COGNITIVE ASPECTS OF CONCEPTUAL MODELING DIAGRAMS: AN EXPERIMENTAL STUDY

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF INFORMATICS

OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

ÖZKAN KILIÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

SEPTEMBER 2007

Approval of the Graduate School of Informatics

Prof. Dr. Nazife Baykal
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Deniz Zeyrek
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Onur Demirörs
Co-Supervisor

Asst. Prof. Dr. Bilge Say
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Cem Bozşahin (METU, COGS)

Asst. Prof. Dr. Bilge Say (METU, COGS)

Assoc. Prof. Dr. Onur Demirörs (METU, IS)

Assoc. Prof. Dr. Kürşat Çağıltay (METU, CEIT)

Asst. Prof. Dr. Annette Hohenberger (METU, COGS)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this wok.

Name, Last name: Özkan KILIÇ

Signature: ___________________

ABSTRACT

COGNITIVE ASPECTS OF CONCEPTUAL MODELING DIAGRAMS: AN EXPERIMENTAL STUDY

Kılıç, Özkan
M.Sc., Department of Cognitive Science
Supervisor: Asst. Prof. Dr. Bilge Say
Co-Supervisor: Assoc. Prof. Dr. Onur Demirörs

September 2007, 148 pages

This thesis is about diagrammatic reasoning and error-finding in conceptual modeling diagrams. Specifically, the differences of the cognitive strategies and behaviors of notation-familiar participants versus domain-familiar participants working on conceptual modeling diagrams are inspected. The domain-familiar participants are experienced in the topic being represented, but they do not have any formal training in software development representations. On the other hand, the notation-familiar participants are educated in software representations, but unfamiliar with the topic represented. The main experiment and the follow-up experiment also aim to study how some properties of diagrams affect the error-finding behaviors. The participant groups' performances in the main experiment are investigated and compared by the analysis of verbal protocol data and eye movement data. The combination of the two different methods enhances detailed analyses. In the follow-up experiment, only eye movement data is involved to evaluate how some properties of diagrams affect problem-solving. By means of both experiments, it is concluded that diagrammatic complexity has a negative effect on reasoning whereas the degree of causal chaining improves diagrammatic reasoning. In the main experiment, some differences in the diagrammatic reasoning processes between the groups are observed, too. The notation-familiar participants are observed to be more successful in error-finding although they are unfamiliar with the topic. This study underlines the interaction of cognitive science and software engineering by integrating eye movement data, verbal protocol analysis and performance data into the cognitive inspection of software engineering notations.

Keywords: Diagrammatic Reasoning, Eye Movements, Verbal Protocol Analysis

ÖZ

KAVRAMSAL MODELLEME DİYAGRAMLARININ BİLİŞSEL ÖZELLİKLERİ: DENEYSEL BİR ÇALIŞMA

Kılıç, Özkan
Master, Bilişsel Bilimler Bölümü
Tez Yöneticisi: Yrd. Doç. Dr. Bilge Say
Ortak Tez Yöneticisi: Doç. Dr. Onur Demirörs

Eylül 2007, 148 sayfa

Bu çalışma grafiksel muhakeme ve kavramsal modelleme diyagramlarında hata bulma ile ilgilidir. Özellikle, konuya aşina olan katılımcılar ile gösterimlere aşina olan katılımcıların kavramsal modelleme gösterimleri üzerinde çalışırken sergiledikleri bilişsel strateji ve davranışlardaki farklar çalışılmıştır. Konuya aşina olan katılımcılar konu hakkında bilgi sahibidirler ama yazılım mühendisliği gösterimleri konusunda eğitimleri yoktur. Öte yandan, gösterimlere aşina katılımcılar yazılım mühendisliği gösterimlerinde eğitim almışlardır ama konuya aşina değillerdir. Ana deney ve takipçi deney, ayrıca, diagramların bazı özelliklerinin hata bulma davranışlarını nasıl etkilediğini çalışmayı amaçlamaktadır. Ana deneydeki katılımcı gruplarının performansları, sözel protokol verileri ve göz hareketi verilerinin analizi yoluyla çalışılmış ve karşılaştırılmıştır. İki farklı metodun birleşimi, detaylı analizleri güçlendirmiştir. Takipçi deneyde diagramların bazı özelliklerinin problem çözmeyi nasıl etkilediğini daha detaylı görmek amacıyla yalnızca göz hareketi verileri kullanılmıştır. Her iki deney yoluyla, diyagramsal karmaşanın muhakeme üzerinde negatif bir etkisi olduğu, diğer yandan nedensel bağ derecesinin grafiksel muhakemeyi geliştirdiği sonucuna varılmıştır. Ana deneyde grupların grafiksel muhakeme işlemlerinde bazı farkların da olduğu gözlenmiştir. Gösterimlere aşina katılımcıların konuya aşina olmamalarına rağmen hata bulma konusunda daha başarılı oldukları gözlenmiştir. Bu çalışma göz hareketi verileri, sözel protocol analizi ve performans verilerinin yazılım mühendisliği gösterimlerinin bilişsel incelemesine entegrasyonu sayesinde bilişsel bilimler ve yazılım mühendisliği arasındaki etkileşimin önemini vurgulamaktadır.

Anahtar Kelimeler: Grafiksel Muhakeme, Göz Hareketleri, Sozel Protokol Analizi

To My Brother Şükrü Kılıç

ACKNOWLEDGMENTS

I express honest gratitude to Asst. Prof. Dr. Bilge Say for her guidance and insight throughout the research. Thanks also go to my co-supervisor Assoc. Prof. Dr. Onur Demirörs for his support. The assistance of Özlem Gümüş, Asst. Prof. Dr. Neşe Alkan, Asst. Prof. Dr. Annette Hohenberger, Assoc. Prof. Dr. Kürşat Çağıltay, Asst. Prof. Dr. Deepti Mishra, Gül Tokdemir, Muhsin Civelek, Cengiz Acartürk, Hacer Karacan and Özge Alaçam is gratefully acknowledged.

I offer sincere thanks to my family and my best friends Bircan Akgül, Mükremin Çakır and Kuzey Ali Solak, the driver, for their support and faith.

The Tobii Eye Tracker in this study is granted by TÜBİTAK (SOBAG 104K098). I would like to thank to Middle East Technical University Human Computer Interaction Research Group (http://hci.metu.edu.tr), and also, BİLKARDEM and the KAMA Project Staff for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Using and Reasoning with Diagrams

Diagrams are used in many fields as design aids or problem-solving devices. Diagrammatic reasoning refers to how diagrammatic (or pictorial) representations can be used in problem-solving and reasoning (Chandrasekaran, Glasgow & Narayanan, 1995). Diagrammatic representations in reasoning processes have been investigated in order to study the interplay of mental representations and external representations[1] (Barkowsky, Freska, Hegarty & Lowe, 2005) and how diagrams help to form cognitive spatial maps. Informed by such studies, researchers came up with some cognitive models, such as models of visual attention (Mozer & Sitton, 1998), memory (Raaijimakers & Shiffrin, 2002), and manipulation processes (Hegarty, 1992), derived from cognitive aspects of diagrammatic reasoning.

Mental representations reside in the working memory and are produced from the interaction of cognitive faculty with the diagrams, which are external representations (Kosslyn, Seger, Pani & Hillger, 1990). This coupled system of mental and external representations plays a key role in the use of

[1] External representations mean the diagrams on physical mediums.

diagrams in problem-solving. The problems are solved through inference and the visual perception process from both the symbolic and the diagrammatic part of the representations. Visual perception is achieved through eye movements, which play an important role in diagrammatic problem-solving because it provides the working memory with the mental image of external representations. Thus, diagrammatic reasoning studies usually involve eye movement data to study the interaction of the mental and external images.

Diagrammatic representations are very common in many fields, like physics, mathematics, architecture, mechanical systems, and many engineering areas. In software engineering, it is very common and almost compulsory for the large systems' designers to prepare diagrams during the development phase. The use of graphical representation together with textual information clarifies the discussions of the design ideas in software development because usually a graphical modeling language, such as Unified Modeling Language (UML), is employed to model entities or objects to be coded as computer programs (Rumbaugh, Jacobson & Booch, 1998). In software development, the development experts usually have to produce software in other expertise domains. Therefore, they have to work with the experts that are familiar with various target domains. In such an example, the representations play the role of a communication channel between the development experts and the domain experts. Hence, an understanding of how different experts' ways of working with diagrams differ is meaningful and important for effectively using such a medium of communication.

Thus, motivated by the need for a better understanding of cognitive aspects of diagrammatic reasoning, both from a foundational cognitive science and an applied perspective, this thesis investigates the effect of domain and notation-familiarity as well as properties of diagrams themselves on error-finding in a certain kind of software engineering diagram. There are few similar studies in a similar vein. For example, Hungerford, Hevner, and

Collins (2004) study the success of software development experts on finding errors in software engineering notations. However, it is mainly a performance-based study, in which the number of errors found, task completion time and demographic data are mainly taken into account. Hungerford et al. (2004) observe the error-finding strategies and successes of experienced software developers on software diagrams. Similarly, many software notation testing techniques are limited and depend on checklist criteria and performances (Juristo, Moreno & Vegas, 2003). Since the notations are derived from other expertise domains and play the role of a communication channel between the developers and domain experts, a more detailed and comparative study is required as in this study.

1.2 The Aim and Scope of the Present Study

In this thesis, a conceptual modeling (see section 2.3.1 Conceptual Modeling, on page 28) tool notation developed within the KAMA project[1] (KAMA, 2005) has been used with a given domain-specific scenario to observe the effect of domain versus notation-familiarity of participants and certain properties of diagrams on error-finding by using eye tracking and verbal protocol data synchronously, in addition to performance data. Certain effects of diagrammatic properties on error-finding behavior have been further confirmed by a follow-up experiment. The details of the material and the rationale behind the choice of methods are explained in Chapter 2 and Chapter 3. In order to fulfill our aim, 10 notation-familiar participants, who are from the software engineering domain, and 10 domain-familiar participants, who are from the military domain, have participated in the main experiment in the Human Computer Interaction Laboratory (HCI) at Middle East Technical University (METU). The participants are asked to find the intentionally placed errors in the diagrams of a scenario. While finding errors

[1] The KAMA tool, "Kavramsal Modelleme Aracı", is a simulation conceptual modeling tool developed for the Turkish Army.

and validating the notation-scenario pair, the participants are asked to think aloud. Moreover, an eye tracking device is employed to obtain eye movement data. Each participant's video is also recorded and subjected to verbal protocol analysis. The synchronization of protocols and eye movement data is combined with performance data, which are time to complete the task and number of errors found, and demographic data. The combination of verbal protocol and eye movement data is used in the statistical analysis to see the effects of familiarity and the effects of some diagrammatic properties of error-finding behavior on conceptual modeling notations.

After the main experiment, some properties of diagrams are found to affect reasoning. These properties are the computational complexity[1] and the degree of causal chaining[2]. In order to distinguish the first property from the computational complexity in computer science, we will use the term, diagrammatic complexity in this thesis (see the section 3.4.4 Diagrammatic Complexity and Degree of Causal Chaining on page 67). In order to test these properties, 24 additional university students are studied in the follow-up experiment, in which only the eye movement data is collected. The aim of the follow-up experiment is to explore the effects of diagrammatic complexity and the degree of causal chaining properties of diagrams through eye movement and performance data.

1.3 Research Questions

The research questions of the study can be stated as follow:

> **Q1:** How do different kinds of domain-familiarity affect diagrammatic reasoning?

[1] Computational complexity refers to the time and space required to finish a task.
[2] The degree of causal chaining refers to the linking of objects in a cause-effect manner.

Q2: How does experience[1] affect the error-finding activity?

Q3: How and which error types affect the success rate of error-finding significantly?

Q4: How do the diagrammatic complexity and the degree of causal chaining affect diagrammatic reasoning?

Q5: Does gender have effect on these properties?

The first three questions belong to the main experiment and the following two questions are for the follow-up experiment. The corresponding hypotheses will be given in later sections.

1.4 Limitations

First of all, the verbal protocol method elicits a sample of the participant's overt expression of cognitive processing and cannot be viewed as a complete report of what a participant is thinking. Concurrent verbal protocols are categorized according to observations and inferences after the main experiment. Eye movement data also have correlations with protocols. The combination of methods and the warm-up task are used to minimize some negative aspects of verbal protocols. Another limitation of the study is the number of participants. Since the experiments are performed at METU and take a long time, it has been difficult to invite especially the domain-familiar participants to the HCI laboratory as such invitations have to be out of regular working hours. Another problem is the age and experience difference. The reason for this difference is that the time required for training the domain-familiar participants is much higher than the time required for the notation-familiar ones. Moreover, it is very difficult to find experienced notation-familiar participants and invite them to METU because the

[1] Experience is measured as the number of professional working years in the expertise fields.

experienced notation-familiar participants are usually at administrative positions of companies. As a result, the age and, consequently, the experience of domain-familiar group are higher than that of the notation-familiar one. The last limitation of the study is the concentration on a single gender. Since it is very rare to find female domain-familiar participant in the military domain and all of the domain-familiar participants in this study are male, male are chosen for the notation-familiar participants for the study, too, in order to preserve the balance. However, in the follow-up experiment, in which 24 participants are employed to assess the diagrammatic complexity and the degree of causal chaining properties of diagrams, 12 female and 12 male university students are invited to see the effect of gender.

The outline of the rest of the thesis is as follows: In Chapter 2, diagrammatic reasoning is discussed in a detailed way and conceptual modeling is introduced. Chapter 3 explains the method, results and discussion of the main experiment. The method, results and discussion of the follow-up experiment is given in Chapter 4. The conclusions and future work resides in Chapter 5.

CHAPTER 2

COGNITIVE AND SOFTWARE ENGINEERING ASPECTS OF DIAGRAMMATIC REASONING

2.1 Significance of Diagrammatic Reasoning

Diagrams are pictorial and abstract representations of information. Maps, line graphs, bar charts, engineering blueprints, and architects' sketches are examples of diagrams, but photographs and video are not. Moreover, diagrams are visuo-spatial representations that utilize spatial and visual properties of components to capture and convey information. Diagrammatic reasoning refers to how diagrammatic (or pictorial) representations can be used in problem-solving and reasoning (Chandrasekaran et al., 1995). The active images held in working memory during a problem-solving task with diagrams are called mental images. The mental imagery allows diagrammatic problem solvers to create, modify and even animate mental images to aid in cognitive activities, like diagrammatic reasoning, through visual information. Chandrasekaran et al. (1995) state that diagrammatic reasoning is not about how raw sensory information in the visual modality is processed to form percepts, which is the subject of theories of image processing and perception. Instead, they describe the issue as representation of diagrams and mental images and the functions played by them in problem-solving. It requires multidisciplinary studies to conduct research in diagrammatic reasoning. For example, design theorists study diagrams as fundamental design aids, and logicians regard diagrams as heuristic aids on the path to the proof, and, artificial intelligence researchers study diagrammatic reasoning as the discrete symbol processing for artificial intelligent problem-solving.

In the real world, there are external diagrammatic representations on a medium. In a problem-solving process with diagrams, the agent needs to infer visual information and construct his/her own internal diagrams or images in the working memory. Chandrasekaran et al. (1995) use the term "visual information" to refer to information that the agent can extract by inspection from an image or from the world by directing visual attention to it. However, design schemas, color coding of the diagrammatic primitives and spatial distribution influence the internal images or visual information of the viewers depending on the personalities, preferences and cognitive capabilities (Kalyuga, Chandler & Sweller, 1997). Furthermore, information distribution, organization, relations among visual components and visual grammar, which is the set of distribution and relation rules among the components, affect problem-solving behavior and duration a lot (Hahn & Kim, 1999). Since what can be visually extracted depends on persons, their training and other personal factors, the exact characteristics of visual information cannot be given. However, shapes, certain simple spatial relations, color, texture and such are the kinds of information included in visual information.

Besides its definition and relation to psychology and other fields, the history and the aim of diagrammatic reasoning studies are worthy of review. How human being infers information from the existing world is a discussed question since the Greeks. Simon (1995) reports that reasoning in language and reasoning from diagrams or other pictorial representations were first differentiated by the Pythagorean discovery of irrational numbers, which have no place among integers or fractions but are essential for representing length of lines in geometry, for example, the ratio of the diagonal to the side of the square. With the Descartes' invention of analytic geometry, diagrammatic representations were able to represent some phenomena better, like irrational numbers. Moreover, the Swiss mathematician Leonhard Euler proposed using circles to illustrate relations between sets and to generate solutions for problems in class logic (Euler, 1768, as cited in Simon, 1995, *p*.9). In the 1880s, John Venn greatly improved this method by using diagrams of overlapping regions (i.e., topological models) to illustrate truth-conditions of propositions (Venn, 1881, as cited in Simon, 1995, *p*.9). In this period of time, diagrammatic representations and proofs were very fashionable. However, in 1900s Dedekind's symbolic mathematics later became popular for proofs and made diagrammatic proofs less favorable (Simon, 1995). Thus, diagrammatic and linguistic representations found common ground in 20th century to be studied in order to understand human cognition.

However, there is still a debate between sentential cognitive processes versus diagrammatic cognitive processes of a common problem: Which kind of representations are useful in problem-solving and what are their differences in use of cognitive processes? Larkin and Simon (1987) defined the external representations for a problem as either sentential or diagrammatic representations, and then discussed their use by observing human problem solvers. In their study, they concluded that following features of diagrams are superior or beneficial with respect to sentential forms: First, diagrams can group together all the information used collectively, thus avoiding too much search for problem-solving inferences. Second, diagrams use locations to group information about a single element, avoiding the need to match symbolic labels. Third, diagrams support large amount of perceptual inferences. In order to understand the roles or significance of diagrams in problem-solving and the difference between sentential and diagrammatic representations, cognitive processes and the utilization of working memory in diagrammatic reasoning and proposed models should be reviewed.

2.2 Properties of Diagrams That Affect Cognitive Processes

Simon (1995) states that research in diagrammatic reasoning has two goals: The first goal is to deepen our understanding of ourselves and the way in which we think. In other words, diagrammatic reasoning research enlightens human cognition through thinking with diagrams, how people store and utilize pictorial information during understanding, modifying, reasoning, and creating pictorial phenomenon in their minds. The second goal is to provide an essential scientific base for constructing representations of diagrammatic information that can be stored and processed by computers. If the cognitive processes in diagrammatic reasoning are investigated and scientifically formulated, it will help in the development of software performing diagrammatic reasoning in the field of artificial intelligence. By developing and storing computational diagrammatic information, computers can partially imitate the human way of diagrammatic thinking. This will enable computers to achieve some of the computational efficiencies in their processes as diagrams provide for human beings. Thus, computerized diagrammatic reasoning models will be a reflection of human cognition and may serve as test bed for computational cognitive models.

In order to improve models of diagrammatic reasoning, the diagrams' features, like expressiveness, computability, functionality, and such, should be understood. The diagram should not only be complete by expressing fully all the facts, but also be effective in presenting the information in a way that makes it easy to perceive and reason with (Mackinlay & Genesereth, 1985). This point relates to the name of expressiveness versus computability of diagrams. Expressiveness determines the success and visual properties of diagrams in fully representing the information. In other words, if the viewer needs to fully understand a system represented by diagrams, expressiveness of diagrams is concerned. However, computability determines how information is indexed, what logic is used for indexing, in which ways and how often relations, loops and conditions are embedded into diagrams while expressing information. In other words, diagrams can be better representations in terms of usefulness not because they contain more information, but because the indexing of this information can support extremely useful and efficient computational processes. It can be roughly stated that expressiveness affects understanding of diagrams while computability is dominant in problem-solving with diagrams. Furthermore, when a problem solver using diagrams in the study needs a “better” representation, he or she chooses the more computational one (to take advantages of it), not the more expressive one (Larkin & Simon, 1987; Hahn & Kim, 1999). The reason is that when the help of diagrams in problem-solving is considered, the advantages of diagrams are computational (Larkin & Simon, 1987; Hahn & Kim 1999; Kim, Hahn & Hahn, 2000). The problem solvers use computational representations in order to reduce the solution search space and involve functional diagrammatical components into solution production.

Besides indexing, loops and conditions, logic also improves the computability. Stenning and Lemon (2001) emphasize the function of logic in the analysis of diagrammatic representations. They discuss the combination of logic and psychology to show that diagrammatic representations with logical analogies provide better semantic interpretations. Their argument is consistent with Larkin and Simon (1987) and Simon (1995) because logical and semantic representations imply computational diagrams, which are mathematically expressed and formulated.

Indeed, diagrams are not universally valuable and useful. The expressive and computational powers and their balance make them more beneficial according to the task in which diagrams are used. Since the same information can be given in different representations, expressiveness is something subjective and immeasurable. Even though the different representations are informationally equivalent (i.e., represent the same content), they will not necessarily be computationally equivalent (i.e., be equally easy to use) (Larkin & Simon, 1987) because different diagrammatic representations provide different perceptual cues that affect the amount of search effort that is required for problem-solving with diagrams (Zhang, 1997).

In order to understand the cognitive processes of diagrammatic reasoning, first the roles of human memory in diagrammatic reasoning needs to be discussed. Then, problem-solving by diagrams as well as cognitive and computational aspects of diagrammatic reasoning models will be reviewed.

2.2.1 Working Memory and Diagrammatic Reasoning

Working memory is the system that maintains and stores information in the short term and that underlies human thought processes (Baddeley, 2003). Its basic model was proposed more than 30 years ago and it continues to evolve. Besides its main executive part, the central executive, the working memory model's main components are the phonological loop, the visuospatial sketchpad and the episodic buffer as shown in Figure 1.

Working memory is a necessary cognitive component for temporary knowledge storing, learning, planning and control of action, computation, reasoning, language processing, language acquisition, building up an identity and such. Its components sometimes work cooperatively for different purposes under the control of the central executive. The central executive, which is the most important but least understood part, is the supervisory functional component of the model (Baddeley, 2003). It controls, coordinates and regulates the slavery components and actively participates in memory tasks.

