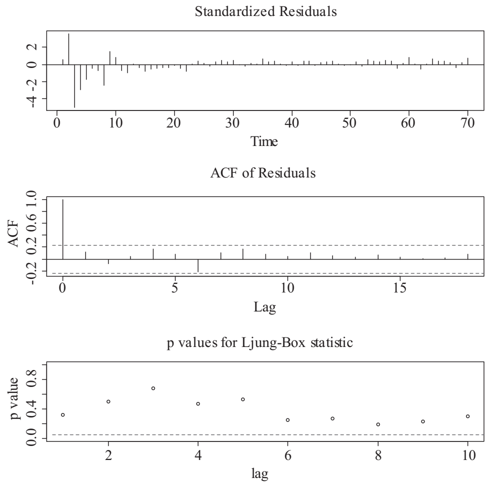


Figure E.3: Analysis of residuals for goodness-of-fit for ARIMA(2, 0, 1) model of picture trial

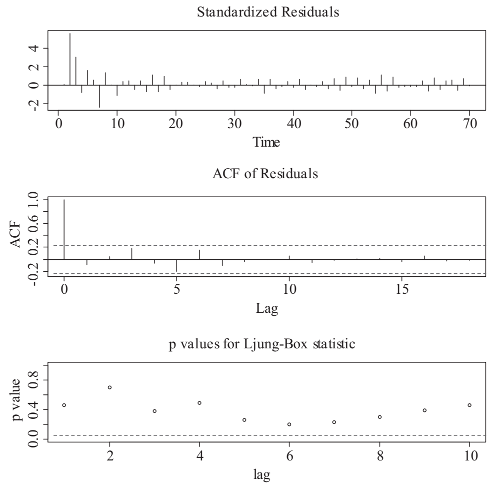


Figure E.4: Analysis of residuals for goodness-of-fit for ARIMA(1, 0, 1) model of word trial.

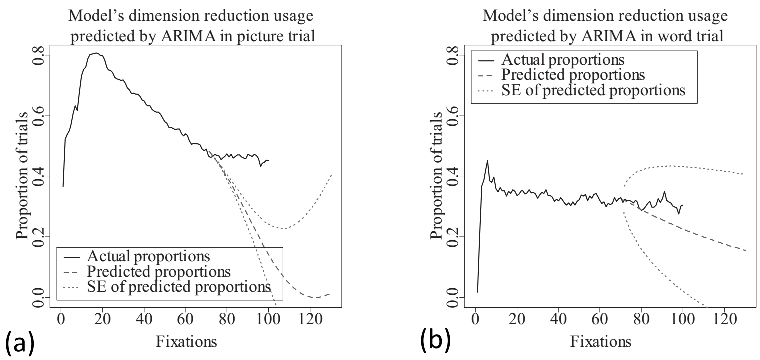


Figure E.5: Models' proportions of dimension reduction usage in word and picture trials as predicted by ARIMA models.

Acknowledgements

On my first date at the University of Groningen, Niels received me in front of the reception at the Bernoulliborg. For a student who spent prior five years in Asian academic culture, it was inconceivable to me that a person, as important as Professor, would personally meet me and give me a tour around the Artificial Intelligence department. After the tour when I finally had a chance to gather my thoughts, I realized that my experience with Prof. Taatgen (this is how I referred to him for the next few weeks after my arrival, undoubtedly, making him uncomfortable) is going to be something totally new.

Over time, I have developed a deep respect for Niels' supervision skills. If you have read acknowledgement sections of dissertations written by Niels' former PhD students then you already know that he is an excellent supervisor. It is not even declarative knowledge anymore. It is procedural knowledge. When someone asks about Niels' supervision skills, a student automatically replies "perfect" without a hint of top-down processing. Previous students left me nothing new to write down about Niels as my supervisor. I can only express my sincere and complete solidarity with other students. However, during the last four years, I have grown to respect Niels not only as a supervisor, but also as a person. Few people are as open minded and understanding as Niels. When talking to Niels, I always had a confidence that I can express my thoughts openly without a fear of being misunderstood or dismissed (Niels certainly perfected the theory of mind). Simply put, Niels is a very good person. I believe that these qualities, in addition to his academic prowess, make Niels such an excellent supervisor. I feel highly fortunate in having had an opportunity to work with Niels. Under his supervision, I have grown not only as a scientist, but also as a person. Indeed, I am very thankful to him for that.