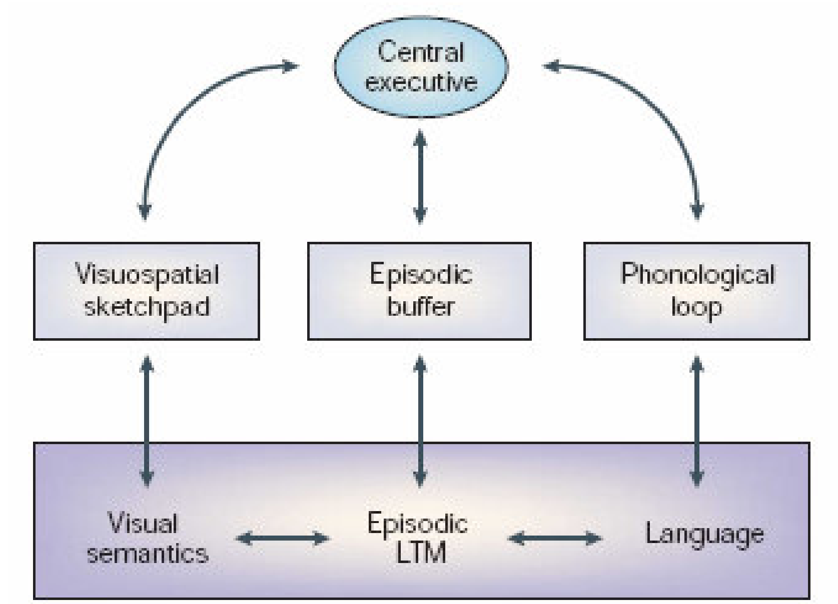


Figure 1 Working Memory Model (Adopted from Baddeley, 2003, p.835)

The first one of the slavery components, the phonological loop, consists of a phonological store that holds memory traces for a few seconds before they fade away, and an articulatory rehearsal process that is analogous to subvocal speech (Baddeley, 2003). It only holds auditory traces and provides the central executive with these traces for auditory imagery. Baddeley, Gathercole and Papagno (1998) state that the phonological loop's main function is the facilitation of language acquisition and auxiliary functions.

Next, the episodic buffer provides an interface between the sub-systems of working memory and long-term memory (Baddeley, 2003). The central executive in working memory activates episodic traces in the long-term memory by means of this buffer. Baddeley (2000) states that the episodic buffer is assumed to be a limited capacity temporary storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive, which is capable of retrieving information from the store in the form of conscious awareness, of reflecting on that information, and, where necessary, manipulating and modifying it.

The final component, the visuospatial sketchpad, is strongly involved in diagrammatic reasoning because it temporarily holds conscious mental images under the control of the central executive. How visual information is used in mental imagery and held in working memory has been investigated. Baddeley and Hitch (1974) first proposed a distinct memory component for visual memory for active use of visual information, namely the visuospatial sketchpad. In Baddeley's final model (2003), the visuospatial sketchpad is the specialized part of the working memory which temporarily holds visual information during pictorial comprehension, recognition, mental animation or problem-solving processes. In their first model in 1974, the visuospatial sketchpad was thought to be a passive storage. In the 2003 model of Baddeley (2003), this part of the working memory became more functional in diagrammatic cognitive processes. During a problem-solving action with diagrams, the conscious imagery in the visuospatial sketchpad is controlled by the central executive, which is the functional processing part of working memory. According to Baddeley's (2003) model, when a diagrammatic problem solver receives visual inputs of the problem, he constructs his mental images and stores them in this buffer. When the solver develops a strategy according to a solution, he might need to alter these images in the buffer. The central executive modifies, controls, and even animates these images in the visuospatial sketchpad during the diagrammatic problem-solving action. If this action needs some auditory inputs or past experience from the long-term memory to find a better way of the solution, the central executive interacts with these other components to solve the problem. Moreover, the temporary images kept in the visuospatial sketchpad sometimes need to be refreshed by sensory input, which is accomplished by gaze fixations, to prevent the decay of visual information. The reason is that depending on the task, if a mental image is not actively employed in a task by the central executive, it will slowly decay to make room for the incoming input. Sperling (1963) stated that the duration of visual memory is 0.5 to 1.0 second; and thus, viewer needs to refresh the visual information. For example, in a study examining viewers' eye fixations as they interpreted graphs, Carpenter and Shah (1998) showed that viewers must continuously reexamine the labels to refresh their memory while performing diagrammatic reasoning.

The visuospatial sketchpad is very important in thinking with diagrams or employing diagrams in problem-solving. For example, Wilson, Baddeley and Young (1999) have a patient with a severe visual short-term memory deficit. Their patient has a very poor performance on a visual recognition memory test, namely the Doors test (Baddeley, Emslie & Nimmo-Smith, 1994), and on retention of checkerboard patterns (Della Sala, Gray, Baddeley & Wilson, 1997). However, the patient could still draw the complex figures she remembers that did not rely on her working memory, but on her long term memory. There are many other similar studies on the role of visual memory for employing diagrams by Hanley and Davis (1995, as cited in Logie & Della Sala, 2005, *p*.85), Luzzatti, Vecchi, Agazzi, Cesa-Bianchi and Vergani (1998, as cited in Logie & Della Sala, 2005, *p*.85) and Carlesimo, Perri, Turriziani, Tomaiuolo and Caltagirone (2001, as cited in Logie & Della Sala, 2005, *p*.85). These studies show that the visuospatial sketchpad is the vital part of the working memory for employing diagrams in comprehension, perception and problem-solving tasks. In this thesis, the diagram utilization in problem-solving is specifically studied. In the following section, the use and aspects of diagrams in problem-solving is introduced.

2.2.2 Problem Solving with Diagrams

People often draw diagrams to solve problems because sentential explanations are sometimes not satisfying for a problem-solving task, as the famous proverb sometimes holds, "A picture is worth a thousand of words." For example, in a single image, pictorial representations provide us with colors, shapes, locations, causality, logical chains, potential actions, intentions, topological or geometric relations and such, which might be highly required depending on the nature of the problem while it might require extensive sentential explanations and memory load to provide the same information. Since some of this information cannot be given by sentential representations in such an easy way, diagrams aid people in the problem-solving task. Research has been reported in cognitive psychology as well as artificial intelligence on the roles of diagrammatic reasoning in human problem-solving. Larkin and Simon (1987) state that one important advantage of diagrammatic representations is that it makes explicit spatial relations that might require extensive search and numerous inference steps to detect a symbol. In other words, a pictorial representation often replaces an inference problem by a recognition problem. Chandrasekaran and Narayanan (1990), Novak and Bulko (1990) also point out the usefulness of diagrams to human problem solvers as a device to aid in visualization, thought experiments or predictions. Diagrams also aid us in the selection of appropriate method to solve a problem as the organizers of cognitive activity (Novak & Bulko, 1990). In some cases, combination of diagrammatic and sentential reasoning might work better. Hegarty (2000) observes that the problem solvers that are allowed to take notes on diagrams to be solved are more successful than pure diagram solvers because the former compensate for limited spatial working memory resources. In other words, the people that are allowed to take notes take advantage of reducing their memory loads because they skip some inner pictorial steps and create verbal landmarks, instead. Thus, she claims that the diagrammatic reasoning ability is limited to the capacity of the visual working memory because the people whose memory load is reduced are more successful diagram solvers in the study.

The success of diagrammatic problem solvers, who are subjects solving problems through diagrams, is also being discussed. It is known that the nature of problem and visual properties of the artifacts are important for diagrammatic problem solvers. Yoon and Narayanan (2004) investigate the predictors of success for diagrammatic problem solvers. For this purpose, they ask their participants to solve some physical problems that are graphically represented. During the task, eye tracking data are also collected. They compare the fixations, response time, causal order of processing, and gaze durations of successful and unsuccessful participants. According to Yoon and Narayanan (2004), spending more time on the task and visually attending to more components do not necessarily lead to success in mechanical reasoning from device diagrams. On the other hand, considering more component pairs that are causally or logically related and attending to longer causal chains of components can lead to better accuracy. Concentrating on critical components for relatively longer time also appears to improve accuracy in problem-solving. Critical components can be defined as functional and logical roles and actions[1] defined in diagrams (Cheng, 1996). Increased shifting of one's focus of visual attention from component to component during problem-solving is marginally and positively related to accuracy, but clearly which components are attended to and in which order are more significant predictors. They claim that successful problem solvers exhibit significantly longer periods of gaze duration and solution time, indicating that the consideration of more causal links along longer lines of action, does indeed suggest that a display that facilitates reasoning along the lines of causal action can enhance comprehension.

Additional to studies that characterize what kind of properties of diagrams elicit more successful problem-solving, there are some studies on how diagrammatic problem-solving differs from symbolic problem-solving. The current symbolic reasoning uses only symbolic forms of representation such as logical axioms, frames, semantic nets, and such. Yet, in a problem-solving activity through diagrammatic reasoning, the information represented explicitly in a symbolic method is not necessarily what is explicit in a picture.

[1] The functional and logical roles and actions are the representations of mathematically definable entities (like loops, conditions and animated components) or logical entities (like and-or-not-exclusion-etc.).

The information represented explicitly in the symbolic data structure is not the type of information that is explicit in diagrams. Moreover, the operations that are permitted on the symbolic data structure are those humans can perform easily with diagrams through mental imagery or actual modification of the drawing with the control of the central executive.

Similarly, Iwasaki (1995) claims that the fundamental problem of people's reasoning about commonsense spatial problems does not fit well with the model of reasoning process with purely symbolic representations. If the pure symbolic reasoning was enough, people would not require diagrammatic representations sometimes. Architecture of a system that can solve spatial problems should include two separate representations for pictorial and symbolic information as well as separate mechanisms for manipulation of such information (Iwasaki, 1995). Her claim supports the role of the central executive and the visuospatial sketchpad in terms of a differentiation into computational and specialized units for holding pictorial and symbolic data and utilizing them during problem-solving with diagrams. Iwasaki (1995) categorizes diagrammatic problem-solving mainly into geometry problem-solving, reasoning about static and dynamic physical problems, like architectural or engineering sketches, and reasoning about non-physical problems, like software engineering representations. In any field, one must have sufficient information about what is represented by the diagram in order to perform diagrammatic problem-solving. By the help of such studies on diagrammatic problem-solving, some models and computer applications are developed.

2.2.3 Diagrammatic Reasoning Example Models and Applications

Research in diagrammatic reasoning needs models and applications in order to understand human thought processes and the role of diagrams in human cognition. To begin with, Qin and Simon (1990) investigate the cognitive roles of diagrams in a model. They report on their exploration of the role of diagrams in understanding scientific concepts as cognitive artifacts. In their research, Qin and Simon (1990) analyzed 6 graduate or undergraduate university students as subjects on their understanding of sections of Einstein's original paper on special relativity, which was published in 1905 with no diagrams. The subjects were asked to read the material and draw images to make it understandable for themselves. The subjects produced some complete or incomplete images during understanding the concepts and deriving the equations in Einstein's paper. Qin and Simon (1990) found close correlations between the quality and accuracy of diagrams that subjects drew and their understanding in Einstein's paper. In other words, while reading Einstein's paper, if some of the subjects drew more expressive, meaningful and incremental representations, which are the clues of the images already in their working memories, they presented better understanding and equation derivation from the paper. They state that mental models[1] consist of the relevant knowledge already held in subjects' memories that shapes their attempts to form new representations needed in problem-solving. These new representations are produced incrementally in the working memories of subjects.

Kosslyn, Seger, Pani and Hillger (1990) discuss the uses of mental imagery in everyday life. They employed 12 undergraduate subjects in keeping a booklet of diary in their daily lives to report imagery whenever used. Later they discussed the images with the subjects. Kosslyn et al. (1990) find that relatively few images were reported to be used in the service of what we took to be the primary purpose of imagery, that is, recall and mental simulation. Kosslyn et al. (1990) explain that most of the images the subjects reported had no recognizable purposes and were not part of a sequence. If this finding is compared with Qin and Simon's (1990)

[1] Mental models are thought to be internal representations in the visuospatial sketchpad derived from external representations by visual information (Chandrasekaran et al., 1995).

statements, it seems that when imagery is used in problem-solving, the behavior is different than when it is used in everyday life because in Qin and Simon (1990)'s study, the images are incremental, sequential and related to the problem to be understood or solved. While deriving Einstein's equations, imagery is used with obvious purpose: to simulate functions in a process and identify its type. On the other hand, in daily life, diagrammatic perception is not sequential and incremental. These studies point out that the humans use diagrammatic reasoning in a problem-solving context incrementally.

In another study, Koedinger and Anderson (1990), use the verbal protocols of 4 expert problem solvers as their basis for formulating a model of human skill in the domain of geometric proof problems. They report that unlike the previous studies of geometric problem-solving, experts tend to focus on key portions of the diagrams, and that they tend to skip steps. Their geometric problem-solving system derived from the experts' verbal data parses geometry problems in perceptual chunks that dramatically cut down the proof search space. Similarly, Ligozat (1999) makes use of diagrammatic reasoning in geometry and topology. He uses the geometric and topological diagrams reported in the literature as reasoning artifacts to develop a computer application performing proofs in terms of three steps: construction, inspection and interpretation. He claims that during construction, inspection and interpretation steps, diagrams provide focus points and help people to narrow down the search pace. Therefore, diagrams are useful in reasoning because the incremental diagrammatic artifacts narrow down the solution space.

As an example of computational models[1], Glasgow and Papadias (1992) present hierarchical models for people using diagrams for understanding or thinking, and explore the use of these models for presenting and reasoning

[1] The term, "Computational model" is used for a computerized application derived from a cognitive model framework.

about images of scenes, maps, and objects. They emphasize computational efficiency and the appropriateness of representations for supporting different types of reasoning. In contrast, Rogers (1995) presents a model of human visual problem-solving, concentrating on the interactions between bottom-up, which is visually driven processing, and top-down, which is expectation-driven processing. Rogers' ultimate goal is the design of better interfaces and support systems for human problem-solving. Rogers shows how her model can be used to represent protocols of radiologists reading chest x-rays.

In another study, Narayanan, Suwa and Motoda (1995) study inferring behavior of mechanical systems from diagrams and develop a computational algorithm for reasoning together with the help of verbal protocols. Narayanan et al. (1995) are concerned with the interaction between bottom-up and top-down processes and problem-solving inferences from diagrams, like Rogers (1995). They explain a model that makes use of verbal protocols and features of image representations. Their model's algorithm combines both rule-based and diagram-based reasoning.

Hegarty (1992) studies how people infer kinematics of mechanical systems through mental animation by the help of reaction-time and eye-fixations and uses these data for computational models. In an experiment on the analysis of pulley systems, Hegarty (1992) finds that eye movements and reaction times suggest that subjects solve these types of problems by piecemeal operations, and that the sequencing of these operations is isomorphic to the causal sequencing of events in the pulley system, and that errors increase as the causal chains lengthened. Hegarty (1992) concludes that eye-fixations on the visual display might relieve the subjects of the need to store a static representation of the system in visual memory, so that more mental capacity can be devoted to the animation. In a similar study, Hegarty (2000) discusses the capacity limits in diagrammatic reasoning concerning mental animations. In Figure 2, a sample mental animation item is given. The question is if the handle is turned in the direction shown, in which direction (A or B) will the box turn?

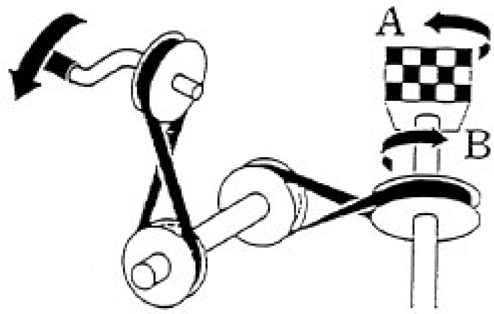


Figure 2 A Sample Mental Animation Item (Adopted from Hegarty, 2000, *p*.195)

In her study, Hegarty states that diagrammatic reasoning and mental animations depends on visuospatial working memory resources (Sims & Hegarty, 1997; Hegarty, 2000). She implements her model in the 3CAPS architecture developed by Just, Carpenter and Hemphill (1996). She concludes that it is possible to simulate human diagrammatic reasoning and mental animation tasks in computer environments regarding the memory capacity.

In another study, Anderson and McCartney (1995) describe how computers can perform diagrammatic reasoning in a detailed way on a battleship game, which is played on a grid paper, in which each square of the grid can be data structures of the diagram. In a similar study, Anderson and McCartney (1994) also define binary operators and functions on this set of grids. After these definitions, they simulate how computers can perform diagrammatic reasoning using such functions, definitions, operators and data structures to play a battleship game. For example, Sawamura and Kiyozuka (2000) combine a computational diagrammatic model with sentential forms, and introduce a computational visual reasoning system with diagrams and sentences, JVenn software. JVenn makes use of Venn diagrams in reasoning.

For example, it translates the shaded intersection areas of Venn diagrams into the formulas in the form of letters. As another example of geometric problem-solving by diagrams, McDoughal and Hammond (1995) describe a computer program POLYA that writes proofs of high school geometry problems by taking advantage of reusing solutions from past problems (Schank & Abelson, 1977; Kolodner, 1993). It applies previously defined sentential proof steps into a newly defined geometry problem to reach a conclusive proof. Finally, Decuyper, Keymeulen and Steels (1995) study the same topic for problems solving in physics.

These studies and applications exemplify the use of diagrammatic reasoning. Although this thesis does not aim to model or develop an application of diagrammatic reasoning, these studies claim that the interaction between sentential forms and diagrammatic expressions, the verbal data, and the eye movement data are the main source for studying diagrammatic reasoning. In this study, their combination is used to improve the understanding of the topic.

2.3 Diagrammatic Reasoning and Software Engineering Diagrams

Capturing, conveying information and utilizing it efficiently is very important in software development because software is a computational simulation of a real-life phenomenon or problem. In software engineering, it is very common and almost compulsory for large systems' designers to prepare diagrams during the development phase. The overall modeling of the system is required to provide developers with a modular view.

Entity-relationship diagrams and data flow diagrams are common artifacts for visualization of the information in software to be developed. Entity-relationship diagrams (ERDs) display objects, their relationships, conditions and hierarchies under certain conventions (Schiffner & Scheuermann, 1979).

Similarly, data flow diagrams (DFDs) represent how data flow throughout the software to aid in development of the software (Guo, 1997). As the scale and complexity of software systems increase, more complex and more representative diagrammatic notations are required. Sometimes pages of sentential explanations, conditions, relations and objects need to be displayed in a single composite diagram with all semantic properties. Therefore, the increased complexity of new diagramming techniques has created different methods and languages such as object oriented graphical design languages. Unified Modeling Language, UML, is the most common example of object-oriented software development languages (Rumbaugh, Jacobson & Booch, 1998). In object oriented software development, the abstract classes of real-life objects and entities are created as the encapsulations and classifications of information. The functions, interactions and flows are coded with respect to these object classes.

Object-oriented representations have properties that are thought to facilitate the communication among those involved in collaborative development tasks (Rosson & Alpert, 1990), and there is some evidence that the use of object oriented representations reduces the need for clarification in the discussion of design ideas (Herbsleb, Klein, Olson, Brunner, Olson & Harding, 1995). Such representations are claimed to clarify the discussion of design ideas because UML is a language that models entities or objects to be coded with graphical and textual notation (Rumbaugh Jacobson & Booch, 1998). By combining text and representations, Unified Modeling Language is used for specifying, visualizing, constructing, and documenting the artifacts of software systems diagrammatically.

Using software engineering diagrams can be studied from a cognitive science point of view because they represent knowledge and relations and they go through some mental processes of developers or users, like understanding, perception, translating, reasoning, integrating and modifying during software development. Although there is an official agreement on some basic standards of object-oriented design notations (Rumbaugh Jacobson & Booch, 1998), there has been hardly any work about validating the cognitive aspects of such diagrammatic notations. Though Eichelberger (2003) discusses the aesthetic and intuitive properties of UML and Blackwell et al. (2001) propose a cognitive framework for notations, they do not experimentally support their views. In their studies, they do not explain their cognitive and aesthetic aspects in an applicable and experimental manner in software engineering. Yet, they employ some findings of other previous studies to support their claims. In other words, experimentally supported studies of software development artifacts from the view point of cognitive science and psychology are required.

There are some experimental studies on software notations performed. For example, Hahn and Kim (1999) study diagrammatic reasoning in a diagram decomposition and integration task with design experts. Their subjects' problem is the integration of different software diagrams to get a unified diagram which can be coded into computer programs. In this problem, subjects are more successful in the integration process if diagrams are more decomposable and have high causal relationships, which means components are causally bounded to each other with 'if' structures. Similarly, the decomposed and linked information representation was expected to provide effective visual cues between different diagrams to understand, use, and search the system (Narayanan, Suwa & Motoda, 1995). This analysis behavior, decomposing, might also have reduced the working memory load because the critical information in integrating the diagrams was more likely to be readily available.

Similarly, Kim, Hahn and Hahn (2000) study the cognitive integration process of multiple diagrams composed of different classes of diagrams, which can be developed by UML specifications. They ask their subjects, which are university students, to integrate these graphs perceptually, i.e. in a way easy to understand, and conceptually, i.e. in a way easy to categorize. The subjects' perceptual processes are analyzed based on their diagram transition between the graphs. The subjects' conceptual processes are analyzed based on their behaviors. Kim et al. (2000) conclude that the use of more diagrams does not necessarily increase analysts' understanding of the target system unless the representation of the diagrams aids in problem-solving. The same conclusion is derived from Green, Petre and Bellamy (1991)'s study. Kalyuga, Chandler and Sweller (1997) state that the grammatical rules and the logical structures, like *and-or-not*, that combine the components of diagrams have been found to influence reasoning with diagrams more other than the layout organization (e.g., color coding of the graphical primitives and spatial orientation).

In another study, Gallant (2002) investigates diagrammatic reasoning with statecharts. Statecharts are software and system engineering artifacts that are used to represent the state of a system with textual and diagrammatical notations. Carrying out some experiments with some novice domain-familiar subjects, he concludes that more understandable design models supported by cognitive science will strengthen the persistence in long term memory for designing with statecharts. Gurr and Tourlas (2000) study the sequential function charts and reasoning in software engineering. Unlike Gallant (2002), they propose the use of natural and intuitive representations, which have artistic and aesthetic concerns. They claim that aesthetic designs improve the software notations. They prepare natural and intuitive sequential function chart representations and try to formulate them. However, their terms, "natural and intuitive" are subjective and immeasurable aspects of representations. Since they are aware of this inability, Gurr and Tourlas (2000) suggest that a more concrete framework for more cognitively plausible engineering representations should be developed and diagrammatic languages should be created upon them.

Hungerford, Hevner and Collins (2004) study software diagrams because reviews and inspections of software artifacts throughout the development life cycle[1] in terms of cognitive science are effective techniques for identifying defects and improving software quality. They study 12 experienced software engineers as subjects. The subjects are given entity relationship diagrams and data flow diagrams with deliberate defects. While they identify the defects, verbal protocol analysis is employed to find out their diagrammatic problem-solving strategies. They compare concurrent and non-concurrent search techniques between diagrams among subjects. Concurrent search technique means switching between different groups of diagrams while searching and understanding diagrams whereas the non-concurrent search technique means sticking to the same group in diagrams. There are different diagrams with errors that are related to each other and some subjects concurrently move among these related diagrams to find errors but some of the subjects do not. After the experiments with the experts, Hungerford et al. (2004) conclude that combining search techniques will cover more defect classes and therefore improve defect detection software engineering diagrams. Hungerford and his colleagues (2004) study inspired this thesis. Defect detection and verbal protocol analysis are common properties of this thesis and their study. They try to relate expertise and search strategies in the success of defect detection. However, each component representation and the connection rules of the components have importance in error-finding besides the search strategies. Therefore, eye movement data is collected to have a better understanding of diagrammatic reasoning. Moreover, most of the software systems are developed for different expertise fields. Software developers mainly need to work with the people that are familiar with different domains. Thus, for a software engineering task, the comparison of the software developers and the people from other domains working on the same problem would provide better understanding of diagrammatic reasoning in the software engineering domain. Consequently, the defect detection in software engineering diagrams are studied with the help of

[1] Software development lifecycle refers to the stages of development of software from the analysis to the end product.

verbal reports and eye movement data of two different groups working on the same problem.

2.3.1 Conceptual Modeling

In this study, software engineering diagrams from conceptual modeling are employed. Therefore, it will be useful to have some introductory knowledge on conceptual modeling. A computer simulation is a computer program which attempts to simulate an abstract model of a particular system. Simulations are powerful parts of modeling in physics, chemistry, biology, human systems, economics, engineering technologies, and such, to provide deeper insight into the operation of those systems. A simulation conceptual model is the simulation developer's way of translating modeling requirements (i.e., what is to be represented by the simulation) into a detailed design framework (i.e., how it is to be done), from which the software, hardware, networks (in the case of distributed simulation), and systems/equipment that will make up the simulation can be built (Pace, 2000; Pace, 1998). Conceptual modeling is mostly used in the military domain of simulations. The following figure (DMSO, 2000) represents the components of conceptual modeling.

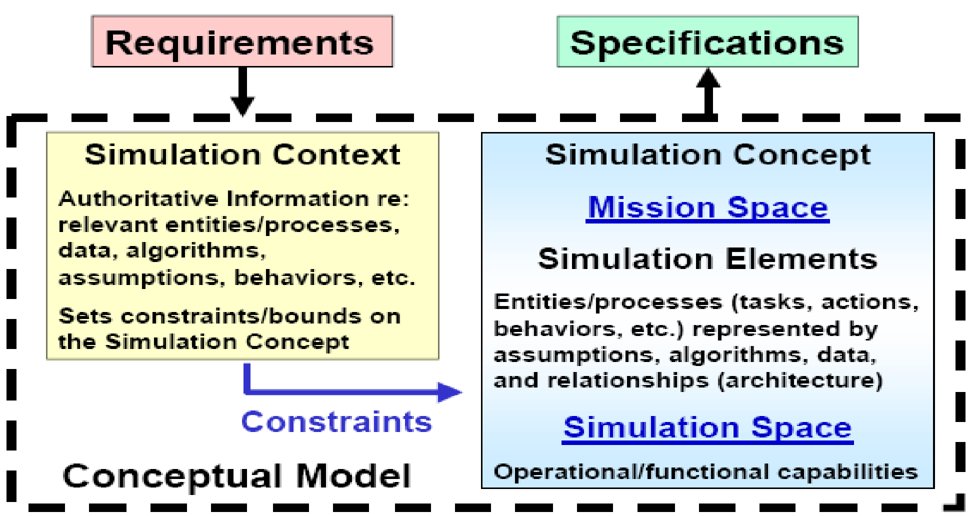


Figure 3 Conceptual Model Components (Adopted from DMSO, 2000)

As can be seen in Figure 3, simulation context provides authoritative information about the domain in which simulation elements to be built. On the other hand, simulation concept describes the developers' concept for the entire simulation. Simulation concept is divided into two spaces: Mission Space and Simulation Space. Mission space is concerned with the representation and Simulation Space is concerned with simulation control. The USA Defense Modeling and Simulation Office (DMSO) defined a conceptual model as user's and developers' common view to an object and to its functions (Pace, 2000). DMSO also proposed High Level Architecture for Federation (simulation) Development and Execution Process[1].

The USA DMSO's conceptual modeling standards and projects are mainly for C4ISR simulations (Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance). C4ISR is a modeling standard for defense systems which aims to improve interoperability between the systems (land, air, sea and space) by computerizing them. In this thesis, the conceptual modeling notations of a C4ISR simulation project, namely KAMA[2], are used. KAMA aims to meet C4ISR needs of the Turkish Army; for example, weapons, missions, military rank and such will be computerized conceptually and operate with other entities from other subsystems (KAMA, 2005).

In order to see how a conceptual modeling notation stands as a software engineering diagram, we have to explain model development in conceptual modeling. A conceptual model is produced by a language, and a language is originated from a meta language as described in Figure 4.

---

[1] This architecture has become a standard for conceptual simulation development practices as suggested by IEEE.
[2] KAMA is "Kavramsal Model Oluşturma Aracı" in Turkish.

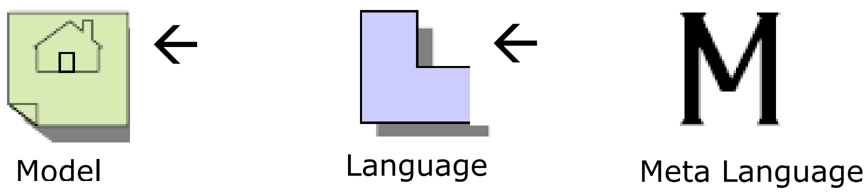


Figure 4 Model, Language and Meta Language (KAMA, 2005)

A meta language (in this case, it is the UML-derived notation) is used to represent all of the attributes and methods of the objects as in real-life. In other words, it creates an abstraction of an object with some graphical and textual information, which is meta model, to help developers to grasp the object to be modeled. Then, a programming language is used to translate the object from the meta model to programming language lines to be executed by computers. The final product is a computerized conceptual model. In this thesis, conceptual modeling meta language notations of a scenario are studied in the experiments.

In KAMA, UML is used as model development meta language as in Figure 4. UML is a language that models entities with graphical and textual notation (Rumbaugh et al., 1999). UML has been observed to be insufficient for modeling in some aspects of the categories of mission space; therefore, it is extended when required. This thesis only makes use of the KAMA tool's representations and sample scenarios[1]. How these materials are developed and in which methodologies are used in the study will be explained in the following chapter.

[1] Part of the project involves the development of a tool for representing and visualizing conceptual modeling in a computer environment but this tool has been left out of the scope of this thesis.

CHAPTER 3

MAIN EXPERIMENT

3.1 Methodology and Design of the Main Experiment

In this chapter, the experiment's methodology and experimental design of the main experiment is explained. A background of eye movement data analysis is given at the beginning. Then, verbal protocol data analysis is introduced, and the combination of these methods are given as the methodology of the study. The Experimental setting and how the experiment is conducted are explained. Finally, experimental data preparations before the analysis are given. The results and the discussion of the main experiment are given at the end of this chapter.

3.1.1 Background on Eye Movement Data Analysis

Eye movement studies have been popular and carried out since 1970s. The human eye covers a visual field of about 200°, but receives detailed information from only 2° (Levi, Klein & Aitsobomo, 1985). This tiny high-resolution field of the retina is called the fovea. During visual perception, visual information is acquired from a limited spatial region surrounding the center of gaze, the fovea. The visual information from the world enters through the fovea and is transported to the visual system, which is the oculomotor system. It is the only way of gathering visual information. In cognitive science, eye movement data are crucial if the domain is related to the visual system and visual information processing.

There are several types of eye movements. Fixations are the rapid eye movements which are directions of gaze referring to the number of focuses to a specific area. Saccades are the most common way of moving the eyes in a sudden, ballistic, and nearly instantaneous period of time (Jacob, 1991). A saccade is an instantaneous gaze duration typically followed by a fixation. Theoretically, saccades are instants between the fixations. However, in most of the studies saccades and corresponding fixations are usually filtered by some software with 40msec or 100msec, or not all of them are related to the aims of researches. In this study, we did not filter the fixations because we assumed that all fixations for searching or thinking are related to the task. The total durations of saccades for a specific period of time or region can be grouped as the gaze duration information. Other classes of information, such as smooth pursuit, vergence, optikinetic nystagmus, torsion and micro-saccades are also explored to understand the oculomotor system (Richardson, Dale & Spivey, 2007). The latter ones are not covered in this study because they are hard to analyze, claimed to be noisy data, and require neurobiological studies and equipment, too.

Since eye movement data are considered to be the way to the visual system, these data can be regarded as "window to the mind" (Johansson, 2004). Eye movements are considered to be useful for learning about on-line cognitive process of diagram-based problem-solving because eye movement provides problem-solving measures that solution time and accuracy cannot address (Grant & Spivey, 2003). For example, early studies show some evidence that eye movements can correspond to inference making (Hunziker, 1970; Lenhart, 1983; Nakano, 1971, as cited in Grant & Spivey, 2003). Recently, eye movements are studied to identify problem-solving in geometric reasoning (Epelboim & Suppes, 2001), reasoning about mechanical systems (Hegarty, 1992; Hegarty & Just 1993), insight problem-solving (Knoblich, Ohlsson & Raney 2001; Grant & Spivey, 2003), image scanning (Noton & Stark, 1971, as cited in Salvucci & Goldberg, 2000), arithmetic (Suppes, 1990), real-world scene perception (Henderson, 2003), reading (Rayner, 1995), language (Richardson, Dale & Spivey 2007; Rayner, 1998; Marian & Spivey, 2003; Farretti, McRea & Hatherell, 2001; Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995), and human-computer interfaces (Ellis, Candrea, Misner, Craig, Lankford & Hutchinson, 1998).

Eye movements not only inform us abut visual perception, but also indicative about visual imagery and intention. For example, Spivey and Geng (2001) record subjects' eye movements while they listen to spoken descriptions of spatiotemporal dynamic scenes in front of a blank screen. The subjects' expectations and imaginations affect their eye movements as if they were looking at the real scene instead of the blank one. Instead of purely using an internal visuospatial sketchpad (Baddeley, 2003), the subject employs an external sketchpad as an additional layout of the auditory input. Similarly Rayner (1998), Tanenhaus and Trueswell (1995) analyze the eye movement patterns of the subjects reading a text to see how cognitive processes related with pronoun resolution, responses to word frequencies, lexical ambiguity resolution, and semantic and discourse context affect their eye movements. In another study, Spivey-Knowlton, Tanenhaus, Eberhard and Sedivy (1998), provide their subjects with a set of objects like a candle, a bag of candy, and a spoon, and asked them to pick up the bag of candy. Spivey-Knowlton and colleagues observe that the subjects fixate the candle unconsciously before the candy is fixated. The eye movement data in Spivey-Knowlton et al. (1998)'s study is affected by similar sounding object names. In brief, interaction between eye movements and cognitive domain does not happen in a one-way fashion. On the other hand, diagrammatic reasoning usually focuses on how eye movement data is deliberately chosen by the attentional system to solve a problem.

In some studies (Hegarty, 1992, 2000; Sims & Hegarty, 1997), the eye movements of the subjects looking at mechanical systems are studied and it is observed that the subjects' attention is related to the fixations and they fixated to the animated, critical or causally linked components while understanding the system. Grant and Spivey (2003) and Just and Carpenter (1985) state that there is a positive correlation between eye movement and the problem-solving process; and that attention guides the eye movements. Grant and Spivey (2003) also state that if a solver's attention shifts to a critical diagram component, i.e. if they fixate to a critical component, they usually solve the problem successfully. Knoblich, Ohlsson and Raney (2001) describe insightful problem-solving via eye movements. They state that an initial period of purposeful problem-solving activity is followed by an impasse, a state of mind in which the problem solver feels that all options have been explored and he or she cannot think of what to do next. They state that during diagrammatic problem-solving, the attentional mechanism is in a search state and many fixations of shorter durations are observed. When the impasse occurs, the number of fixations decreases whereas the durations of these fixations increase. Attention guides eye movements to the critical components, and then resolution of the impasses leads to successful reasoning. These studies claim that there is a strong relation between attentional system and eye movements. However, the relation of eye tracking data to attention-driven cognitive activities like reasoning, searching, and understanding is a still debatable issue that requires more studies.

Since attention plays an important role in visual and cognitive processing, and especially in diagrammatic reasoning, and eye movements indicate the overt behavioral direction of attention in a piece of a diagrammatic element, eye movements also provide information about the attentional system (Findlay, 2004). To understand the phenomenon of visual attention, it is important to distinguish between "where" and "what". Where we look is not always the same as what we look at. For example, it is possible to receive a whole image of a certain object through the fovea, but also to have the attention at a peripheral object, or a certain component of the object. This attentional dichotomy is commonly called *overt* and *covert* attention, where overt attention corresponds to foveal attention and covert attention corresponds to parafoveal attention (Johansson, 2004). This dichotomy is particularly relevant when explaining how we select our attention during diagrammatic problem-solving, and especially in a *bottom-up* explanation (Wolfe, 1998). In a bottom-up explanation, the visual attention selection mechanism can be said to consist of two stages: Initially, we have a pre-attentive stage ("where" to look) and later an attentive stage ("what" to look at). The pre-attentive stage is working in parallel across the entire visual field (parafoveal), however, the attentive stage is limited and can only handle one object at a time (foveal). When an object is processed from the pre-attentive to the attentive stage it is considered to be selected. If this object is a critical component of the diagram, it will most probably help in reasoning. Moreover, this attention selection also means that visual attention is overtly shifted to a new location before the saccade occurs. This selection process of attention has been explained by a number of different metaphors. One of the most common metaphors is Posner's "spotlight", (Duchowski, 2003). The spotlight-metaphor suggests that the attentional mechanism moves in the same manner as a spotlight and that the object in the spot is what we attend to. On the other hand, the top-down explanation of attention perspective states that the interesting features are deliberately selected by a certain interest, i.e. user-driven, and not by some kind of pop-up effect (Wolfe, 1998). This voluntary and task-dependent attention is present when people look at pictures. In other words, they look at them differently with respect to what they are looking for and what their interests are (Yarbus, 1967, as cited in Johansson, 2004).

Besides the discussions about the attentional system and the eye movement data, another issue is how to interpret the mass of eye tracking data. Even for a small study, eye tracking devices can provide thousands of fixations and saccades. Salvucci (1999, 2000), Duchowski (2002, 2003), Wooding (2002), Salvucci and Anderson (1998, 2001), and the designers of some eye tracking software[1] aim to propose some methods and applications for handling such a large amount of data. For example, in this thesis, Clearview® software is used for grouping the eye movement data according to areas of interest[2] and verbal protocols are used for manual categorizations.

Whether the attentional selection mechanism during reasoning is a bottom-up or top-down and despite the massive amount of data, eye tracking fixations and gaze durations are essential data to understand visual cognitive processes like diagrammatic reasoning. To understand a diagram or solve a diagrammatic problem, fixations are made to focus on the critical components, and gaze durations are required to resolve impasses and reach solutions or validations. The difficulty of working with eye fixations is the large number of data, which can be compensated by verbal protocol analysis.

3.1.2 Background on Verbal Protocol Analysis

Thought processes can be described as a sequence of states, each state containing the end products of cognitive processes, such as information retrieval from long-term memory, information perceived and recognized, and information generated by inference (Ericsson & Simon, 1993). The information in a state can be input to a verbalization process and reported orally. There are many complex processes in which the outcome information

[1] Such as Clearview® and The Observer® software.

[2] The area of interest is used as a boundary of eye movement data on a specific region defined by the Clearview® software.

of thinking does not emerge in an observable action. Obviously one of the ways of getting the information is to ask people to “think aloud”. These verbal reports are called verbal protocols (Ericsson & Simon, 1993; Bainbridge & Sanderson 2005).

Verbal protocols that are collected during a task’s duration are called concurrent verbal protocols, which are considered as verbalization of thoughts, retrieved from working memory. On the other hand, retrospective verbal protocols are gathered from long-term memory after the completion of an action. They are considered as verbalization of specific information such as reasons and explanations (Ericsson & Simon, 1993). Ericsson and Simon (1993) state that concurrent and retrospective verbal protocols should be collected and studied for accuracy and completeness of protocols and inferences for a task.

Verbal protocol analysis has been used in research about process tracing[1] (Todd & Benbasat, 1987), knowledge acquisition (Van Someren, Barnard & Sanberg, 1994, as cited in Benbunan-Fich, 2001), model formulation (Goodwin, 1987, as cited in Benbunan-Fich, 2001), thinking and decision making behavior (Schweiger, 1983), computer-aided architectural design (Gero & Tang, 1999), user-interface design (Know, Bailey & Lynch, 1989), usability (Benbunan-Fich, 2001), problem-solving (Heydemann, 1986; Deffner, 1989; Chen, 1999), sorting tasks (Hoc & Leplat, 1983, as cited in Ericsson & Simon, 1993), accounting (Belkaoui, 1989, as cited in Ericsson & Simon, 1993), narrative writing (Randsdell, 1995), and diagrammatic reasoning (Hungerford, Hevner & Collins, 2004; Hahn & Kim, 1999; Iwasaki 1995).

[1] Process tracing shows how a decision is made according to stages of an action (Todd & Benbasat, 1987).

The main question related to the use of verbal protocols as experimental data is, "Do reported activity reflect the actual knowledge and processes of our cognitive faculty?" Since speech itself is an outcome of mental activities, can it provide the researchers with the data of an ongoing process? Can every process be verbalized? Some of these main problems and others are discussed by Ericsson and Simon (1993), and by Bainbridge and Sanderson (2005). Besides these questions, there are some basic problems of verbal protocol gathering. First, when a subject is asked to think-aloud, the task itself and the way the task is done change. For example, if there are alternative ways to perform a task and a subject is asked to think-aloud, he or she might choose the way that can be described easily. The second problem is related to the tasks with time constraints. If people are asked to solve problems while thinking-aloud under time limitations, they may not mention about some of the inner stages of the process because many things may quickly pass through their minds. Third, giving verbal protocols can be socially biased. People can select what is appropriate to say because of social tension. Another problem is related to the tasks that are done in nonverbal ways. For example, perceptual-motor skills like swimming or the manipulation of an image cannot be verbalized by people entirely, or the verbalization of a perceptual-motor skill depends on subject's vocabularies. Finally, especially in problem-solving tasks, memories of past events, skills on functions or causal relations and expertise knowledge about components will affect the process. Yet, verbal protocols may not have to include these data, or such data will require to be inferred by the researcher. In the following paragraphs, these criticisms are discussed.

There is, however, strong experimental and argumentative evidence that verbal protocols are useful if they are used cautiously. First of all, verbal protocols are not claimed to be thoughts themselves but they can be good complementary sources of cognitive information (Ericsson & Simon, 1993). Verbal protocols give explicit information about the sequence of the items considered, and from this, the strategy being used is inferred as well as the working-memory contents (Bainbridge & Sanderson, 2005). For example, when Ericsson & Simon's (1993) famous question, which is the multiplication of 24 with 36, is asked, the concurrent verbalization can be, "36 times 24… 4 times 6… 24… 4… carry the 2… 12… 14... 144," and so on. Subjects are not verbalizing their thoughts while performing the task, and they do not describe or explain what they are doing. They, instead, simply verbalize the information they are busy with while generating the answer. This information already resides in their working memories, and is used as input by their central executive (Baddeley, 2003) and the verbalization process shows how information is processed. Some inference mechanisms should be used to label the protocols[1]. For example, from the verbal data given in the example, one can infer that the subject, firstly, makes use of the symmetric property of multiplication and changes the order of the numbers from 24 x 36 to 36 x 24 so that the multiplier becomes smaller. Then, the subject starts the multiplication process from the rightmost decimal digits; and for later use, stores the carry. The subject does not verbalize how he or she evaluates the result 12, but it is multiplication of 3 and 4. Afterwards, he or she adds the previously carried number, 2, to 12, to get 14 and so on. Thus, one can, for examples, label the possible protocols as "Order, StartRight, MultiplyAloud, Carry, MultiplySilent, UseCarry" and so on. When the orders, occurrences and frequencies of such possible protocols are studied, it will yield to the verbal protocol analysis. However, if the subject is later asked to describe and explain his or her thoughts, additional thoughts and explanation from the long term memory have to be accessed to produce some auxiliary descriptions, which might differ from concurrent verbal data. In this thesis, concurrent verbal protocols are collected, but retrospective verbal protocols are not. Yet, in the interview questions it is assessed whether concurrent verbal protocols are valid in terms of inferences and the statistics.

[1] Protocols are assigned and short name of segments whose frequencies and occurrences compose the verbal protocol analysis.

When subjects are asked to think-aloud, Ericsson and Simon (1993) arguably claim that concurrent verbal protocols do not strongly affect the way in which a problem-solving task is done, but it increases the time required for the task to be completed. They state that this time increase does not result from the change of thought process, but from the time required for vocalization. Norris (1990) examines five groups of subjects taking a test of critical thinking with a silent control group. In other words, besides the silent control group, one group continuously gives concurrent verbal reports, while three groups give immediate verbal reports, which are produced after different periods of time when they are asked why they choose their answer. No differences in performance between any of these groups are found, except time. Similarly, Randsdell (1995) compares think-aloud protocols of a group of college students with a silent control group, both working on the narrative writing of college students. She reports no difference in their performance. Biggs, Rosman and Sergenian (1993) compare decision making of a group, thinking aloud with a silent control group about investment-related decision making task. They report no difference in the decision making process. In another decision making study, Biehal and Chakavarti (1989) have subjects choose a pocket calculator from several alternatives on two occasions. The first choice is made among four alternatives, and half of the subjects verbalize their choices while the rest is silent. The second choice is made with additional four alternatives and all of the subjects verbalize their decision making. Yet, before the second stage of choices, the subjects are given a warm-up task that encourages description, justification, and explanation. No reliable difference in decision outcome between the groups is observed for either the first or the second decision. However, when the verbal protocols of the second choice are analyzed in a detailed way, there are reliable differences between the first and second groups thinking-aloud because a well-designed warm-up activity makes the subjects provide more information.

Bowers and Snyder (1990) compare 48 subjects working on 12 different tasks on computers of a think-aloud group with a silent control group. They report no difference in solutions, number of steps and erroneous completions. Sanderson (1990) has subjects work with a mechanical factory system for an hour for five days under silent conditions, and then instructed them to think aloud for 5 days while still working on the same task. No difference in performance is reported. In problem-solving tasks, Schooler, Ohlsson and Brooks (1993) study the effect of retrospective and concurrent verbalization on problem-solving and reported no difference with respect to a silent control group. In another study, Rossi, Daneluzzo, Tomassini, Struglia, Cavallaro, Smeraldi, et al (2006) claim that the verbalization strategy on the Wisconsin Card Sorting Task[1] performance in schizophrenic patients has positive effects. Itoh (2005) studies the effect of verbalization on recognition of faces and concludes from the experiments that the verbalization process interferes with subsequent face recognition only in terms of recognition time when memory for the target person is good, whereas verbalization enhances recognition when memory is poor.