I express my deep gratitude to Jelmer and Ben. As someone who knew next to nothing about cognitive science, I had many ignorant questions. However, Ben and Jelmer have always done their best to teach me and shown an inhumane amount of patience. Just like with Niels, Ben and Jelmer were the people with whom I could openly share my ides and thoughts, something I deeply appreciated all the time.

I would like to thank Prof. Gray, Prof. Roellfsema and Prof. Roerdink for agreeing to spend a significant amount of their time reading my thesis and providing valuable comments and for generously agreeing to come to my defense.

I would like to acknowledge the current and former members of the Cognitive modeling groups. More than once, I had listened to interesting talks during the group meetings that were both inspiration and useful for my own research. Among Cogmod members, I would like to specifically mention Hedderik and Steffi. When I needed help setting up an experiment or advice on a particular research question, Hedderik was always ready to provide a helping hand. Steffi volunteered to proofread and comment on several of my papers. Without any doubts, Steffi contributed significantly to the quality of my papers.

Finally, I express my gratitude to my colleagues at the Artificial Intelligence department. It was a truly unique experience, and I can state with a pride that I studied at the AI department of the RUG. I would like to thank specifically Harmen and Olarik for standing by me as my 'paranimfen'.

References

Addyman, C., & French, R. M. (2012). Computational modeling in cognitive science: A manifesto for change. *Topics in cognitive science , 4* (3), 332-341.

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. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review , 111* (4), 1036-1060.

Anderson, J. R., Fincham, J. M., Schneider, D. W., & Yang, J. (2012). Using brain imaging to track problem-solving in a complex state space. *NeuroImage , 60*, 633-643.

Anderson, J., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological Science , 2* (6), 396–408.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature , 501* (7465), 97-101.

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology , 513* (5), 532-541.

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.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience , 4* (10), 829-839.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science , 3* (1), 2-13.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences , 20* (4), 723-742.

Banks, A. P., & Millward, L. J. (2009). Distributed mental models: Mental models in distributed cognitive systems. *The Journal of Mind and Behavior , 30*, 249-266.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and aging , 23* (4), 765–777.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta psychologica , 129* (3), 387-398.

Borst, J. P., & Anderson, J. R. (2013). Using Model-Based functional MRI to locate Working Memory Updates and Declarative Memory Retrievals in the Fronto-Parietal Network. *Proceedings of the National Academy of Sciences USA , 110* (5), 1628-1633.

Borst, J. P., Taatgen, N. A., & Van Rijn, H. (2011). Using a Symbolic Process Model as input for Model-Based fMRI Analysis: Locating the Neural Correlates of Problem State Replacements. *NeuroImage , 58* (1), 137-147.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge university press.

Braine, M. D., & O'Brien, D. P. (1991). A Theory of "If": A Lexical Entry, Reasoning Program, and Pragmatic Principles. *Psychological Review , 98* (2), 182-203.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence , 47*, 139–159.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive Modeling.* Sage Publications.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review , 100* (3), 432-459.

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review , 108*, 847–869.

Byrne, R. M., & Johnson-Laird, P. N. (1989). Spatial Reasoning. *Journal of Memory and Language , 28*, 564-575.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. *Mapping the mind: Domain specificity in cognition and culture* , 169-200.

Carreiras, M., & Santamaria, C. (1997). Reasoning About Relations: Spatial and Nonspatial Problems. *Thinking and Reasoning , 3* (3), 191-208.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action.* Amsterdam: North-Holland.

Chapman, G. B. (1991). Trial order affects cue interaction in contingency judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition , 17*, 837–854.

Clark, A. (1997). *Being there: Putting brain, body and world together again.* Cambridge, MA: MIT Press.

Coney, J. (1988). Individual differences and task format in sentence verification. *Current Psychology , 7* (2), 122-135.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods , 42* (3), 692-700.

Crystal, D. (1997). *The Cambridge Encyclopedia of Language.* Cambridge: Cambridge University Press.

Cummins, D. D. (1996a). Dominance hierarchies and the evolution of human reasoning. *Minds and Machines , 6* (4), 463-480.

Cummins, D. D. (1996b). Evidence for the innateness of deontic reasoning. *Mind & Language , 11* (2), 160-190.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition , 23* (5), 646-658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition , 19* (3), 274-282.

Destefano, M., Lindstedt, J. K., & Gray, W. D. (2011). Use of complementary actions decreases with expertise. In L. Carlson, C. Hölscher, & T. Shipley (Ed.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2709-2714). 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.