These studies and discussions show that verbal protocols can be used as a source to understand cognitive activities because verbal protocols show how information is used in mental processes. Verbal protocols do not yield the thoughts, but the way of thinking and choices of solutions. It only increases the time for completing a task because of verbalization time. For perceptual motor skills that cannot be verbalized, like swimming, verbal protocol analysis is not appropriate. It is mainly for decision making and problem-solving tasks. In order to enrich the verbal data, it is required to give an appropriate warm-up task. This warm-up task will also reduce the social tension of the subjects. For the social criticism of protocols, which is that the social tension on the subject affects verbalizations, socially-biased verbal protocols are used in interactional sociolinguistics (Schiffrin, 1994). For example, when the task is being done under the control of a superior authority of the participant, verbal reports will be biased.

[1] A widely used test of abstract thinking, planning, and ability to alter mental set as circumstances require (Oxford Reference Online, 2006)

In brief, there are some issues requiring consideration about eye tracking and verbal protocol analysis. Eye tacking is more directly indicative of perceptual and attentional cognitive processes and concurrent verbal protocol analysis gives us an indication of how the reasoning process occurs in a problem-solving task. Note that in our task, there are no time pressures and neither is the task based on perceptual-motor skills. Thus, we used both methods in a complementary way (Schiffrin, 1994). Many tools, like The Observer®[1], and researchers aim to combine different data modalities and methodologies. This is also suggested in the literature. Salvucci and Anderson (1998, 2000) and Salvucci (1999) propose to combine verbal protocols with eye tracking data to enrich and more deeply understand the results. They propose an automated eye-tracking protocol analysis method and Salvucci (2000) introduces a tool, EyeTracer, which identifies fixations with respect to the predefined protocols. In a previous study, Carpenter, Just and Schell (1990) compare a condition of think aloud along with eye-movement recording to a conventional administration of the Raven Progressive Matrices Test[2] and use eye tracking data for protocols. The point is that eye tracking data supports verbal protocol analysis so that labeling of protocols[3] and inferences can be performed more accurately. It further helps, for example, which non-verbalized informational visual elements are used in a specific segment of the protocol. On the other hand, verbal protocols enable researchers to categorize and manage massive amount of eye tracking data. By the categorization of eye tracking data, it is possible to perform qualified statistical analysis. After combining such different data modalities with other information, like performance, solution time, age and experience, it is possible to make statistical tests among different data modalities.

---

[1] The Observer® is software for the analysis of different data modalities like speech, video, eye tracking, and hearth rate (The Observer Reference Manual, 2003).
[2] A nonverbal intelligence test requiring inductive reasoning about abstract geometric patterns (Oxford Reference Online, 2006)
[3] In this study, "protocol" and "label" are used interchangeably.

In this study, a pilot experiment has been performed on 4 participants to observe the effect of verbalization on the specific design of this task. In the pilot study, one domain-familiar subject and one notation-familiar subject performed the task silently. Another domain-familiar subject and a notation-familiar one complete the experiments while thinking aloud. The main difference observed between the verbalizing and non-verbalizing subjects is the completion times, as expected. Number of errors and eye tracking patterns show no significant difference. Moreover, the silent group provides no data other than number of errors found and eye movements. Thus, with protocol analysis, we are able to pinpoint exactly what participants are doing when and where on the diagram, when we synchronize protocol data with eye movement data.

3.2 Setting of the Main Experiment

In this section, first the hypotheses of the main experiment are given. Then, the setting, material and the conducting of the main experiment are introduced.

3.2.1 Hypotheses of the Main Experiment

Although Chandrasekaran et al. (1995) state that familiarity affects diagrammatic problem-solving positively, since we have two different domains of familiarity and there are not enough comparative studies on this issue, the hypothesis is defined as there is no significant difference in error-finding behaviors between two different domain-familiar groups. We need to compare their success rates to analyze the hypothesis. Two groups of people, the domain-familiar participants versus the notation-familiar participants, working on the same problem are studied in order to see the effect of familiarity on diagrammatic reasoning and the cognitive differences of error-finding strategies of the groups.

We also want to explore the question how experience affects the error-finding activity. Since Chandrasekaran et al. (1995) and Hungerford et al. (2004) state that experience increases the success rate of error-finding, we hypothesize that experience improves the error-finding activity of both groups. The correlation between the success rates of error-finding and the experiences of participants is also analyzed in the main experiment.

After the main experiment, it is observed that the success rates in error-finding differ between the groups with respect to certain types of errors. We wanted to explore how and which error types affect the success rate of error-finding significantly. Although it was not hypothesized pre-experimentally, the hypothesis, which is "The success rates of error-finding differ with respect to error types", is tested after the main experiment. Since the initial question is not to test error-finding differences with respect to error types, the number of errors observed in each type is not balanced. However, there are significant differences observed between the groups with respect to error types. The research questions and the corresponding hypotheses of the main experiment are summarized as follow:

Q1: How do different kinds of domain-familiarity affect diagrammatic reasoning?

> **H1:** There is no significant difference in error-finding behaviors of two domain-familiar groups.

Q2: How does experience affect the error-finding activity?

> **H2:** Experience improves the error-finding activity of both groups.

Q3: How and which error types affect the success rate of error-finding significantly?

H3: Success rates of error-finding differ with respect to error types.

The main experiment setting and the materials are given in the following section.

3.2.2 Setting up and Conducting the Main Experiment

In this study, there are two groups of participants. One group is composed of 10 notation-familiar participants, while the other has 10 domain-familiar participants. The demographic data of the participants are given in the Results and Discussions chapter. The participants are asked to work on a scenario (see Appendix A) and diagrams (see Appendix B) belonging to this scenario. The participants are tested individually in the experiment, and their transportation expenses to reach the experiment facility, HCI Lab., are met.

The scenario from the military domain includes the marksman state diagram[1], the stoppage removal activity diagram[2], the marksman mission training activity diagram, and the marksman mission ending and evaluating activity diagram's verbal explanations on a separate document. The marksman state diagram represents the overall states in which a marksman can be in a firing mission. The actions for stoppage removal are represented in the stoppage removal activity diagram. The marksman mission training activity diagram represents how to tutor a marksman in a firing mission.

---

[1] A state diagram shows an object's or an actor's possible states and transitions among states.
[2] An activity diagram is a special state diagram where most of the states are action states and most of the transitions are triggered by completion of the actions.

Finally, marksman mission ending and evaluation of the marksman is given in the marksman mission ending and evaluating activity diagram. These diagrams are derived from Civelek's (2006) graduation project and drawn by Microsoft Visio®.

There are 14 intentionally embedded errors in the diagrams different from the scenario. The participants' task is to read the scenario and compare it with the diagrams and find errors. While finding errors, the participants are asked to think aloud. When an error is found, the participants are required to explain why it is an error.

The experiments are performed in the Human Computer Interaction (HCI) Laboratory at Middle East Technical University (Figure 5 and Figure 6). This laboratory is a medium established to design, utilize, and evaluate interactive technologies like web sites and other computer software. There is a special computer, Tobii 1750 Eye Tracker®, in the laboratory, which is an eye tracking device (Figure 5).

The most important feature of Tobii 1750 is that it is almost an ordinary 17'' TFT computer monitor without any visible or moving "tracking devices". The eye tracking cameras are hidden under optic panels, which never disrupt the participants' attention. Besides eye movement data, it is possible to record the voice and video of the participants in the HCI laboratory. The researcher can control the experiment in the observer room. The Clear View program is used to record and categorize the eye tracking data.

Figure 5 The Experiment Room and Tobii at HCI Lab., METU

Figure 6 The Observer Room at HCI Lab., METU

The experimental material consists of a script with the scenario, the demographic data form, the dictionary of symbols and abbreviations, the error report form, the interview questions for after the experiment and the informed consent form (see Appendix A). The scenario is given to the participant one day before the experiment to have him read the scenario beforehand in order to become familiar with it. Before the experiment starts, the participant is given the informed consent form and brief information. He is also given the dictionary of symbols and abbreviations before the experiment gets started. When the participant is ready, he participates in a warm-up activity (see Appendix A) which is composed of the multiplication of 17 and 12, saying the name of 10 animals and describing the process of tea-making[1]. When the participant is ready, the eye tracking calibration is made to configure the data of the participant's eyes, like distance, radius of gazes, and response time, in order to ensure the quality of the data. This is an automatic process performed by Tobii®. During the experiment, the participant reads the scenario from the paper with one or few lines and compares the lines with the diagrams on the device's screen (see Appendix B). The participant is asked to verbalize every step and whenever he gets silent for more than about 1 minute, he is asked to keep talking. When the participant claims to find an error, he explains it and takes short notes on the error report form. The participant is free to end the experiment whenever he wants to. After the experiment, the interview is performed to collect the participants' comments and support the concurrent protocol. Although we do not collect retrospective verbal reports, the interviews are helpful in the findings, categorizations and inferences of the concurrent verbal reports.

The HCI laboratory's setting allows the eye tracking data of each participant, their speeches, and videos to be recorded. Eye movement data categorized with respect to regions and protocols are required to be synchronized. Synchronization of these data is a challenging task before the analysis (see section 3.3 Data Processing of Main Experiment for Analysis on page 49). Together with other data forms, these are the data for the analysis of the main study.

---

[1] Tea-making process is chosen to make the participants familiar with the verbalization of a process, which is a very common process.

After the pilot study, the main experiment is performed using this combined data collection method. After the analysis, in order to test some properties of the diagrams, an additional follow-up experiment is conducted with 24 participants.

3.3 Data Processing of Main Experiment for Analysis

As the first step of data processing before the analysis for the main experiment, the concurrent verbal protocols' labels are determined as **read_reason**, **find_error**, **explain**, **pause**, **find_non_error**, **accept_as_correct**, **misaccept_as_correct**, **revise_decision**, **comment_on_notation**, **nrs**, **comment_on_scenario**. Before determining the protocols, some sample video segments are coded by two independent coders manually. Then, the codes are compared and discussed. The resulting protocols are refined as such labels.

represents the protocol and their explanations.

Table 1 Verbal Protocol Segments and Definitions

Segment Labels / Protocols	Definitions
read_reason	Refers to the period event[1] in which the participant reads the scenario and performs a reasoning action to verify the information represented or to find errors in the diagrams.

[1] An event that happens in a period of time.

Table 1 (cont.)	
find_error	Refers to the point event[1] in which the participant reports an error after reasoning in an erroneous part of the diagrams.
find_non_error	Refers to the point event in which the participant reports an error after reasoning in a non-erroneous part of the diagrams.
accept_as_correct	Refers to the point event in which the participant verifies a non-erroneous part of the diagrams after reasoning.
misaccept_as_correct	Refers to the point event in which the participant verifies an erroneous part of the diagrams as correct after reasoning.
revise_decision	Refers to the point event in which the participant revises his previous decision after reasoning.

[1] An event that happens in an instant of time.

Table 1 (cont.)	
explain	Refers to the period event in which the participant explains his error-finding after find_error or find_non_error actions.
pause	Refers to the period event in which the participant pauses for a period of time.
comment_on_notation	Refers to the period event in which the participant comments on notations and representations of the scenario.
comment_on_scenario	Refers to the period event in which the participant comments on sentential explanations of the diagrams.
nrs	Refers to the period event, namely non-relevant segment, in which the participant actions are uncategorized or not in the scope of the study. For example, the periods before the experiment starts and interruptions of the experiment are **nrs**.

The periodic protocols have start and end times. On the other hand, the point protocols are instantaneous activities with only a single timestamp. Each participant's video is segmented according to Table 1. Video segmentation process was first started to be made by the help of annotation software, like the HyperResearch and the Observer®. Both products were experimented with. However, due to some usability problems and the limitations on the number of defining new events, the tools were abandoned because the size of the data of a single participant was huge, and the navigations and limitations resulted in difficulties. Then, the segmentation was made by hand with the help of multimedia tools and MS Excel.

After the segmentation of video of concurrent verbal reports according to the protocols defined in Table 1, the diagrams are divided into sub areas. The division of the diagrams is made according to the scenario. Each statement corresponds to a portion of the diagrams. For example, the part shown in Figure 7 corresponds to the portion of the scenario that belongs to the marksman mission training activity diagram as, "At the beginning, the trainer selects the fire group according to the *M*.F.P.C. (Mechanical firing proficiency card) to produce the fire group."

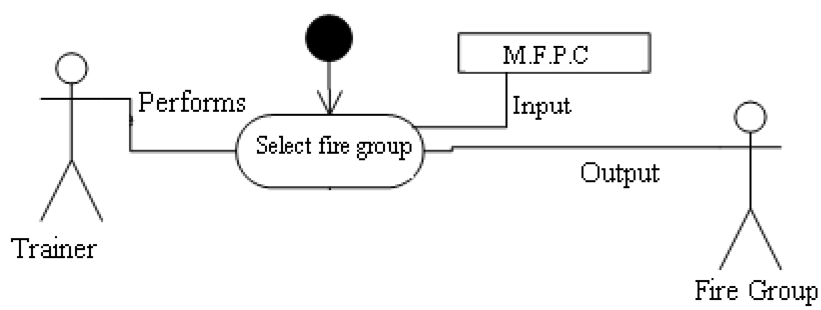


Figure 7 A Sample Portion of the Marksman Mission Training Activity Diagram

In this representation, the trainer is the performer of the action. The *M*.F.P.C. card is the input and the fire group is the output of the activity. The other example in Figure 8 belongs to the scenario portion, "If the cartridge leaves its bed, then the cartridge shooter is in defective state and the marksman changes the cartridge bed."

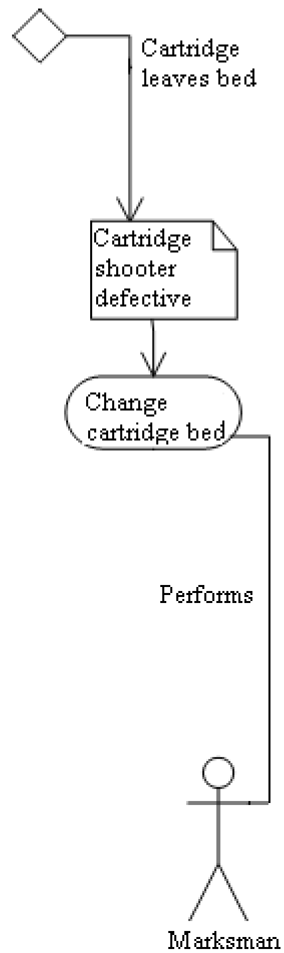


Figure 8 A Sample Portion of the Stoppage Removal Activity Diagram

In a final example from the marksman state diagram in Figure 9, the representation corresponds to the statement, “When the marksman is in stoppage removal stage, if he is unable to remove stoppage, then he changes his state to unload the gun.”

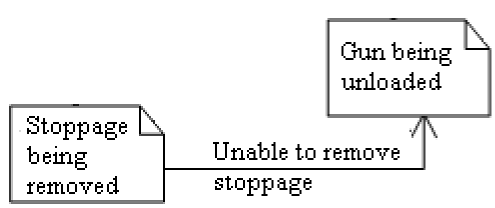


Figure 9 A Sample Portion of the Marksman State Diagram

The examples that are given above explain the division process of the diagrams. The naming conventions of each sub-part also need explanations. The marksman state diagram is named as **p1**, and the stoppage removal activity diagram is named as **p2**. The marksman mission training activity diagram is assigned to **p3**, and the marksman mission ending and evaluating activity diagram is **p4**. In the follow-up experiment, there are diagrams of the motor vehicle tax law and the debt collection and bankruptcy left no partitioned because we are only interested in the sub-sections with the diagrammatic complexity and high degree of causal chaining. Each of the diagrams in the main experiment is partitioned. If a sub-part is a non-erroneous area, it is named with the letter-**a**, and also a number is added as a suffix to the letter according to its order in the scenario. If sub-part is an erroneous area, it is named with the letter-**e**, and again a number is added to it according to the order. Then, the name of the corresponding diagram, p1, p2, p3 or p4, is put as prefix ending with an underscore to the beginning of the name. For example, the first non-erroneous area of the marksman state diagram, which is **p1**, is named as **p1_a1**, the second one is **p1_a2**. Similarly, the first erroneous area of the marksman mission training activity diagram, which is **p3**, is named as **p3_e1**. All representations of sub-parts can be seen in Appendix B. Table 2 shows the distribution of sub-parts among diagrams.

Table 2 Distribution of Sub-parts among Diagrams

Diagram name	**Prefix**	**Number of non-erroneous parts**	**Number of erroneous parts**	**Sum**
The marksman state diagram	p1	8	3	11
The stoppage removal activity diagram	p2	17	3	20
The marksman mission training activity diagram	p3	9	4	13
The marksman mission ending and evaluating activity diagram	p4	9	2	11
	TOTAL			**55**

These areas are also defined as area-of-interests. By the Clear View software of Tobii, one can define such area-of-interests visually. After defining the areas, the software provides the researcher with the data of the total fixation and gaze duration of each participant. The total fixations and gaze durations of each sub-part are analyzed. The results are discussed in the analysis and discussion section.

After receiving the fixation and gaze duration data of each area-of-interest, it is important to synchronize the eye tracking data and the video segmentation with protocols. Synchronization is a highly challenging and very time-consuming task. A program is written in C programming language to do this task within the scope of this thesis. This program receives each participant's video segmentation with starting and ending times of protocols and the eye movement data of each participant, and then synchronizes the segments and eye movement data. Besides the number of fixations and the gaze duration of each area-of-interest, each segment's fixations and gaze durations are identified and classified by the help of this program.

By the help of the methodology defined above, verbal protocol data is combined with eye tracking data to take advantage of both. The eye movement data of segments help in categorization and inference of the protocols. On the other hand, the protocols are very helpful in categorization and management of the thousands of fixations and durational information. For example, it is observed that a participant starts reasoning by reading the lines from the textual scenario, and searches the critical components on the screen related to the scenario. In searching periods, the number of fixations is usually high while the gaze durations are low. When the critical components are found, an impasse occurs with the low number of fixations and the longer period of durations. Then, the participant reaches to a conclusion protocol, like **find_error** or **accept_as_correct**. The period that starts from reading the text from the scenario and searching to the impasse and impasse resolution is named as **read_reason**. Eye movement data helps here to determine the **read_reason** segment even when the participant is silent. The protocol, on the other hand, helps in categorization of huge number of eye movement data. This mutual interaction is combined with the performance data, like time to finish the task, number of errors, and success rate with respect to error types, and with the demographic data, like age and experience. These various data are analyzed by the help of SPSS v.13.

In the analysis, we look if there is a significant difference in groups in terms of protocols. Moreover, tests are made to see if there are differences in terms of error types between the groups. The eye movement data differences with respect to some diagrammatic properties and sub-parts are also tested. Finally, the hypothesis about the correlation between performance data and demographic data is tested. The analysis and discussion sections report the analyses and results related to hypotheses and discuss and evaluate the findings of the main experiment.

3.4 Results of the Main Experiment

For the main experiment, firstly, the demographic results are summarized. Then, the results of verbal protocols are followed by the results of eye tracking data. Specifically, the relations of error types, diagrammatic complexity, the degree of causal chaining and diagrammatic familiarity to diagrammatic reasoning are analyzed by using SPSS v.13.

3.4.1 Demographic Data

Before the experiments, a questionnaire is given to the participants to gather demographic data (see Appendix A). There are 20 male participants in the study. 10 of them are the domain-familiar participants, who are familiar with the scenario, and 10 of them are notation-familiar participants, who are familiar with the representations.

The notation-familiar group's members have degrees from computer sciences or information sciences. They are also accustomed to the use of diagrams in their fields of expertise. The notation-familiar group's age mean is 25.60 years (*SD*=1.65). The notation-familiar participants' experience mean is 3.50 years (*SD*=1.90). The domain-familiar group's members have degrees from various fields (mostly, military academy graduates) other than computer sciences or information sciences. They are not familiar with the use of software engineering diagrams. The domain-familiar group's age mean is 32.10 years (*SD*=3.45). The domain-familiar group's experience mean is 10.5 years (*SD*=3.50).

Age and experience differences between the groups result from their specific expertise fields. The domain-familiar group's training period takes longer time; thus, the age and experience of the intended KAMA users, which are the domain-familiar participants, are relatively high compared to the notation-familiar group. Furthermore, it is difficult to find and invite elderly and experienced notation-familiar participants to the study. There are significant differences between the groups as introduced in the following sections.

3.4.2 Verbal Protocol Data

From Table 1, it can be seen that one can perform the **pause**, **comment_on_notation**, **comment_on_scenario**, and **nrs** activities at any time. The other actions are bounded to the **read_reason** activity and they can be outlined as in the Figure 10. This figure is intended as a summary for the observed patterns of dependence in the participants' protocols. That is all of them are found to obey this pattern. The pattern is not imposed as a design constraint on the protocols pre-experimentally, and after the analysis of all participants' segments, it is concluded that all of them obey the pattern repeatedly in Figure 10.

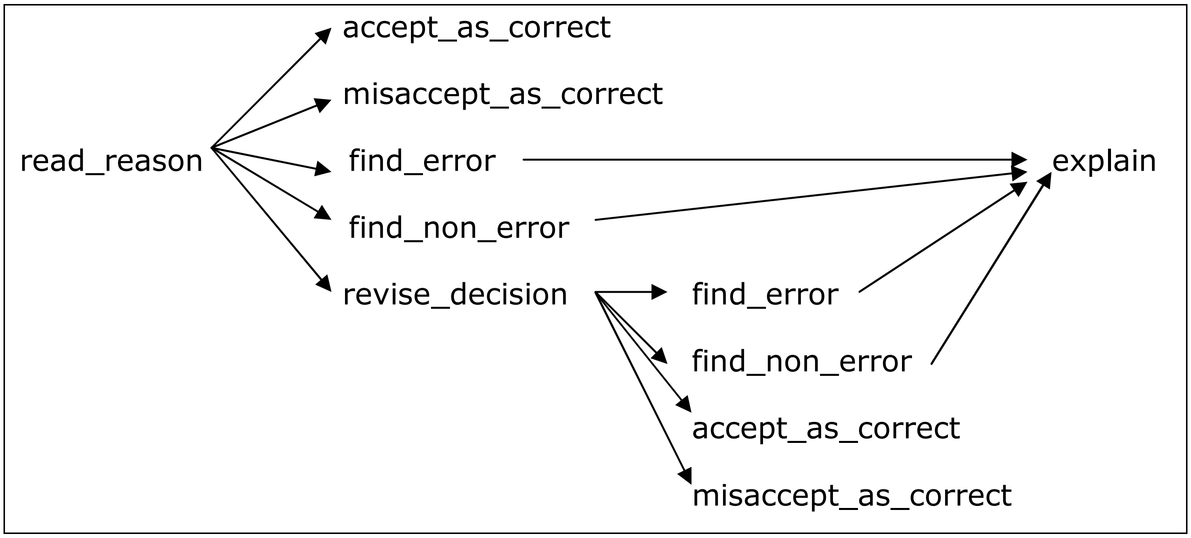


Figure 10 Order of Reasoning Related Segments

The mean duration of total segments, which is also the mean time for the experiment duration, for the notation-familiar group is 30:07 (*SD*=3:40). The domain-familiar group's average time is 26:25 (*SD*=4:31). It shows that notation-familiar group spends more time on notation than the domain-familiar group. Although independent-samples t-test shows no significant difference in the total duration of diagrammatic reasoning experiments of groups ($t(18)=2.004$, $p=.060$), it can be said that if the sample size is increased, the total duration of the experiments between groups may tend to differ significantly. In the following section, verbal protocols of the segments are compared between the groups.

The means and standard deviations of segments in two groups are given in Table 3.

Table 3 Mean Numbers of Segments in Groups

	Domain Familiar Group (N =10)		Notation Familiar Group (N =10)	
	Mean	SD	Mean	SD
Number of read_reason	54.40	2.41	60.30	5.81
Number of explain	9.60	4.58	13.30	3.16
Number of find_error	5.90	2.13	9.20	2.70
Number of accept_as_correct	38.70	2.91	43.70	3.53
Number of find_non_error	3.90	3.14	4.10	2.28
Number of misaccept_as_correct	7.60	2.17	5.60	2.27
Number of comment_on_notation	.40	.70	1.20	1.03
Number of comment_on_scenario	.60	.97	1.10	.88
Number of pause	.80	1.03	.30	.67
Number of revise_decision	1.00	.67	1.80	1.47

In order to find out the segment differences between the groups, the means of the total segments between the groups are statistically tested. Since there is not enough information in the literature about the diagrammatic reasoning differences between the groups, no significances in all of the protocols with respect to the groups are assumed according to the hypothesis stated before as there is no significant difference between the groups in diagrammatic reasoning activity.

ANOVA is used to test the difference between the total numbers of segments between the groups. The mean number of segments for the domain-familiar participants is 123.10 (SD=7.56) and 140.60 (SD=14.95) for the notation-familiar participants as given in the Figure 11. There is a significant difference between the groups ($F(1,18)=10.906$, $p=.004$).

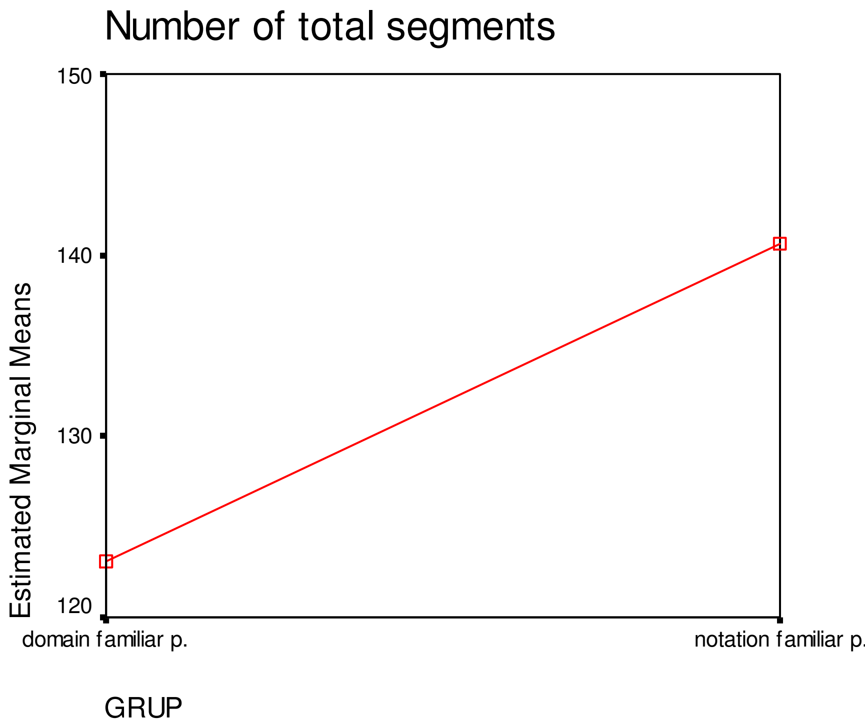


Figure 11 Mean Number of Total Segments

In order to understand in more detail the significant difference between the numbers of segments of each group, a further analysis of segments is performed. The segments **accept_as_correct**, **misaccept_as_correct**, **find_error**, and **find_non_error** are grouped together as **error-related** segments. The rest of the segments are left as they are, which are **read_reason**, **explain**, **pause**, **revise_decision**, **comment_on_notation** and **comment_on_scenario** activities. MANOVA is applied to test the difference in these various segments between the groups. The covariance matrices of MANOVA differ significantly (Wilk's λ=.313, $F(7,12)=3.762$, $p=.022$) which means that the two groups have a significant effect on all of the segments combined. The consequent ANOVA is applied for each segment to see the separate- effects on each segment. The Bonferroni correction is applied to the seven separate follow-up t-tests for each segment type which results in a new p-level of p=0.007 (0.05/7=0.007). The error-related segments (M=56.10, SD=2.60 for domain-familiar group, M=62.60, *SD*=5.68 for notation-familiar group) are significantly different between the groups ($F(1,18)=10.824$, $p=.004$). Moreover, the **read_reason** segment ($F(1,18)=8.788$, $p=.008$) is very close to significant difference. The **explain** ($F(1,18)=4.424$, $p=.050$) and the **comment_on_notation** segments ($F(1,18)=4.114$, $p=.058$) are close to significance, too. The rest of the segments are not significant.

ANOVA is also applied to the number of errors reported by each group. The mean number of reported errors for domain-familiar participants is 5.90 (*SD*=2.13) and 9.30 (*SD*=2.79) for notation-familiar participants. A univariate ANOVA for number of reported errors by each group is also significant ($F(1,18)=9.373$, $p=.007$). When an ANCOVA (Analysis of Covariance) is applied to test the effect of group on number of reported errors with "experience" as a covariate, experience cancels the significant effect of groups F values ($F(1,17)=2.901$, $p=.107$). Thus, although the experience of groups are different (the notation-familiar participants' experience *M*=3.50 years, *SD*=1.90, the domain-familiar group's experience M=10.5 years, *SD*=3.50), group explains the difference.

Despite its insignificance, the **pause** activity is performed 8 times by the domain-familiar group and 3 times by the notation-familiar group. All of the **pause** activities are done at the beginning of the diagram part-2, the stoppage removal, and the diagram part-3, the fire mission training.

Even though it is not significantly different, the **revise_decision** activity is performed 18 times by the notation-familiar participants group and 10 times by the domain-familiar participants. There are 12 erroneous and 43 non-erroneous subparts in the diagrams (see Appendix B). 14 of the **revise_decision** actions are performed in the erroneous areas (most frequently in p3_e3). Erroneous areas are significantly subject to the **revise_decision** activity. Independent-samples t-test is applied to test the hypothesis. The result is significant, too, $t(53)=2.623$, $p<.05$. The non-erroneous areas ($M=.33$, $SD=.68$) on the average are subject to the **revise_decision** activity less than the erroneous areas ($M=1.17$, $SD=1.70$, Cohen's d=.86).

By the observation of the verbal protocols' frequencies, distributions and the comparisons of means, the **error_related** segments are significantly different between the groups. The **read_reason**, **explain**, and **comment_on_notation** activities' trends are for significance. In other words, if the sample size is improved, these activities might probably differ, The **revise_decision** activity is significant not between the groups but between the erroneous and non-erroneous areas. These results will be discussed in the discussion section.

3.4.3 Error Comparisons

There are 14 errors in the diagrams. The errors are categorized as follow: Actor connection (AC) errors, Missing or redundant component (MRC) errors, Interchange (IC) errors, Wrong connection (WC) errors. Since the initial aim was not to test the success rates with respect to error types, the number of errors in each category was not balanced. Table 4 presents the errors, their types, success of the participants and distributions of the errors in the diagrams' parts.

Table 4 Distribution of Errors among the Participants

type:	wc	wc	mrc	ic	ic	ac	ac	mrc	ic	ac	ac	mrc	ac	ac	
no:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	
in:	p1	p1	p1	p2	p2	p2	p3	p3	p3	p3	p3	p3	p4	p4	**sum**
d1	0*	0	0	1*	1	0	0	0	1	0	0	0	0	0	3
d2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	5
d3	1	1	0	1	1	0	0	0	0	0	0	1	0	0	5
d4	1	0	1	1	0	0	0	0	0	0	0	0	0	0	3
d5	1	1	1	1	1	0	0	0	1	0	0	1	0	0	7

Table 4 (cont.)															
d6	0	1	1	1	1	0	0	0	1	0	0	0	0	0	5
d7	0	0	1	1	1	0	1	1	1	1	0	1	1	0	9
d8	0	1	0	1	0	0	1	0	1	1	0	1	1	0	7
d9	0	1	0	1	1	0	0	0	1	1	0	0	1	0	6
d10	1	1	1	1	1	0	0	0	1	1	0	1	1	0	9
n1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	6
n2	0	0	1	1	1	0	1	1	1	0	0	1	1	1	9
n3	0	1	1	1	1	1	1	1	1	0	0	1	1	1	11
n4	0	0	0	0	1	0	1	0	1	1	0	1	1	0	6
n5	0	1	1	1	1	0	1	0	1	0	0	1	1	0	8
n6	1	1	1	1	1	0	1	1	1	1	1	0	1	1	12
n7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
n8	0	1	1	1	1	0	0	0	0	1	0	1	1	0	7
n9	1	1	1	1	1	0	1	1	1	1	0	1	1	1	12
n10	0	1	1	1	1	0	1	0	1	0	0	1	1	0	8

*0 means error not found.

*1 means error found.

In Table 4, the *type*-row denotes the type of the error and *no*-row is for id-number of the errors. The *in*-row shows whether the error is in part-1 (The marksman state diagram), part-2 (The stoppage removal activity diagram), part-3 (The marksman mission training activity diagram) or part-4 (The marksman mission ending and evaluating activity diagram) of the diagrams (see Appendices B). The prefix d represents the domain-familiar participants whereas n represents the notation-familiar participants. If a participant is successful in finding an error, the corresponding cell, which is the intersection of error-id and the participant-id, is set to 1. The *sum*-column shows the total number of errors found by the participants. Error types are defined as follow:

<u>Actor Connection:</u> In the actor connection type, the errors are in the activity representations without actor connection or with wrong actors. For example, in Figure 12, the activity, "Inform about the mission", should have been connected to the trainer actor, not the marksman according to the scenario.

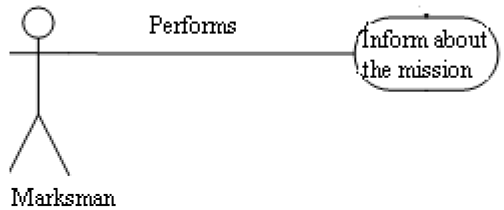


Figure 12 An Example of AC Type Error

Missing or Redundant Component: In this type, the representations with incomplete or redundant parts are categorized together. This type of errors consists of all activity representations with missing or redundant input objects or output objects. Although reasoning most probably differs with respect to whether a component is missing or redundant, their eye search patterns are observed to be almost the same. While solving either a missing component or a redundant component, the participants' eyes usually fix around the activity symbol and search in clock-wise or counter clocks-wise directions. The participants' error-finding attitudes are also similar. For example, in Figure 13, the output "explosion fire" is the redundant component but the rest of the output is given in the scenario.

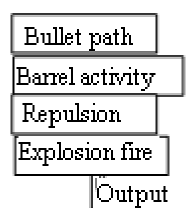


Figure 13 An Example of MRC Type Error

Interchange: In the interchange type, the activity representations are not in same the order that in the scenario. Consecutive activity representations are interchanged. For example, in the scenario it is given as, "In case of defective spring, replace the string. In case of filthy mechanism, clean the rifle." Yet, in Figure 14, it can be seen that the corresponding activity diagrams are interchanged erroneously.

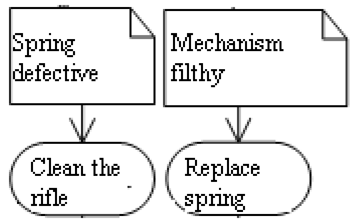


Figure 14 An Example of IC Type Error

<u>Wrong Connection:</u> Wrong connection type errors may occur in state diagrams. In these errors, the states diagrams are connected to the wrong succeeding states. For example, in the scenario it is stated as, "At the beginning, the marksman is in the state of waiting for orders." However, in Figure 15, the starting symbol is erroneously connected to the state of loading the gun.

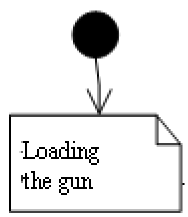


Figure 15 An Example of WC Type Error

Table 5 shows the participants' success rates with respect to error types. The comparison of the success rates of the groups with respect to error types is discusses in the discussion section.

When Table 4 is studied, it can be concluded that error-6, error-11 and error-14, which are all actor connection errors, have the lowest success rates. Furthermore, although error-4, error-5 and error-9 are of same type, the success rate in discovering error-9 is relatively low. These are discussed in the discussion section.

Table 5 Success Rates of Participant with respect to Error Types

type:	WC (Total:2)	MRC(Total:3)	IC (Total:3)	AC (Total:6)
d1	0	0	3	0
d2	2	1	2	0
d3	2	1	2	0
d4	1	1	1	0
d5	2	2	3	0
d6	1	1	3	0
d7	0	3	3	3
d8	1	1	2	3
d9	1	0	3	2
d10	2	2	3	2
n1	2	2	2	0
n2	0	3	3	3
n3	1	3	3	4
n4	0	1	2	3
n5	1	2	3	2
n6	2	2	3	5
n7	2	3	3	6
n8	1	2	2	2
n9	2	3	3	4
n10	1	2	3	2

The hypothesis, "the success rate of finding errors differ with respect to error types between the groups", is tested by t-tests. The success rate of the wrong connection type errors is not significantly different between the groups contradicting with the hypothesis. The independent-samples t-test is not significant, $t(18)=.000$, $p>.05$. This result implies that success rate in the wrong connection type errors does not vary between the groups significantly. Similarly, the success rate of the interchange type errors is not significantly different between the groups according to the independent-samples t-test, $t(18)=.739$, $p>.05$, as in the hypothesis. This result implies that the success rate in the interchange type errors does not vary between the groups significantly.

The success rate of the missing or redundant component type errors is tested. Independent-samples t-test is applied and it is significant as given in the hypothesis, $t(18)=3.051$, $p<.05$. The domain-familiar group ($M=1.20$, $SD=.68$) on the average are less successful in finding the MRC type errors than the notation-familiar group ($M=2.30$, $SD=.67$). The 95% confidence interval for the difference in means is ranging from -1.86 to -.34 in finding the missing or redundant component type errors. The effect size (d) was .36 indicating a medium effect of being domain or notation-familiar on finding the missing or redundant component type errors (see Cohen, 1988).

The other result corresponding to the hypothesis is obtained from the success rate of the actor connection type errors. Independent-samples t-test is applied. The test is significant, $t(18)=2.928$, $p<.05$. The domain-familiar group ($M=1.00$, $SD=1.33$) on the average are less successful in finding the AC type errors than the notation-familiar group ($M=3.00$, $SD=1.70$). The 95% confidence interval for the difference in means is ranging from -3.44 to -.56 in finding the actor connection type errors. The effect size (d) was .31 indicating small effect of being domain or notation-familiar on finding the actor connection type errors (see Cohen, 1988).

These differences will be discussed in the discussion section.

3.4.4 Diagrammatic Complexity and Degree of Causal Chaining

Diagrammatic complexity refers to the time and space required to finish a task; which is an algorithm to solve a problem (Borodin, 1972; Fortnow & Homer, 2003). Parallel flows of actions, loops returning back to themselves, and nested logical statements are the source of diagrammatic complexity because they increase the time required to finalize a for a task. They increase not only the number of steps needed for a task, but also space required for memory. In this study, there is a sample of a diagrammatically complex sub-part, **p4_a5**, in the diagram-**p4**, the marksman mission ending and evaluation activity diagram, which is given in Figure 16.

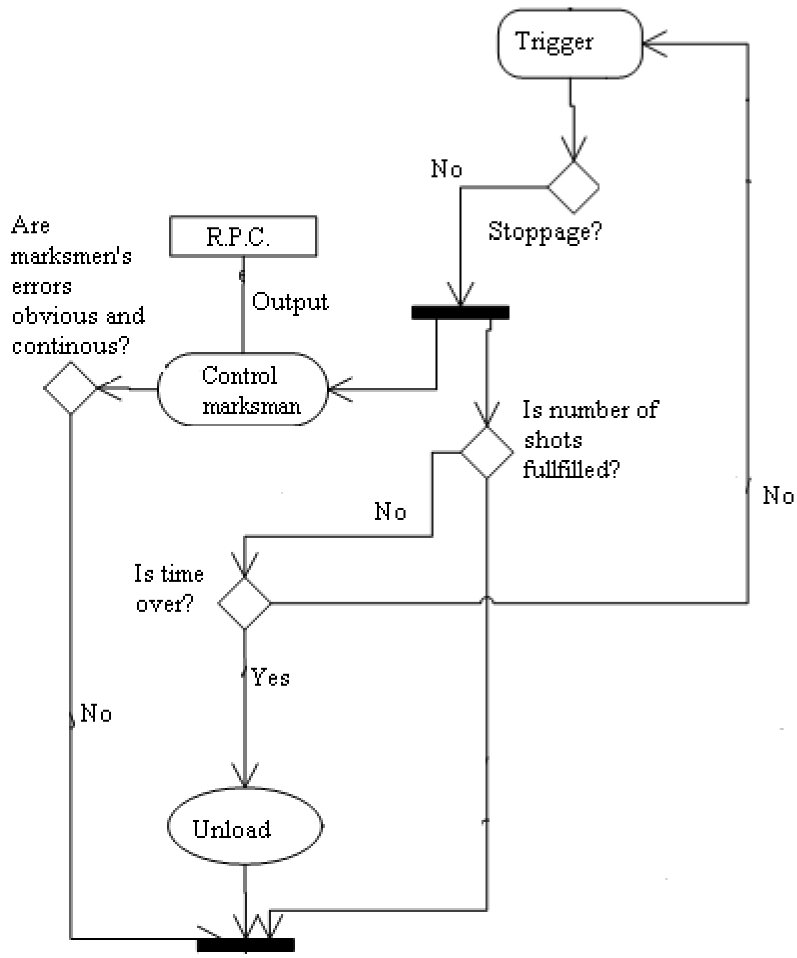


Figure 16 Diagrammatically Complex Sub-part p4_a5

There are parallel flows after the decision symbols of this diagram. Some of the exits of decision symbols return back to a previous decision symbol, and eventually to themselves. This portion is a diagrammatized version of diagrammatic complexity, and it takes a lot of time for all of the participants to reason about it.

Yoon and Narayanan (2004) state that the successes of people working on diagrams depend on the critical components and the causal chain of the diagrams. If diagrammatic components are combined with each other, with a decisional element in a single 'if-then' structure, they are said to form a causal chain. All of the elements of the stoppage removal activity diagram, **p3**, are bound to each other in the form of causal chaining. As opposed to the difficulty of understanding **p4,** which results from its diagrammatic complexity, **p3** is very understandable because of its high degree of causal chaining.

In order to test the difference between the areas with high diagrammatic complexity and the area with a high degree of causal chaining, another portion of the diagram is extracted from **p3** as in Figure 17.

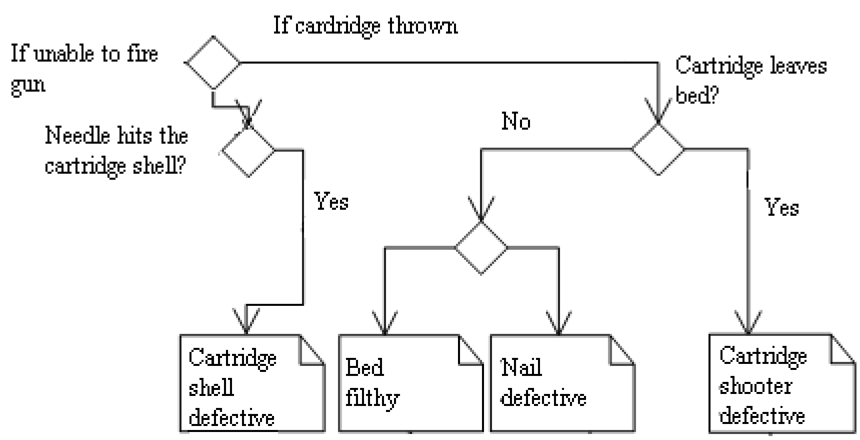


Figure 17 The Sub-part of p3 with a High Degree of Causal Chaining

Since the most important element of computational and causal issues is the decision element, the number of decision elements is kept the same, namely 4, among the diagrams regarding the scenario. One can easily infer that the subpart with high degree of causal chaining has more crowded informational elements than the diagrammatically complex subpart does. The total gaze durations and the number of fixations of these areas are compared.

The diagrammatically complex area on average attracts longer gaze durations (*M*=23.36s, *SD*= 9.17s) than the area with a high degree of causal chaining (*M*=16.71s, *SD*=8.13s). Also, the diagrammatically complex area (*M*=183.40, *SD*=76.83) on average has more eye fixations than the area with the high degree of causal chaining (*M*=52.35, *SD*=24.76).

MANOVA is used to test the effect of area type on the combination of the two dependent variables durations and fixations on these fields. The highly significant result of MANOVA (Wilk's λ=.411, $F(2,37)=25.567$, $p=.000$) requires further separate ANOVAs in order to test for the single effects of the fixations and the durations. Both the gaze durations ($F(1,38)=5.887$, $p=.020$) and the number of fixations ($F(1,38)=52.711$, $p=.000$) of the diagrammatically complex area and the area with a high degree of causal chaining are significantly different.

MANOVA is used to analyze the effect of area type (the diagrammatic complexity vs. the degree of causal chaining) and group on the combination of eye gaze duration and number of fixation. The combined effect is significant (Wilk's λ=.757, $F(2,35)=5.633$, $p=.008$). Yet, ANOVA shows that only the single effect of area type is significantly different in terms of gaze duration. ($F(1,36)=8.550$, $p=.006$) but not in terms of number of fixations. Therefore, the difference in eye movement data of these fields is explained by the diagrammatic type of the areas, not by the domain of expertise (see Figure 18 and Figure 19).