Doya, K. (2008). Modulators of decision making. *Nature Neuroscience , 11* (4), 410-416.

Drewitz, U., & Brandenburg, S. (2012). Memory and Contextual Change in Causal Learning. *Proceedings of the 11th International Conference on Cognitive Modeling* (pp. 265-270). Berlin: TU-Berlin Academic Press.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology , 48*, 269-297.

Entertainment Software Association. (2013). *Essential facts about the computer and video game industry: sales, demographic and usage data.*

Farkaš, I. (2012). Indispensability of Computational Modeling in Cognitive Science. *Journal of Cognitive Science , 13*, 401-435.

Fernandez, E., Bolea, J. A., Ortega, G., & Louis, E. (1999). Are neurons multifractals? *Journal of neuroscience methods , 89* (2), 151-157.

Formisano, E., Linden, D. E., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., et al. (2002). Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron , 35* (1), 185-194.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science.* The MIT Press.

Gigerenzer, G; Todd, P M; Group, the ABC Research. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Glass, B. D., Maddox, W. T., & Love, B. C. (2013). Real-Time Strategy Game Training: Emergence of a Cognitive Flexibility Trait. *PloS One , 8* (8), e70350.

Goertz, G., & Starr, H. (Eds.). (2003). *Necessary conditions: Theory, methodology and applications.* Rowman & Littlefield Pub Incorporated.

Green, C. S., & Bavelier, D. (2004). The cognitive neuroscience of video games. In P. Messaris, & L. Humphreys (Eds.), *Digital media: Transformations in human communication* (pp. 211–223). New York: Peter Lang.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and Blickets: Effects of Knowledge on Causal Induction in Children and Adults. *Cognitive Science , 35*, 1407–1455.

Gust, H., Krumnack, U., Kühnberger, K. U., & Schwering, A. (2008). Analogical reasoning: A core of cognition. *Zeitschrift für Künstliche Intelligenz (KI), Themenheft KI und Kognition , 1*, 8-12.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral Brain Sciences , 21* (6), 803–831.

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.

Hedden, T., & Zhang, J. (2002). What do you think I think?: Strategic reasoning in matrix games. *Cognition , 85* (1), 1-36.

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics , 65* (1), 58-71.

Hong, S. K., & Drury, C. G. (2002). Sensitivity and validity of visual search models for multiple targets. *Theoretical Issues in Ergonomics Science , 3* (1), 85-110.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica (Amsterdam) , 26* (2), 107–129.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics , 63* (2), 272-285.

Humphrey, G. K., & Lupker, S. J. (1993). Codes and operations in picture matching. *Psychological Research , 55*, 237–247.

Ingrand, F. F., Georgeff, M. P., & Rao, A. S. (1992). An architecture for real-time reasoning and system control. *EEE Expert: Intelligent Systems and Their Applications (IEEE Press) , 7* (6), 34–44.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence , 20* (11), 1254-1259.

Jacob, M., & Hochstein, S. (2008). Set recognition as a window to perceptual and cognitive processes. *Perception & Psychophysics , 70* (7), 1165-1184.

Jeffrey, R. (1981). *Formal logic: Its scope and limits (2nd ed.).* New York, NY: McGraw-Hill.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 26* (3), 683-702.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness.* Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (2004). The history of mental models. *Psychology of reasoning: Theoretical and historical perspectives* , 179-212.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist , 58* (9), 697–720.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-computer interaction , 4* (12), 391-438.

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). Philadelphia, PA: Drexel University.

Kieras, D. (2011). The persistent visual store as the locus of fixation memory in visual search tasks. *Cognitive Systems Research , 12* (2), 102-112.

Kim, M., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science , 6* (6), 376–380.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition , 83* (2), B35-B42.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science , 18*, 513-549.

Kiselev, V. G., Hahn, K. R., & Auer, D. P. (2003). Is the brain cortex a fractal? *Neuroimage , 20* (3), 1765-1774.

Klanderman, D. B., Moore, M. W., Maxwell, M. S., & Robbert, S. K. (2013). Creating Problems and Their Solutions: Service-Learning Through Trinity Mathematics Triathlons, Math Nights, and Math Centers. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies , 23* (6), 563-571.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience , 30*, 57-78.