Besides the significant gaze durations and fixations of the diagrammatically complex area, the number of **find_non_error** activity of it is also high. Table 6 shows the distribution of **find_non_error**s among the sub-parts, of which **p4_a5**, the diagrammatically complex one, frequently deceived the participants to be erroneous, though it is not.

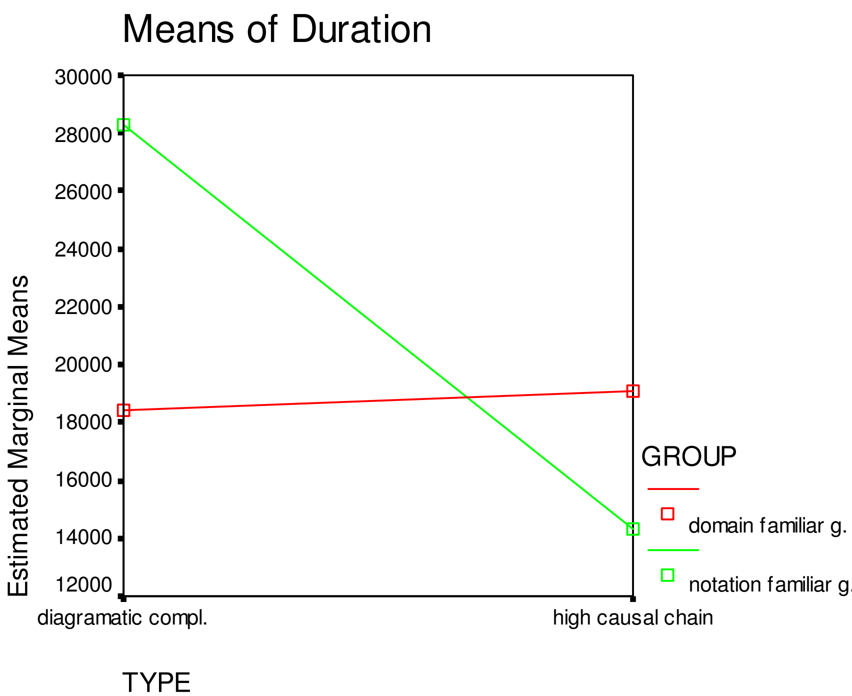


Figure 18 Means of Duration on Sub-parts

,

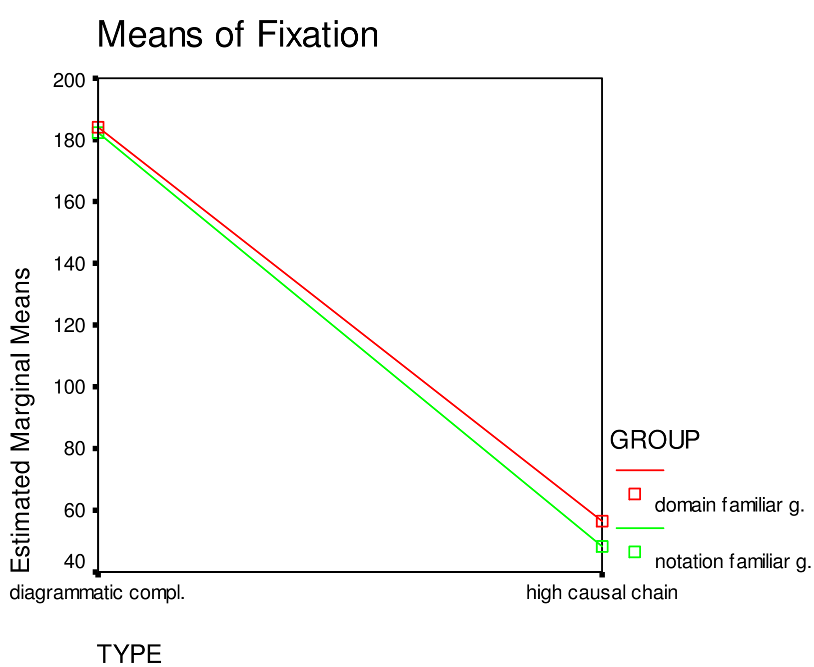


Figure 19 Means of Fixations on Sub-parts

Table 6 The Areas Subject to find_non_error Activity

Area	The domain-familiar participants (N=10)	The notation-familiar participants (N=10)	Sum
p1_a2	2*	0	2
p1_a4	1	4	5
p1_a5	2	0	2
p1_a7	2	4	6
p1_a8	1	1	2
p2_a14	0	1	1
p2_a2	4	0	4
p2_a6	1	0	1
p3_a10	1	1	2
p3_a2	0	2	2
p3_a3	2	1	3
p3_a4	3	1	4
p3_a5	2	1	3
p3_a8	0	5	5
p3_a9	0	1	1
p3_e2	2	2	4
p4_a3	2	1	3
p4_a5	15	11	26
p4_a8	1	0	1
p4_e2	0	1	1
p4_e2	0	2	2
Sum	41	39	80

* The numbers in cells refer to the number of find_non_error activities

This interesting finding is specifically tested in the follow-up experiment, which is discussed later.

3.4.5 Eye Tracking Data

In this study, eye tracking data of the participants are also collected. These are fixation numbers[1] and gaze durations[2]. The domain-familiar participants group's mean number of eye fixations is 2269.30 (*SD*=616.81) and the mean

[1] Number denotes how many times a subject's eye are focused on a region.
[2] Number denotes how many miliseconds a subject's eye look at a region.

of gaze durations is 586.938 seconds (*SD*=104.03s). On the other hand, the notation-familiar participants group's mean number of eye fixations is 2379.00 (*SD*=510.08) and the mean of gaze durations is 734.74 seconds (*SD*=128.07s). To analyze eye tracking data, fixation numbers and gaze durations of participants should be compared between groups.

MANOVA is used to test the effect of group on the combination of the two dependent variables, namely number of fixations (the domain-familiar participants group's *M*=2269.30 fixations, *SD*=616.81, the notation-familiar participants group's *M*=2379.00 fixations, *SD*=510.08), and gaze durations of groups (the domain-familiar participants group's *M*=586.938s, *SD*=104.03s, the notation-familiar participants group's *M*=734.74s, *SD*=128.07s). The test shows a significant difference between the groups (Wilk's λ=.518, $F(2,17)=7.922$, $p=.004$). Consequent ANOVAs are used to explore the separate effects. Only the gaze duration data is significantly different between the groups ($F(1,18)=8.025$, $p=.011$) but the numbers of fixations do not differ between the groups significantly. When "experience" is a covariate in the test, it removes the combined effect of the two variables and also of the separate effect of gaze duration. In other words, experience removes the significance in gaze duration.

When eye tracking data is analyzed with respect to sub-areas, p4_a5, which is a diagrammatically complex area, has the longest gaze duration (*M*=50.13s, *SD*=21.39s) and the biggest number of eye fixations (*M*=180.90, *SD*=78.83). This is going to be discussed in discussion sections.

In order to see the differences in durations and fixations of groups, the sub-areas are also tested. Independent-samples t-test is applied to all areas' gaze durations and fixations. The test shows that the durations of p3_a4, which is an erroneous area of type MRC, ($t(18)=2.349$, $p<.05$) and p4_a4 ($t(18)=-2.809$, $p<.05$) are significantly different between the groups.

Domain familiar group's mean for p3_a4 is 17.31 seconds (*SD*=17.31s) and for p4_a4 is 18.43 seconds (*SD*=2.73). On the other hand, notation-familiar group's mean for p3_a4 is 23.88 seconds (*SD*=2.20s) and for p4_a4 is 28.30 seconds (*SD*=2.21s).

Finally, also the fixation numbers of p4_a4 are also significantly different between the groups. Independent-samples t-test is applied and with $t(18)=2.789$, $p<.05$ the difference is seen. The domain familiar group's mean is 57.20 fixations (*SD*=8.38) and the notation-familiar group's mean is 88.50 (*SD*=7.47). The difference between durations and fixations of notation-familiar group for p3_a4 is clustered around the activity symbol to search the missing input object. Moreover, for p4_a4, durations and fixations around actor representations create the difference between the groups.

The eye tracking data support the results of the verbal protocol data. For example, if there is a longer duration for a **read_reason** activity, then its eye tracking data is consistent with it, too. This correspondence is revisited in the discussion section.

3.4.6 Interview Results

After the experiments, the participants are interviewed (see Appendix A). In this section, the interview results will be reviewed and compared to the errors. The participants report that the state diagram is easier to understand than the activity diagrams. The participants are asked to grade the understandability of each diagram from 1 (difficult) to 5 (easier). The correlation between the grading of each diagram and the number of errors found in the corresponding diagram are different between the groups. A Pearson correlation is applied. For the domain-familiar group, there is a significant negative correlation between the number of errors found in part-4 and the grades of part-4, $r(8)=.802$, $p<.05$. Although it is not significant, there is also a negative correlation between the number of errors found in part-4 and the grades of part-4 for the notation-familiar group. The diagram part-4 includes the diagrammatically complex sub-part. For the notation-familiar group, there is a significant positive correlation between the number of errors found in part-3 and the grades of part-3, $r(8)=.867$, $p<.05$. The domain-familiar group's correlation is also positive for part-3, despite its insignificance. The diagram part-3 has high degree of causal chaining.

For both groups, there are positive correlations between the grades and the number of errors found for diagrams part-1, part-2, and part-3. However, there is a negative correlation between the participants' grades and the number of errors found in diagram part-4. In other words, although participants are successful finding the errors in part-4, they assign lower grades to this part, contrary to part-3. They complain about crowded informational components. However, part-2 and part-3 include more informational elements than part-4 does. Therefore, the number of errors found might not be the major concern for understandability of diagrams. It is diagrammatic complexity and causal chain degree that is main concern about understandability of the diagrams.

Apart from the grading, the participants' experiences are also compared to the number of errors found. The hypothesis is that experience has positive effect on error-finding. Although Pearson correlation is not significant, but tends to significance, there is a difference between the groups: As oppose to the hypothesis, the correlation between the experience and the number of errors found is negative for the domain-familiar group, whereas it is positive for the notation-familiar group as hypothesized. The same proposition holds for age despite its insignificance, too. This is discussed further in the discussion section.

In the questionnaire, the participants' problem-solving strategies are asked, too. They read the scenario by one or a few lines and searched each line on the diagrams. The participants' reported problem-solving strategies are same as the observed strategies. Although diagram parts are related to each other, the participants do not perform parallel search among the diagrams to understand or find errors; i.e., they do not employ the concurrent search techniques among the diagrams; and they follow the representations linearly, instead. The participants also complain about the crowded informational elements, their distributions and the similarity between the main flow line and the lines connecting actors and objects to the activities. The participants state that they get lost when these lines are too long without any visual landmark that differs visually from the surrounding visual elements, which is also supported by the eye tracking data.

3.5 Discussion of the Main Experiment

Since the main experiment includes verbal protocol data, eye movement data, error types and the parts with the diagrammatic complexity and the degree of causal chaining, each of them are discussed in a different sub-section as follows.

3.5.1 Discussion of Verbal Protocol Findings

The pilot study and the main experiment show that concurrent verbal protocols do not change the thought processes, but it gives clues about the processes, and increases the completion time because of verbalization as Ericsson and Simon (1995), Norris (1990) and Randsdell (1995) state.

The results section shows that there are differences in diagrammatic reasoning activities of the notation-familiar and the domain-familiar groups. Error related segments shows significant difference between two groups. First of all, although the scenario is related to the domain-familiar group, the notation-familiar group who is familiar with diagrammatic representations is significantly more successful in error-finding. In other words, as Hegarty (2000) states, the problem solvers that can employ textual and graphical representations together are more successful. This result enhances the argument that if a problem is diagrammatized and even if the topic is relatively non-familiar to a group of participants, diagrammatic representational familiarity improves error-finding and correcting. Similarly, the **accept_as_correct** activities are different, too. Diagrammatic reasoning is not only about finding and solving errors, but also about verifying the information represented by diagrams. The notation-familiar group works more on diagrams, and they accept the scenario and its representation as correct, sentence-by-sentence. On the contrary, the domain-familiar group sometime reads a block of sentences, and tries to verify the statements. Since they are not familiar with the representation, they omit verification through diagrams or combine a set of sentences into a single verification.

As a result of the difference in the error related segments, **read_reason** activities are expected to be different. In the results section, the **read_reason** activities of the notation-familiar group, indeed, are also observed to be more frequent. Since the notation-familiar group is familiar with diagrams, they tend to investigate diagrams more thoroughly, and they try hard to find errors. On the other hand, domain-familiar group refers more frequently to the scenario, which is error free, and tends to omit diagrammatic problem-solving.

Although the **pause** activities are performed 8 times by domain-familiar group and 3 times by notation-familiar group, it is not significant. All of pause activities occur at the beginning of the diagram part-2, the stoppage removal, and the diagram part-3, the fire training mission. The details and cognitive processes of these **pause** activities can not be explained. However, the diagram part-2 is the first activity diagram that the participant sees, which is in a parallel flow. The diagram part-3 is the first activity diagram in a linear flow manner. Therefore, it can be proposed that whenever a participant comes across a different type of diagram part, he goes through an unspecific **pause** stage. This can be a period in which the participants try to get familiar with the new type of representation or receive the initial sketch of the new representation to help succeed in the reasoning activity.

In the results section, it can be further seen that the **read_reason**, **explain** and the **comment_on_notation** activities' are close to significance in two groups. In other words, if the sample size is increased, these activities might probably differ, too. Since the domain-familiar participants are less successful in finding errors, their **explain** activities are less frequent and they spend less time one explaining. The notation-familiar group comments on the representations and provide suggestions during the experiments more frequently. Since the domain-familiar participants are not familiar with the diagrammatic representations, they keep silent or prefer commenting on scenario. In brief, verbal protocol analysis shows us that diagrammatic reasoning, finding errors and verifying diagrams are different in the two groups. Although the notation-familiar group is non-familiar with the scenario, they are more successful.

The difference in the number of reported errors is removed when experience is a covariate. The reason is that the correlation between the success rates and experience is positive for the notation-familiar group and negative for the domain-familiar group. This means that the significant difference in success rates is a result of the expertise domain, not experience.

3.5.2 Discussion of Error Findings

The missing or redundant component and the actor type error-findings show significant differences between the groups. The missing or redundant component type errors are missing or redundant representations of input or output objects of an activity diagram. The notation-familiar group is successful translating a sentential explanation into a diagrammatic explanation. Therefore, they are more successful in finding such errors. Similarly, the domain-familiar group tends to omit checking actor connections. They are unsuccessful in actor connection errors, unlike notation-familiar participants.

The interchange and the wrong connection type errors are related to the flow of the scenario or activities; i.e., verbs, while the missing or redundant component and the actor type errors are related to participants and objects of the sentences. For example, one can translate a sentence, "…George writes a letter by a pen…" into a diagrammatic representation as in Figure 20.

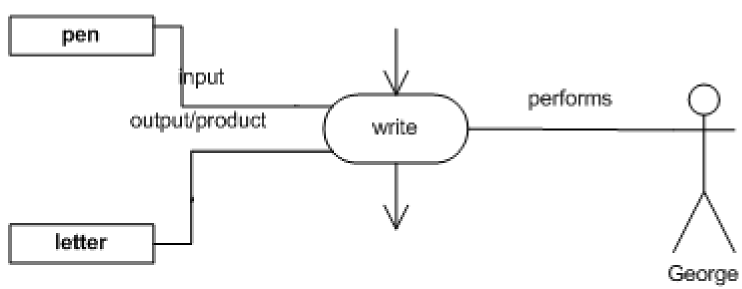


Figure 20 Sample Representation of a Sentence

It can be claimed from the results of the analyses that if there was an error in the activity symbol, *write,* or in its flow connections, both groups would most probably be successful in finding the error. On the other hand, if there were an error in the input/output objects or in the actor connection, the notation-familiar group would be more successful. In brief, if a sentential explanation from a domain is erroneously translated into a diagrammatic representation, the domain-familiar participants' successes in detecting errors will be low if the error is in the subject or object representations with respect to the notation-familiar participants. For a textual sentence to be understand, firstly the verb is in the focus of attention as a functional nucleus unit. Then, the attention mechanism selects the thematic relations and moves to the arguments of the verb, which are the subject and object. Similar condition is valid for understanding of the diagram representing a sentence and it is supported by the finding of this study. Since the first attention focus is toward the verb and the thematic roles are in the center of focus not the syntax, the IC and WC type errors are not significantly different between the groups. However, the notation-familiar group is more successful in the error types AC and MRC which are related to the arguments of the action.

When Table 4 analyzed, it can be seen that error-6, error-11 and error-14, which are all actor connection errors, have the lowest success rates. In error-6, the actor connection line is missing and very few participants notice this error. It also has the shortest gaze duration and fewest fixations. Parallel to the activity symbol of error-6, there are non-erroneous activities. Actor connection lines and main flow lines are very similar to each other. This might be the reason for neglecting the connection in error-6. In error-11 and error-14, the actor connection lines and main flow lines are also similar. Therefore, if the component connection lines are identical to the main flow lines in terms of its visual and spatial properties, the participants tend to neglect the connections. Such studies should be repeated with more participants and more errors in a more detailed way to strengthen these results.

Furthermore, although error-4, error-5 and error-9 are of same type, the success rate in error-9 is relatively low. These errors are of the interchange component errors. In diagram part-2, where error-4 and error-5 reside, the main flow is in a parallel manner and with a high causal chain. On the other hand, in diagram part-3, the main flow is linear. Therefore, it can be concluded that if the interchange type of error occurs in a linearly flowing representation, the success rate in finding the error will be low compared to the representation flowing in a parallel way.

When the results of the correlation between the diagrams' grades and the number of errors found in the corresponding diagrams are analyzed, it can be proposed that a diagram's understandability is not mainly related to the number of errors found in the diagram. Although the participants find high number of errors in the diagram part-4, they assign the lowest grade, which is 1, to it. They complain about crowded representations. However, the diagram part-3 is almost twice as crowded as the diagram part-4, and it receives the highest grades from the participants. The reason for the diagram part-4's grades to be low is that its diagrammatic complexity is high. The diagrammatically complex sub-part of the diagram part-4 increases the load in the central executive of the working memory of the participants during diagrammatic reasoning. On the other hand, although the diagram part-3 is more crowded and the number of errors found varies among the participants, its grade is high. The reason is that the diagram part-3's degree of causal chaining is high. It is easy to process the components in the diagram part-3. When eye tracking data of both diagrams are investigated, the diagrams part-3's total gaze duration's average is 160.18 seconds (*SD*=6.45s), which is the highest duration. In this case, high gaze duration is not related to difficulty of the diagram but high number of informational element because as the number of informational elements increases, the participant needs to look at the diagrams more to refresh the mental images in the visuospatial sketchpad.

Finally, if the experience and the number of errors found are compared between groups, it is observed that the correlation between experience and the number of errors found is positive in the notation-familiar participants and negative in the domain-familiar participants. It can be concluded that for the notation-familiar group, if the experience increases, their familiarity with the diagrammatic representations and diagrammatic problem-solving increases, too. Therefore, experience has a positive effect on the notation-familiar group, as also reported in the study of Hungerford et al. (2004). On the other hand, as the experience of the domain-familiar participants group increase, they become more familiar with the sentential representation. It can be tentatively proposed that if a domain-familiar person gains experience in his domain without the use of diagrams, his diagrammatic reasoning ability decreases. The results of the analyses also show the success rate difference between the groups is not because of experience but because of expertise fields.

3.5.3 Discussion of Diagrammatic Complexity and Degree of Causal Chaining

Similar to the diagrammatic complexity property of algorithms, the diagrammatic complexity of diagrams increases the working memory load, too. This creates difficulty in reasoning and problem-solving as Larkin and Simon (1987) conclude from their study. High numbers of **find_non_error** protocols, eye fixations and gaze durations are clustered on the diagrammatically complex part, **p4_a5**. This supports the idea that diagrammatic complexity increases the difficulty in understanding the diagrams independent of the expertise fields. When one carefully investigates the sub-part **p4_a5**, it can be seen that although informational elements are not very crowded in this field, many eye movements are observed around critical components like decision elements, logical connections and inclusive input/output representations. As in Noon and Narayanan's (2004) statements, the degree of causal chaining improves the understanding of diagrams and fixating the critical components is required in problem-solving (Cheng, 1996; Grant & Spivey, 2003; Kalyuga, Chandler & Sweller, 1997).

In the diagrammatically complex part, the participants many times referred to the scenario and their impasses took longer times and sometimes these impasses resulted in **find_non_error** activities. In the reasoning process of this section, the participants tended to get silent and looked at the components. On the other hand, the eye tracking data of the areas with degree of causal chaining is usually for retrieving visual information and reasoning processes were decomposed. The participants were able to decompose the representations and work incrementally. However, the diagrammatically complex area is difficult to decompose. Although p2 is much more crowded than p4 in terms of informational elements, p4 is difficult to reason with. Very few **find_non_error** activities are observed in this part. The interview results also agree on that p4 is the most difficult and deceptive one.

3.5.4 Discussion of Eye Tracking Data Findings

When the total gaze durations are studied, similar to verbal protocols' findings, it can be seen that the domain-familiar participants refer to the diagrams less than notation-familiar ones. They usually look at the diagrams less and focus on the sentential explanations more. There is no difference in number fixations of the groups. Yet, the durations are different. The domain-familiar group's fixations are like a search pattern which is frequent with shorter gaze durations. On the other hand, the notation-familiar group's fixations are frequent with longer durations. This explains the non significance in number of fixations and the significance in gaze duration.

The difference in eye tracking data supports the difference in **read_reason** activities of the groups. Eye movement data show that Chandrasekaran et al.'s (1995) term "visual information" is inferred from the inspection of an external image. Eye fixations are required to reexamine labels to refresh the content of the working memory with respect to the related diagrammatic components of the problem, as stated by Findlay (2004) and Carpenter and Shah (1998).

When the eye tracking patterns are studied, it can be seen that frequent and short fixations are employed in the searching of diagrammatic components, while rare and long fixations indicate the impasses as stated by Knoblich, Ohlsson and Raney (2001). The notation-familiar group fixates on the sub-parts with longer durations. On the other hand, the domain-familiar group fixates with shorter duration. They usually perform search behavior, as stated by Bonatti, Frot, Zangl and Mehler (2002). Whether it is shorter or longer, the fixations are required to refresh and support the mental imagery operations in the visuospatial sketchpad as stated by Logie and Della Sala (2005). Most of the **read_reason** activities' eye-tracking patterns follow the same steps of fixations. Moreover, the participants' eyes usually start with fixating the activities, then its connections and finally the objects and actors. This can be arguably interpreted as the participants first attend to the activities and flow and then pay less attention to actors and input/output objects. This claim is also supported by the findings of error type comparisons. For example, p3_a4 and p4_a4 show significantly different gaze durations and p4_a4 also shows different fixations between the groups. p3_a4 consists of a missing or redundant component type error. The notation-familiar group tends to search for the missing component and to look at the component and actor connections with longer duration. p4_a4 has its actor away from the activity, which is not fixated by domain-familiar participants. These significances also support the error type differences between the groups.

The Clearview® software puts the eye tracking data of the participants in groups and produces overall pictures of the general gaze duration patterns. These pictures are called hotspots. The general gaze duration hot spot pictures of groups are give in Appendix D, in which red color refers to the areas with the longest duration, green for the areas with medium length duration, and yellow for the shortest duration.

CHAPTER 4

FOLLOW-UP EXPERIMENT

4.1 Setting of the Follow-up Experiment

In this chapter, first the hypotheses of the follow-up experiment are given. Then, the setting, material and the conducting of the experiment are introduced.

4.1.1 Hypotheses of the Follow-up Experiment

After the main experiment, certain diagrammatic properties are observed to affect error-finding behaviors. These diagrammatic properties are the diagrammatic complexity and the degree of causal chaining. In order to study the question, how these properties affect error-finding behavior, the follow-up experiment was performed. Before the follow-up experiment, the hypothesis is defined as the diagrammatic complexity affects diagrammatic reasoning negatively, and the high degree of causal chaining affects diagrammatic reasoning positively. This hypothesis is tested with the help of eye tracking data in the follow-up experiment. In the follow-up experiment, we also wanted to explore whether gender affects these properties and the success rates of error-finding or not. Then, other hypothesis of the follow-up experiment is: The success rates of error-finding do not differ significantly with respect to the gender of the participants. The gender factor of the follow-up experiment is balanced and tested with respect to this hypothesis. The research questions and the corresponding hypotheses of the main experiment are summarized as follow:

Q1: How do the diagrammatic complexity and the degree of causal chaining affect diagrammatic reasoning?

> **H1:** The diagrammatic complexity affects diagrammatic reasoning negatively, and the high degree of causal chaining affects diagrammatic reasoning positively

Q2: Does gender have effect on these properties?

> **H2:** Gender does not affect these properties.

The setting and materials of the follow-up experiment are also given in the following section.

4.1.2 Setting up and Conducting the Follow-up Experiment

The aim of the follow-up experiment is to test how the diagrammatic complexity and the degree of causal chaining properties of diagrams affect eye movement and problem-solving in diagrams. Therefore, 24 university students are tested in the follow-up experiment. The scenario of the follow-up experiment is about the motor vehicle tax law and the debt collection and bankruptcy. In Turkey, the motor vehicles' owners have to pay annual taxes with respect to some properties of the vehicles, and the steps of the motor vehicle tax payment action are represented in a diagram in the follow-up experiment. The debt collection and bankruptcy scenario explains the legal actions to be taken in case of debt collection and bankruptcy. The diagrams of the follow-up experiment are prepared in the same way by using Microsoft

Visio® as it is done in the main experiment (see Appendix C). In the follow-up experiment, only the eye tracking methodology is collected. The results of main experiment and follow-up experiment are reported in the analysis and discussion section.

4.2 Results of the Follow-up Experiment

24 participants in the follow-up experiment are tested to study some of the findings of the main experiment. Half of the participants are male while the other half is female students from various departments of METU and they have no formal training in software engineering notations. The mean age of the participants is 20.12 years (*SD*=1.18).

There are 4 errors in the first diagram, the motor vehicle tax law, and 5 errors in the second one, the debt collection and bankruptcy task. The mean of the number of errors found is 4.04 (*SD*=1.30). The hypothesis is that there is no significant difference in number of errors found with respect to gender. MANOVA shows no significant difference in number of errors found with respect to gender.

In the follow-up experiment, only eye tracking data are collected. In order to test the hypothesis on the negative effect of the diagrammatic complexity and the positive effect of degree of causal chaining on diagrammatic reasoning in term of eye movement data, 2 sub-parts are extracted. Figure 21 shows the diagrammatically complex sub-part of the follow-up experiment, while Figure 22 corresponds to the high degree of causal chaining. The number of decision elements and informational elements are kept the same regarding the scenario.

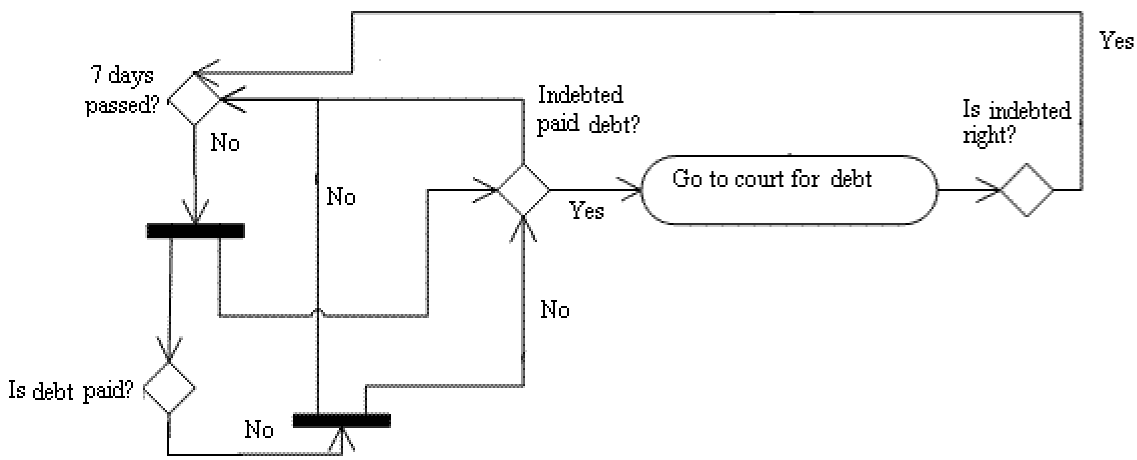


Figure 21 The Diagrammatically Complex Sub-part of the Follow-up Experiment

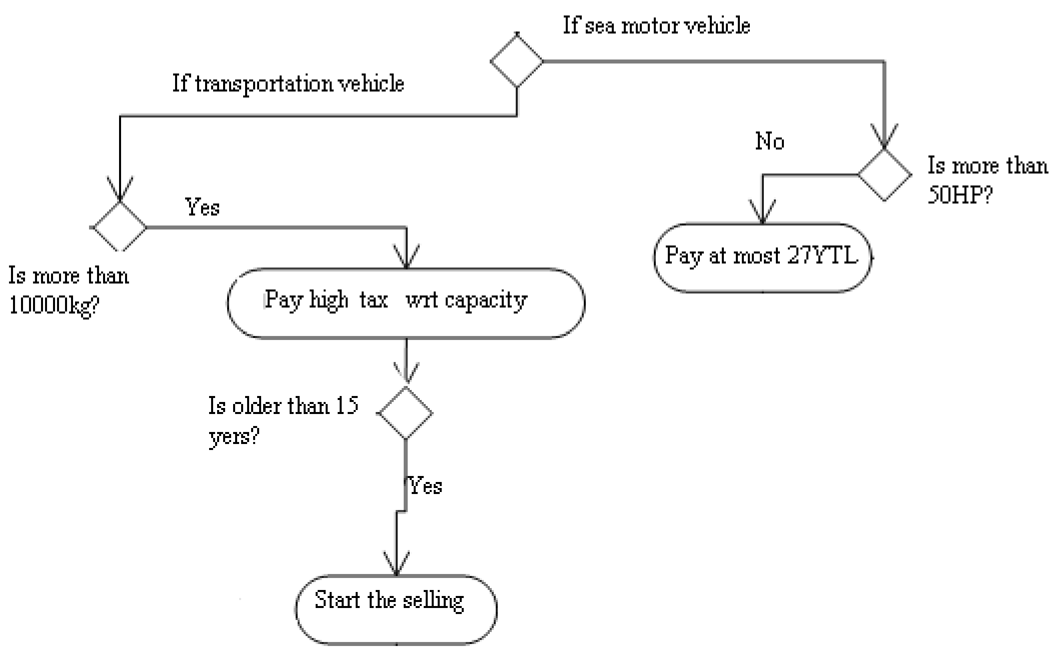


Figure 22 The Sub-part with Causal Chaining of the Follow-up Experiment

The diagrammatically complex area (*M*=72.87s, *SD*=21.84s) on the average has significantly more gaze durations than the area with causal chaining (*M*=53.28s, *SD*=23.21s). Also, the diagrammatically complex area (*M*=174.88, *SD*=47.57) on the average has significantly more fixations than the area with high degree of causal chaining (*M*=130.83, *SD*=59.75).

The MANOVA shows a significant combined effect of fixations and durations (Wilk's λ=.835, $F(2,45)=4.450$, $p=.017$) with respect to area types. Subsequent ANOVAs are performed to analyze the separate effects. Both gaze durations ($F(1,46)=9.066$, $p=.004$) and number of fixations ($F(1,46)=7.981$, $p=.007$) of the diagrammatically complex area and the area with a high degree of causal chaining (in the follow-up experiment) are significantly different.

When the eye movement data of these areas are tested with respect to gender of the participants, no significant difference is observed. In other words, when the gender is a factor in the MANOVA, there are no significant single and combined effects on the eye movement data.

In the follow-up experiment, the participants also noted down some non-errors. In total 12 non erroneous areas are reported and 8 of them are from the diagrammatically complex sub-part, Figure 21.

This results support the main experiment in a way that the diagrammatic complexity affects reasoning negatively while the degree of causal chaining does positively independent of gender.

4.3 Discussion of the Follow-up Experiment

The follow-up experiment supported the findings of the general experiments. First of all, it shows that the diagrammatic complexity is a factor that makes software engineering diagrams difficult to reason with as it is stated by Larkin and Simon (1987). It increases the gaze durations and fixations, too. If the same number of elements is connected to each other with the same degree of causal chaining, it makes the diagrams easy to work with as in Yoon and Narayanan (2004) and Hahn and Kim (1999) studies. Since the diagrammatic complexity increases the load in the working memory, such fields are subject of reporting errors, although they are not.

The follow-up experiment also shows that reasoning with diagrammatic complexity and degree of causal chaining is not affected by the gender.

4.3 General Discussion

In this part I will summarize the discussion section of the three main topics, the differences in the verbal protocols, the error types, and the diagrammatic complexity and degree of causal chaining need consideration.

First of all, the notation-familiar group members tend to employ diagrammatic reasoning more than the domain-familiar group. As a result of this, the amount of eye tracking data of the notation-familiar participants is higher as well as some of the protocols. Since they work on diagrams more, reasoning, error-finding and validation activities they show are significantly more than the domain-familiar ones. They also tend to comment on notation more, as expected. From the positive correlation between experience and error-finding, it can be claimed that the diagrammatic reasoning ability improves as it is used and becomes more familiar. When the problem-solving strategies of the participants are observed, it can be seen that diagrams are used in the form of incremental chunks in problem-solving, as Qin and Simon (1990) state. The participants read the text sentence by sentence and search the corresponding chunk. Eye fixations for the purpose of searching are more frequent and shorter. When a chunk is found, less frequent and longer fixations, namely impasses, are observed to solve the problem as in the study of Knoblich, Ohlsson and Raney (2001). The participants move to the next chunk and incrementally make use of the previous ones to reduce search and solution time.

When the error types are considered, it can be concluded that the participants first focus on the activities, then, the flow, and finally the actors and the input/output actors. If a sentential explanation is diagrammatized, the attention starts from the component that represents the verb, then the attention moves to the flow, which is related to order of the action, and finally to the subject and objects of the sentence. Since the notation-familiar group is more experienced in working with such representations, although they focus last on the actor and object connections, they are still successful in reasoning with them. On the other hand, the domain-familiar members' attention decreases when they eventually look at the actor and object components and because of their non-familiarity in notational issues, they tend to miss such errors. The eye movement data also supports this attention decrease when it comes to arguments of the action in the domain-familiar group.

As the final discussion, the computational properties of the diagrams play a more important role in using them in mental activities than visual properties as reported by Larkin and Simon (1987). Despite the twice more crowded number of visual elements of p3, it is reported to be easier to understand than p4. The negative correlation between the participants' grading and the number of errors found supports this result, too. Eye movement data and the significantly higher number of finding non errors also show that the diagrammatically complex areas are difficult to reason about. The number of errors found in diagrammatic reasoning and eye movement data of the diagrammatically complex areas versus the areas with high degree of causal chaining are not affected by gender as can be concluded from the follow-up experiment.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In this study, the diagrammatic reasoning of software engineering notations is studied in a domain-familiar and a notation-familiar group. It is concluded that some functional and mathematical properties, like the diagrammatic complexity and the degree of causal chaining of diagrams, can affect the success in reasoning. The diagrammatic complexity increases the difficulty in the understanding of diagrams independent of gender and expertise, and it might result in finding of non existent errors. Besides, this study supports further investigation of how spatial properties of visual input affect the working of the visuospatial sketchpad in the working memory.

It can also be concluded from the difference in the reasoning processes in the two groups in the main experiment that the attention mechanism for understanding a representation of a sentential explanation starts from the verb-activity and the flow-connection representations first. It then shifts to the object-input/output and the actor-subject representations. If the participant working on diagrammatic representations is more familiar with a domain, he is found to be more successful. Moreover, for the notation-familiar group, a positive correlation between experience and the number of errors is found. Thus, it can be concluded that the diagrammatic reasoning ability improves if the people practice more the use of diagrams.

When the methodology is reviewed, it can be concluded that combining eye tracking data with verbal protocol data not only enhances the analysis of the data, but also helps in making inferences and categorization. A debated topic, the verbal protocol analysis, is more justifiable if supported by eye tracking data. A similarly challenging issue, the management of an extensive amount of eye movement data, is simplified and categorized by the help of protocols. Studies concerning such cross-data modalities should be encouraged. In this respect, the data analysis tools should be improved in a way that they can combine, synchronize and infer with respect to different data modalities. The human mind works across modalities, like auditory, linguistic and visual data. Thus, this proposal of integrated modality analysis will be helpful in cognitive studies.

As a contribution to the simulation conceptual modeling notation and tool development project (KAMA), we prepared a feedback report summarizing the comments of all the participants to the military scenario. For example, the intersecting connection lines and the similarity of the main flow line and the object connection lines in the diagrams are reported to complicate the understanding of the representations. Moreover, random and imbalanced distribution of the diagrammatic components, large area of white spaces, and the distant components and corresponding textual information are stated to affect diagrammatic reasoning of the representations negatively. Such contributions could be extended as more methodological incorporations if this kind of work is carried out and improved on software engineering notations.

Finally, this thesis has significance in the interaction of software engineering and cognitive science. On the cognitive aspects of software engineering diagrams, there are very few studies (Hungerford et al., 2004). The current studies are usually dependent on performance data, like time and success rate. Yet, this thesis integrates eye movement data, verbal protocol analysis and performance data into the cognitive inspection of software engineering notations, specifically for simulation conceptual modeling notations. Such detailed studies on the cognitive analysis of software representations will definitely improve the quality and success of software systems.

5.2 Future Work

If the number of participants is increased and the gender and experience periods are balanced, some limitations will be overcome, and different results might be concluded.

The concurrent eye movement data of the users of the KAMA tool will be a useful study to assess on-going processes in terms of both cognitive sciences and human-computer interaction.

A final future study can be made to test the properties of diagrammatic reasoning in different tasks other than error-finding. Whether the diagrammatic complexity and the degree of causal chaining properties of diagrammatic reasoning are effective in tasks other then error-finding, or how these properties get affected perceptually, e.g., by using colors, can be studied in further studies. UML representations of complex sentences with many relative clauses can also be studied further.

REFERENCES

Anderson, *M*. & McCartney, R. (1995). Developing a Heuristic via Diagrammatic Reasoning. *Proceedings of the Tenth ACM Symposium on Applied Computing* (pp. 227-231). Nashville: Tennessee.

Baddeley, A.D. & Hitch, G.J. (1974). Working Memory. In G.Bower (Ed.), *The Psychology of Learning and Motivation, 8,* (pp. 47-90). New York: Academic Press.

Baddeley, A. D., Gathercole, S. E. & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. *Psychology Review 105*, 158–173.

Baddeley, A.D., Emslie, E. & Nimmo-Smith, I. (1994). *The Doors and People Test.* Bury St Edmunds, Suffolk, UK: Thames Valley Test Company.

Baddeley, A. D. (2000). The Episodic Buffer: A New Component for Working Memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. D. (2003) Working Memory: Looking Back and Looking Forward. *Nature Reviews Neuroscience*, 4 (10): 829-839.

Bainbridge, L. & Sanderson, *P*. (2005). Verbal Protocol Analysis. In J.R. Wilson & N. Corlett (Eds.), *Evaluation of Human Work, Third Edition.* London: Taylor & Francis.

Barkowsky, T., Freska, C., Hegarty, *M*. & Lowe, R. (2005). Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance. In *Proceedings of the 2005 AAAI Symposium* (pp. 7-10). Manlo Park, California: AAAI Press.

Blackwell, A.F., Britton, C., Cox, A., Green, T.R.G., Gurr, C., Kadoda, G. et al. (2001). Cognitive Dimensions of Notations: Design Tools for Cognitive Technology. In *M*. Beynon, C.L. Nehaniv, & K. Dautenhahn (Eds.), *CT 2001 Lecture Notes on Artificial Intelligence 2117*, (pp. 325-341). Berlin Heidelberg: Springer-Verlag.

Benbunan-Ficsh, R. (2001). Using Protocol Analysis to Evaluate the Usability of a Commercial Web Site. *Information & Management, 39,* 151-163.

Biehal, G. & Chakravarti, D. (1989). The Effect of Concurrent Verbalization on Choice Processing. *Journal of Marketing Research, 26,* 84-96.

Biggs, S.F., Rosman, A.J. & Sergenian, G.K. (1993). Methodological Issues in Judgment and Decision-Making Research: Concurrent Verbal Protocol Validity and Simultaneous Traces of Process Data. *Journal of Behavioral Decision Making, 6,* 187-206.

Bonatti, L., Frot, E., Zangl, R. & Mehler, J. (2002). The Human First Hypothesis: Identification of Conspecifics and Individuation of Objects in the Young Infant. *Cognitive Psychology, 44,* 388-426.

Borodin, A. (1972). Computational Complexity and the Existence of Complexity Gaps. *Journal of the ACM, 19 (1),* 158-174

Bowers, V.A. & Snyder, H.L. (1990). Concurrent versus Retrospective Verbal Protocol for Comparing Window Usability. In *Proceedings of the Human Factors Society-34th Annual Meeting, 2* (pp. 1270-1274). Santa Monica, CA: The Human Factors Society.

Carpenter, *P*. A. & Shah, *P*. (1998). A Model of the Perceptual and Conceptual Processes in Graph Comprehension. *Journal of Experimental Psychology: Applied, 4,* 75–100.

Chandrasekaran, B., Glasgow, J. & Narayanan N. H. (1995). Introduction. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 15-27). Massachusetts: AAAI/MIT Press.

Chandrasekaran, B. & Narayanan, H. (1990). Towards a Theory of Commonsense Visual Reasoning. *Proceedings of the 10th Conference on Foundations of Software Technology and Theoretical Computer Science, Lecture Notes in Computer Science 472* (pp. 388-409). Springer-Verlag.

Chen, C. (1999). A Protocol Analysis Model for Investigating Computer Supported Problem-Solving Activities. *Information Technology, Learning, and Performance Journal, 17(2),* 35-44.

Cheng, *P*. C-H. (1996). Functional Roles for the Cognitive Analysis of Diagrams in Problem Solving. *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 207-212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Civelek, *M*. (2006). Modeling a Sample Mission Space of T.A.F. by Using KAMA-C4isrmos, Graduation Project, Department of Information Systems, METU, 2006.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Defense Modeling and Simulation Office (DMSO). (2000). Conceptual Model Development and Validation. *RPG Special Topic,* 11/30/00.

Deffner, G. (1989). Interaction of Think Aloud, Solution Strategies and Task Characteristics? An Experimental Test of the Ericsson and Simon Model. *Sprache und Kognition, 9,* 98-111.

Della Sala, S., Gray, C., Baddeley, A.D. & Wilson, L. (1997). *The Visual Patterns Test: A New Test of Short-term Visual Recall.* Bury St Edmunds, Suffolk, UK: Thames Valley Test Company.

Duchowski, A. T. (2002). A Breadth-First Survey of Eye Tracking Applications. *Behavior Research Methods, Instruments, and Computers, 34(4),* 455-470.

Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory and Practice.* London: Springer-Verlag.

Eberhard, K.*M*., Spivey-Knowlton, *M*.J., Sedivy, J.C. & Tanenhaus, *M*.K. (1995). Eye Movements as a Window into Real-time Spoken Language Comprehension in Natural Context. *Journal of Psycholinguist Research, 24(6),* 409-436.

Eichelberger, H. (2003). Nice Class Diagrams Admit Good Design? *Proceedings of the 2003 ACM Symposium on Software Visualization*, 159-ff.

Ellis, S., Candrea, R., Misner, J., Craig, C. S., Lankford, C. *P*., & Hutchinson, T. E. (1998). Using Eye Tracking Data to Help Build Better Web Pages. Poster presented at the 42nd Annual Meeting of the Human Factors and Ergonomics Society.

Epelboim, J. & Suppes, *P*. (2001). A Model of Eye Movements and Visual Working Memory during Problem Solving in Geometry. *Vision Research, 41,* 1561-1574.

Ericsson, K.A. & Simon, H.A. (1993*). Protocol Analysis: Verbal Reports as Data, Revised Edition.* Cambridge, MA: Bradford books/MIT Press.

Faretti, T.R., McRea, K. & Hatharell, A. (2001). Integrating Verbs, Situation Schemas, and Thematic Role Concepts. *Journal of Memory and Language, 44,* 516-547.

Findlay, J. *M*. (2004). Eye Scanning and Visual Search. In J.*M*. Henderson, F. Farreira, (Eds.), *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp.59-104). NY: Psychology Press.

Fortnow, L. & Homer, S. (2003). A Short History of Computational Complexity. *Bulletin of EATCS, 80,* 95-133.

Gallant, R. (2002). Diagrammatic Reasoning with Statecharts: A Cognitive Approach for Software/System Engineers. *The 22nd Convention of Electrical and Electronics Engineers in Israel,* 98- 100.

Gero, J.S. & Tang, H.H. (1999). Concurrent and Retrospective Protocols and Computer-Aided Architectural Design. In J. Gu & Z. Wei (Eds), *CAADRIA'99.* (pp. 403-410). Shanghai: Shanghai Scientific and Technological Literature Publishing House.

Glasgow, J. & Papadias, D. (1992). Computational Imagery. *Cognitive Science, 16 (3),* 355-394.

Grant, E.R. & Spivey, *M*.J. (2003). Eye Movements and Problem Solving: Guiding Attention Guides Thought. *Psychological Science, 14(5)*, 462-466.