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.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence , 33* (1), 1-64.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and movements in the control of activities of daily living. *Perception , 28*, 1311–1328.

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* (pp. 2378-2383). Austin, TX: Cognitive Science Society.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics–Doklady , 10*, 707-710.

Loftus, g. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance , 4*, 565-572.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *ACTA Psychologica , 75* (1), 55-74.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex , 48* (9), 1138-1154.

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science , 14* (3), 582-590.

Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly , 2* (4), 245–265.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics , 2*, 547-552.

Maloy, R. W., Edwards, S. A., & Anderson, G. (2010). Teaching Math Problem Solving Using a Web-Based Tutoring System, Learning Games, and Students' Writing. *Journal of STEM Education: Innovations and Research , 11* (1), 82-90.

Markram, H., & Perin, R. (2011). Innate neural assemblies for lego memory. *Frontiers in neural circuits , 5*.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Science , 6*, 255–260.

Meijering, B., Van Maanen, L., Van Rijn, H., & Verbrugge, R. (2010). The Facilitative Effect of Context on Second-Order Social Reasoning. In R. Catrambone, & S. Ohlsson (Ed.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 1423-1428). Austin, TX: Cognitive Science Society.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review , 104* (1), 3-65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review , 104* (4), 749-791.

Michon, J. A. (1967). The game of JAM: An isomorph of Tic-Tac-Toe. *The American Journal of Psychology , 80*, 137-140.

Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General , 125*, 370–386.

Myung, J., & Pitt, M. (2010). Cognitive modeling repository. *Cognitive Science* .

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* , 443-453.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory , 6*, 391-399.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Nyamsuren, E., & Taatgen, N. A. (2013d). From Casual Deduction to Spatial Relations: Bottom-up and Top-down Reasoning Unified. In R. West, & T. Stewart (Ed.), *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 203-208). Ottawa: Carleton University.

Nyamsuren, E., & Taatgen, N. A. (2013a). Pre-attentive and Attentive Vision Module. *Cognitive Systems Research , 24*, 62-71.

Nyamsuren, E., & Taatgen, N. A. (2013b). Set as instance of a real-world visual-cognitive task. *Cognitive Science , 37* (1), 146-175.

Nyamsuren, E., & Taatgen, N. A. (2013c). The synergy of top-down and bottom-up attention in complex task: going beyond saliency models. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3181-3186). Austin, TX: Cognitive Science Society.

Orban, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences , 105* (7), 2745-2750.

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences , 108* (13), 5419-5424.

Phillips, W. A. (1983). Short-Term Visual Memory. *Royal Society of London Philosophical Transactions Series B , 302*, 295-308.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General , 113*, 426-442.

Pomplun, M., & Ritter, H. (1999). A three-level model of comparative visual search. *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 543–548). Mahweh, NJ: Lawrence Erlbaum Associates.

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.

Ponsoda, V., Scott, D., & Findlay, J. M. (1995). A probability vector and transition matrix analysis of eye movements during visual search. *Acta Psychologica , 88*, 167-185.

Pothos, E. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences , 28*, 1–49.

Potter, M., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology , 81*, 10–15.

Proverbio, A. M., Zotto, M. D., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience , 8*, 24.

Rajasekar, A., Lobo, J., & Minker, J. (1989). Weak generalized closed world assumption. *Journal of automated reasoning , 5* (3), 293-307.

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.

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

Rensink, R. A. (2000b). Seeing, sensing, and scrutinizing. *Vision research , 40* (10), 1469-1487.

Rensink, R. A. (2000a). The dynamic representation of scenes. *Visual cognition , 7* (1-3), 17-42.

Rensink, R. A. (2007). The modeling and control of visual perception. In W. D. Gray, *Integrated Models of Cognitive Systems* (pp. 132-148). New York: Oxford.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 34* (2), 328-342.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review , 90* (1), 38-71.

Roberts, M. J. (1993). Human Reasoning: Deduction Rules or Mental Models, or Both? *The Quarterly Journal of Experimental Psychology , 46A* (4), 569-589.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research , 1* (4), 201–220.

Salvucci, D. D. (2010). Cognitive supermodels. *Invited talk, European ACT-R Workshop.* Groningen, The Netherlands.

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.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review , 1*, 101–130.

Schooler, L. J., & Hertwig, R. (2005). How Forgetting Aids Heuristic Inference. *Psychological Review , 112* (3), 610-628.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgment. *Quarterly Journal of Experimental Psychology , 37B*, 1–21.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (pp. 229–261). San Diego, CA: Academic Press.

Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2004). Children's casual inferences from direct evidence: Backward blocking and Bayesian reasoning in preschoolers. *Cognitive Science , 28*, 303-333.

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. (2013). The nature and transfer of cognitive skills. *Psychological Review , 120* (3), 439-471.

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). Bamberg: Universitätsverlag Bamberg.

Taatgen, N. A., Van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review , 114* (3), 577-598.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics , 51*, 599–606.

Thompson, V. A. (1996). Reasoning from false premises: The role of soundness in making logical deductions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale , 50* (3), 315-319.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science , 1*, 46-54.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology , 12*, 97–136.

Treue, S. (2003). Visual attention: the where, what, how and why of saliency. *Current Opinion in Neurobiology , 13*, 428-432.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review , 108* (3), 550–592.

van Diepen, P. M., De Graef, P., & d'Ydewalle, G. (1995). Chronometry of foveal information extraction during scene perception. *Eye Movement Research - Mechanisms, Processes, and Applications , 6*, 349–362.

Verguts, T., & Fias, W. (2009). Similarity and Rules United: Similarity- and Rule-based Processing in a Single Neural Network. *Cognitive Science , 33*, 243-259.

Wais, P. E., Rubens, M. T., Boccanfuso, J., & Gazzaley, A. (2010). Neural mechanisms underlying the impact of visual distraction on long-term memory retrieval. *Journal of Neuroscience , 30* (25), 8541-8550.

Weitzenfeld, J. S. (1984). Valid reasoning by analogy. *Philosophy of Science Association , 51*, 137-149.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 28*, 872–892.

Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition , 29*, 1106–1118.

Winston, P. H. (1980). Learning and reasoning by analogy. *Communications of the ACM , 23* (12), 689-703.

Wintermute, S. (2012). Imagery in cognitive architecture: Representation and control at multiple levels of abstraction. *Cognitive Systems Research , 19-20*, 1-29.

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford.

Wolfe, J. M. (1992). Parallel guidance of visual attention. *Current Directions in Psychological Science , 1* (4), 124-128.

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science , 9* (1), 33-39.

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research , 37* (1), 25-44.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance , 15*, 419–433.

Wolfe, J., & Horowitz, T. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience , 5* (6), 495-501.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2007). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience , 1* (8).

Xu, Y., & Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature , 440* (7080), 91-95.

List of Publications

Journal Articles

Nyamsuren, E. & Taatgen, N. A. (accepted). The Human Reasoning Module. *Biologically Inspired Cognitive Architectures*.

Nyamsuren, E. & Taatgen, N. A. (2013). The effect of visual representation style in problem-solving: a perspective from cognitive processes. *PLoS ONE 8 (11): e80550. doi:10.1371/journal.pone.0080550.*

Nyamsuren, E. & Taatgen, N. A. (2013). Pre-attentive and Attentive Vision Module. *Cognitive Systems Research*, 24, 62-71.

Nyamsuren, E. & Taatgen, N. A. (2013). Set as instance of a real-world visual-cognitive task. *Cognitive Science*, 37 (1), 146-175.

Peer-Reviewed Conference Papers

Nyamsuren, E. & Taatgen, N. A. (2013). The synergy of top-down and bottom-up attention in complex task: going beyond saliency models. *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3181-3186). Austin, TX: Cognitive Science Society.

Nyamsuren, E. & Taatgen, N. A. (2013). From Casual Deduction to Spatial Relations: Bottom-up and Top-down Reasoning Unified. *Proceedings of the 12th International Conference on Cognitive Modeling* (pp. 203-208). Ottawa: Carleton University.

Nyamsuren, E. & Taatgen, N. A. (2012). Pre-attentive and Attentive Vision Module. *Proceedings of the 11th international conference on cognitive modeling* (pp. 211-216). Berlin: TU-Berlin Academic Press.

Nyamsuren, E. & Taatgen, N. A. (2011). Top-down Planning and Bottom-up Perception in a Problem-solving Task. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2685-2690). Austin, TX: Cognitive Science Society.

Abstracts

Nyamsuren, E. & Taatgen, N. A. (2011). Modeling Bottom-up Visual Influence in Decision Making. *Interdisciplinary Workshop on Cognitive Neuroscience, Educational Research and Cognitive Modeling*. Delmenhorst, Germany.