Green, T.R.G., Petre, *M*. & Bellamy, R.K.E. (1991). Comprehensibility of Visual and Textual Programs: A Test of Superlativism against the "Match-Mismatch" Conjecture. In J. Koenemann-Belleveau, T.G. Moher & S.*P*. Robertson (Eds.), *Empirical Studies of Programmers: 4th Workshop* (pp. 121-145). Norwood, NJ: Ablex Publishing.

Guo, *M*. (1997). Automatic Transformation from Data Flow Diagram to Structure Chart*. Software Engineering Notes,* 22 (4), 44-49.

Gurr, C. & Tourlas, K. (2000). Towards the Principled Design of Software Engineering Diagrams. *Proceedings of International Conference on Software Engineering,* 509-518.

Hahn, J. & Kim, J. (1999). Why Are Some Diagrams Easier to Work With? Effects of Diagrammatic Representation on the Cognitive Integration Process of Systems Analysis and Design. *ACM Transactions on Computer-Human Interaction, 6 (3)*, 181–213.

Hegarty, *M*. (1992). Mental animation: Inferring motion from static diagrams of mechanical systems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18 (5),* 1084-1102

Hegarty, *M*. (2000). Capacity Limits in Diagrammatic Reasoning. In *M*. Anderson, *P*. Cheng & V. Haarslev (Eds.), *Theory and Application of Diagrams* (pp. 194-206). Berlin: Springer.

Hegarty, *M*. & Just, *M*.A. (1993). Constructing Mental Models of the Machines from Text and Diagrams. *Journal of Memory and Language, 32,* 717-742.

Henderson, J.*M*. (2003). Human Gaze Control during Real-world Scene Perception. *Trends in Cognitive Science, 7(11),* 498-504.

Herbsleb, J.D., Klein, H., Olson, G.*M*., Brunner, H., Olson, H.S. & Harding, J. (1995). Object-oriented analysis and design in software project teams. *Human- Computer Interaction, 10*, 249-292.

Heydemann, *M*. (1986). The Relation between Eye-movements and Think Aloud for Raven Matrices, *Psychologische Beitrage, 28,* 76-87.

Hungerford, B. C., Hevner, A. R. & Collins, R. W. (2004). Reviewing Software Diagrams: A Cognitive Study. *IEEE Transactions on Software Engineering 30*, 2, 82-96.

Itoh, Y. (2005). The Facilitating Effect of Verbalization on the Recognition Memory of Incidentally Learned Faces. *Applied Cognitive Psychology, 19,* 421–433.

Iwasaki, Y. (1995). Problem Solving with Diagrams. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 657-667). Massachusetts: AAAI/MIT Press.

Jacob, R.J.K. (1991). The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At is What You Get. *ACM Transactions on Information Systems, 9 (*3) ,152-169.

Johansson, R. (2004). *Eye Movements during Visualizations of Pictures and Verbal Descriptions*. Master's Thesis. Cognitive Science, Lund University, Lund, Sweden.

Just, *M*.A. & Carpenter, *P*.A. (1985). Cognitive Coordinate Systems: Accounts of Mental Rotation and Individual Differences in Spatial Ability. *Psychological Review, 92,* 137-172.

Just, *M*. A., Carpenter, *P*. A. & Hemphill, D. D. (1996). Constraints on Processing Capacity: Architectural or Implementational. In D. *M*. Steier & T. *M*. Mitchell (Eds.), *Mind matters: A tribute to Allen Newell.* Mahwah, N.J.: Erlbaum.

Juristo, N., Moreno, *M*. & Vegas, S. (2003). Limitations of Empirical Testing Technique Knowledge. In N. Juristo and A.*M*. Juristo (Eds.), *Lecture Notes on Empirical Software Engineering, 12* (pp. 1-38). Singapore: World Scientific.

Kalyuga, S., Chandler, *P*. & Sweller, J. (1997). Levels of Expertise and User-Adapted Formats of Instructional Presentations: A Cognitive Load Approach. *Proceedings of the User Modeling: Proceedings of the Sixth International Conference (UM 97)*, 261-272.

KAMA. (2005). C4ISR Uygulamalari Ve Kavramsal Model Altyapı Çalışmalari Raporu (Kama-C4isrmos-Kamaçr-S00), 2 November 2005.

Kim, J., Hahn, J., and Hahn, H. (2000). How Do We Understand a System with (So) Many Diagrams? Cognitive Integration Processes in Diagrammatic Reasoning. *Information. Systems Research 11, 3,* 284-303.

Knoblich, G., Ohlsson, S. & Raney, G.E. (2001). An Eye Movement Study of Insight Problem Solving. *Memory & Cognition, 29,* 1534-1555.

Know, S.T., Bailey, W.A. & Lynch, E.F. (1989). Directed Dialogue Protocols: Verbal Data for User Interface Design. In *Proceedings of CHI'89* (*p*.283-287). Texas: ACM.

Koedinger, K. R. & Anderson, J. R. (1990). Abstract Planning and Perceptual chunks: elements of expertise in geometry. *Cognitive Science, 14,* 511-550.

Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo, California: Morgan Kaufmann.

Kosslyn, S., Seger, C., Pani, J. R. & Hillger, L. A. (1990). When Is Imagery Used in Everyday Life: A Diary Study. *Journal of Mental Imagery, 14 (3-4),* 131-152.

Larkin, J. & Simon, H. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cogn*itive Science, 11,* 65-99.

Levi, D.*M*., Klein, S.A. & Aitsobomo, A.*P*. (1985). Vernier Acuity, Crowding and Cortical Magnification. *Visual Research, 25(7),* 963-977.

Ligozat, G. (1999). Reasoning with Diagrams: The Semantics of Arrows. In I. Imam, Y. Kodratoff, A. El-Dessouki & Moonis Ali (Eds.), *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems IEA/AIE-99* (pp. 236-245), Springer.

Logie, R.H. & Della Sala, S. (2005). Disorders of Visuospatial Working Memory. In *P*. Shah & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 81-120). New York: Cambridge University Press.

Mackinlay, J. & Genesereth, *M*. R. (1985). Expressiveness and Language Choice. *Data Knowledge Engineering, 1(1),* 17–29.

Marian, V. & Spivey, *M*. (2003). Bilingual and Monolingual Processing of Competing Lexical Items. *Applied Psycholinguistics, 24(2),* 173-193.

McDougal, R.F. & Hammond, K.J. (1995). Using Diagrammatic Features to Index Plans for Geometry Theorem Proving. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 691-709). Massachusetts: AAAI/MIT Press.

Mozer, *M*.C. & Sitton, *M*. (1998). Computational Modeling of Spatial Attention. In H. Pashler (Ed.), *Attention* (pp. 293-341). London: UCL Press.

Narayanan, N.H., Suwa *M*. & Motoda H. (1995). Hypothesizing Behaviors from Device Diagrams. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 501-534). Massachusetts: AAAI/MIT Press.

Norris, S.*P*. (1990). Effect of Eliciting Verbal Reports of Thinking on Critical Thinking Test Performance. *Journal of Educational Measurement, 27,* 41-58.

Novak, G.S. & Bulko, W.C. (1990). Understanding Natural Language with Diagrams. *In Proceedings of the 5th International Conference on Artificial Intelligence* (pp. 467-470). San Francisco: Morgan Kaufmann.

Oxford Online Reference. (2006). A Dictionary of Psychology. Andrew *M*. Colman. Oxford University Press, 2006. Oxford Reference Online. Oxford University Press. Orta Dogu Teknik University. Retrieved on 27 August 2007 from http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t87.e8937

Pace, D.K. (1998). Impact of Simulation Description on Conceptual Validation. *Proceedings of the fall 1998 Simulation Interoperability Workshop*, 14-18.

Pace, D. K. (2000). Conceptual Model Development for C4ISR Simulations. *Proceedings of the 5th International Command and Control Research and Technology Symposium*, 24-26.

Qin, Y. & Simon, H. (1990). Imagery and Problem Solving. *Proceedings of the 12th Annual Conference of the Cognitive Science Society (pp. 646-653).* Hillsdale, NJ: Erlbaum.

Raaijimakers, L.R.W. & Shiffrin, R.*M*. (2002). Models of Memory. In H. Pashler and D. Medin (Eds.), *Steven's Handbook of Experimental Psychology, 2* (pp. 43-76). NY: Wiley.

Randsdell, S. (1995). Generating Think-aloud Protocols: Impact on the Narrative Writing of College Students. *American Journal of Psychology, 108(1),* 89-98.

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. 3-21). New York: Elsevier.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124(3),* 372-422.

Richardson, D.C., Dale, R. & Spivey, *M*.J. (2007). Eye Movements in Language and Cognition: A brief introduction. In *M*. Gonzalez-Marquez, I. Mittelberg, S. Coulson & *M*.J. Spivey (Eds.), *Methods in Cognitive Linguistics* (pp. 323-344). Ithaca, NY: John Benjamins Publishing Company.

Rogers, E. (1995). A Cognitive Theory of Visual Interaction. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 481-500). Massachusetts: AAAI/MIT Press.

Rossi, A., Daneluzzo, E., Tomassini, A., Struglia, F., Cavallaro, R., Smeraldi, E. & Stratta, *P*. (2006, January 26). The Effect of Verbalization Strategy on Wisconsin Card Sorting Test Performance in Schizophrenic Patients Receiving Classical or Atypical Antipsychotics. *BMC Psychiatry, 6 (3).* Retrieved 1 August, 2007 from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1373618

Rosson, *M*.B & Alpert, S.R. (1990). The Cognitive Consequences of Object-oriented Design. *Human- Computer Interaction, 5,* 345-379

Rumbaugh, J., Jacobson, I. & Booch, G. (1998). *The Unified Modeling Language Reference Manual*. Addison-Wesley

Salvucci, D.D. (1999). Mapping Eye Movements to Cognitive Processes. Doctorate Thesis. Computer Science, Carnegie Mellon University, Pittsburg, the USA.

Salvucci, D.D. (2000). An Interactive Model-Based Environment for Eye-Movement Protocol Analysis and Visualization. In *Proceedings of the Eye Tracking Research & Applications Symposium 2000* (pp. 57-63). Palm Beach Gardens, FL, USA: ACM.

Salvucci, D. D., & Anderson, J. R. (1998). Tracing Eye Movement Protocols with Cognitive Process Models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 923-928). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salvucci, D. D., & Anderson, J. R. (2001). Automated Eye-movement Protocol Analysis. *Human-Computer Interaction, 16*, 39-86.

Salvucci, D.D. & Goldberg, J.H. (2000). In *Proceedings of Eye Tracking Research & Applications Symposium* (pp.71-78). Palm Beach Gardens, FL: ACM.

Sanderson, *P.M*. (1990). Verbal Protocol Analysis in Three Experimental Domains Using SHAPA. In *Proceedings of the Human Factors Society-35th Annual Meeting, 2* (pp. 1280-1284). Santa Monica, CA: The Human Factors Society.

Sawamura, H. & Kiyozuka, K. (2000). In *M*. Anderson, *P*. Cheng & V. Haarslev (Eds.), *Diagrams 2000 (*pp. 271–285). Berlin Heidelberg: Springer-Verlag.

Schank, R.C. & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, N.J.: Lawrence Erlbaum.

Schiffrin, D. (1994). Approaches to Discourse. Oxford: Blackwell.

Schiffner, G. & Scheuermann, *P*. (1979). Multiple Views and Abstractions with an Extended-Entity Relationship Model. *Computer Languages, 4*, 139-154.

Schooler, J. W., Ohlsson, S. & Brooks, K. (1993). Thoughts Beyond Words: When Language Overshadows Insight. *Journal of Experimental Psychology: General, 122,* 166–183.

Schweiger, D.*M*. (1983). Is the Simultaneous Verbal Protocol a Viable Method for Studying Managerial Problem Solving and Decision Making? *Academy of Managerial Journal, 26(1),* 185-192.

Simon, H. (1995). Foreword. In J. Glasgow, N.H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives* (pp. 11-13). Massachusetts: AAAI/MIT Press.

Sims, V. K. & Hegarty, *M*. (1997). Mental Animation in the Visual-spatial Sketchpad: Evidence from dual-task studies. *Memory & Cognition, 25,* 321-332.

Sperling, G. (1963). A Model for Visual Memory Task. *Human Factors*, 5, 19-31.

Spivey *M*.J. & Geng, J.J. (2001). Oculomotor Mechanisms Activated by Imagery and Memory: Eye Movements to Absent Objects. *Psychological Research/Psychologische Forschung, 65(4),* 235-241.

Spivey-Knowlton, *M*., Tanenhaus, *M*., Eberhard, K. & Sedivy, J. (1998). Integration of Visuospatial and Linguistic Information in Real-time and Real-space. In *P*. Olivier & K. Gapp (Eds.), *Representation and Processing of Spatial Expressions* (pp. 201-214). Mahwah, NJ: Erlbaum.

Stenning, K. & Lemon, O. (2001). Aligning Logical and Psychological Perspectives on Diagrammatic Reasoning. *Artificial Intelligence Review, 15,* 29–62.

Suppes, *P*. (1990). Eye-movement Models for Arithmetic and Reading Performance. In E. Kowler (Ed.), *Eye Movements and their Role in Visual and CognitiveProcesses* (pp. 455-477). New York: Elsevier.

Tanenhaus, *M*.K. & Trueswell, J.C. (1995). Sentence Comprhension. In J.L. Miller & *P*.D. Eimas (Eds.), *Speech, Language and Communication, 11,* (pp. 217-262). San Diego, CA, US: Academic Press.

The Observer Reference Manual Version 5.0. Wageningen, the Netherlands: Noldus Information Technologies, May 2003.

Todd, *P*. & Benbasat, I. (1987). Process Tracing Methods in Decision Support System: Exploring the Black Box. *MIS Quarterly, 11(4),* 493-512.

Wilson, B., Baddeley, A.D. & Young, A.W. (1999). LE, A Person Who Lost Her 'Mind's Eye.' *Neurocase*, 5, 119-127.

Wolfe, J. *M*. (1998). Visual Search. In H. Pashler (Ed.), *Attention (*pp. 13-71*)*. London, UK: University College London Press.

Wooding, D.S. (2002). Fixation Maps: Quantifying Eye-movement Traces. In *Proceedings of ETRA'02* (pp.31-36). New Orleans, Louisiana: ACM.

Yoon, D. & Narayanan, N.H. (2004). Predictors of Success in Diagrammatic Problem Solving. In A. Blackwell et al. (Eds.), *Diagrams 2004* (pp. 301-315). Berlin Heidelberg: Springer-Verlag.

Zhang, J. (1997). The Nature of External Representations in Problem Solving. C*ognitive Science, 21(2),* 179-217.

APPENDICES

APPENDIX A. Written Documents of the Main Experiment

The Warm-up Task

Yüksek Sesle Düşünme Hazırlık Etkinliği

Bu çalışmada, kavramsam modelleme çizimlerine yerleştirmiş olduğumuz kasıtlı hataları bulmanızı ve bu hataları bulurken yüksek sesle düşünmenizi istiyoruz. Hataları ararken ve hata olduğunu düşündüğünüz bir şey bulurken aklınıza gelen her şeyi yüksek sesle dile getirmenizi istiyoruz.

Çalışma sırasında bu odada yalnız olacaksınız ve uzun süre yüksek sesle düşünmeye arar verirseniz, "Lütfen konuşmaya devam eder misiniz?" diye uyarılacaksınız. Hangi çizimde ne kadar süre harcayacağınız ve ne zaman çalışmaya son vereceğinize siz karar vereceksiniz.

Çalışmaya başlamadan önce, yüksek sesle düşünme çalışması yapalım:

İlk örneği ben yapacağım, diğer iki çalışmayı siz yapacaksınız:

1. **İsveç'in başkenti neresidir?**

   İlk aklıma gelen Avrupa haritası. İskandinavya'ya yoğunlaşıyorum. Norveç, İsveç, Finlandiya aklıma geliyor. Oslo'yu anımsıyorum. Oslo Norveç'in başkentiydi. Şimdi aklıma İsveç'te verilen Nobel Ödülleri geliyor. Ödüller Stockholm'de veriliyor. Cevap Stockholm.

Şimdi sıra sizde.

2. 12 x 17 kaçtır?

3. 10 tane hayvan adı söyleyiniz. (Sayısını tutmayın, ben sizin için tutarım).

4. Nasıl çay demlersiniz? Anlatınız.

Scenario

Deney Senaryosu

Açıklama: Bu çalışmada, KAMA projesi çerçevesinde geliştirilen kavramsal model oluşturma notasyonunun bilişsel özellikleri incelenmektedir. Bu amaçla atıcı durum diyagramı, tutukluk giderme iş akış diyagramı, atış iş akış diyagramı, atış ve atış sonrası değerlendirme iş akış diyagramı ve bu diyagramların sözel anlatımları kullanılacaktır. Diyagramların işlevleri ve sözel anlatımları aşağıda verilmiştir. Deney sırasında göreceğiniz diyagramlar, verilen sözel senaryodan <u>farklı ve kasıtlı hatalar</u> içerecektir. Katılımcıdan istenen, senaryo ve çizim arasındaki bu farklılıkları bulması ve bulurken ayrıntıları deney sırasında anlatılacak olan sesli düşünme metodunu kullanmasıdır. Bu sırada bir göz takip cihazı da katılımcının göz hareketlerini kaydedecektir. Katılımcıların iş, tecrübe, yaş vb. betimleyici bilgileri haricinde kesinlikle isim, adres gibi kişisel bilgileri istenmeyecek ve toplanan bilgiler gizli tutulacaktır.

Atıcı Durum Diyagramı: Atış görevi sırasında atıcının bulunduğu durumları gösteren diyagramdır. Durumlar arasında geçiş eylemler ve emirler doğrultusunda gerçekleşmektedir.

Tutukluk Giderme İş Akış Diyagramı: Bu diyagram tutukluk giderme işinin hangi koşullar altında nasıl yapıldığını göstermektedir. Her eylemin bir aktörü olmalıdır. Tutukluk giderme işinin aktörü atıcıdır.

Atış İş Akış Diyagramı: Atış eğitimi görevinin nasıl yapıldığını anlatan diyagramdır. Atış iş akışının aktörleri eğitici, nezaretçi ve atıcıdır.

Atış ve Atış Sonu Değerlendirme İş Akış Diyagramı: Atış eğitimi içerisindeki atış etkinliğinin ve atış değerlendirmesinin nasıl yapıldığını anlatan diyagramdır. Atış ve atış sonu değerlendirme iş akışının aktörleri eğitici, nezaretçi ve atıcıdır.

Atıcı Durum Senaryosu

Atıcı başlangıçta emir bekleme durumundadır. Eğiticiden gelen silahı doldur emri doğrultusunda, silah doldurma durumuna geçer. Silah doldurulunca emir bekleme durumuna döner. Emir bekleme durumundayken eğiticiden emir gelirse, silah boşaltma durumuna geçer. Silah boşaltılınca emir bekleme durumuna geri döner. Atıcı emir bekleme durumunda, eğiticinin emriyle nişan vaziyetine geçer ve nişan vaziyeti alınınca emir bekleme durumuna döner. Atıcı eğiticiden gelen emir doğrultusunda, ateş etme durumuna geçer. Ateş etme durumunda atışını bitirirse emir bekleme durumuna geri döner. Ateş etme durumunda silah tutukluk yaparsa, tutukluk giderme durumuna geçer. Tutukluk giderilirse, ateş etme durumuna geri döner. Tutukluk giderilemezse, silah boşaltma durumuna geçer. Silah boşaltılınca, emir bekleme durumuna döner. Atıcı emir bekleme durumundan, eğitici emriyle hedefe gitme durumuna geçer. Eğitici hedefi kontrol eder. Bu kontrol sonucunda, atıcı başarılı durumda veya başarısız durumda olabilir.

Tutukluk Giderme İş Akışı Senaryosu

Tutukluluk 3 şekilde gözlenir: Namluya fişek sürülemiyordur, silah ateş almıyordur veya kovan dışarı atılamıyordur. Her bir durum için farklı tutukluk giderme yöntemi izlenir. Eğer namluya fişek sürülemiyor ise ve fişek şarjörden çıkmıyor ise şarjör oturmamış olabilir, şarjör arızalı olabilir veya fişek arızalı olabilir. Şarjör oturmamış ise atıcı şarjörü oturtur. Şarjör arızalı ise atıcı şarjörü değiştirir. Fişek arızalı ise atıcı fişeği değiştirir. Eğer namluya fişek sürülemiyor ise ve fişek şarjörden çıkıyor ise yine fişek arızalı olabilir, fişek yatağı kirli olabilir, YGY (yerine getiren yayı) arızalı olabilir veya mekanizma kirli olabilir. Fişek arızalı ise atıcı fişeği değiştirir. Fişek yatağı kirli ise atıcı yatağı temizler. YGY arızalı ise atıcı YGY'nı değiştirir. Mekanizma kirli ise atıcı tüfeği temizler.

Silahın ateş almadığı tutukluk durumunda, iğne kapsüle yeterince çarpmıyordur veya iğne çarpıyordur. İğne kapsüle yeterince çarpmıyor ise iğne veya yayı arızalı olabilir. Bu durumda atıcı iğne ve yayı değiştirir. İğne çarpıyor ise fişek arızalı olabilir. Bu durumda atıcı fişeği değiştirir.

Kovanın dışarı atılamadığı tutukluk durumunda, kovan yataktan çıkmıyordur veya çıkıyordur. Kovan yataktan çıkmıyor ise yatak kirli olabilir veya tırnak arızalıdır. Yatak kirli ise atıcı yatağı temizler. Tırnak arızalı ise atıcı tırnağı değiştirir. Kovan yataktan çıkıyor ise kovan atacağı arızalıdır. Bu durumda atıcı kovan yatağını değiştirir.

Atış İş Akışı Senaryosu

Eğitim atışında, atış grubundaki herkesin bir mekanik nişancılık tekamül kartı vardır ve bir önceki görevdeki başarı durumunu gösterir. Eğitici, atış grubu içerisinden atıcı ve nezaretçiyi bu karta göre seçer. Eğiti atışla ilgili bilgi verir. Atıcı, eğiticinin ateş hattına marş emri ile ateş hattına gider. Nezaretçi, eğiticinin hedef tak emri ile hedef kağıtlarını takarak hedefleri oluşturur. Atıcı, eğiticinin YDNV (Yatarak Destekli Nişan Vaziyeti) al emri ile nişan vaziyeti alır. Aynı anda aynı emirle nezaretçi de atıcının yanına çöker. Eğitici emri ile kontrol nezaretçiye geçer ve nezaretçi atıcının nişan vaziyetini kontrol eder ve hatalarını düzeltir. Daha sonra nezaretçi şarjöre fişekleri doldurur. Ardından gelen eğiticinin doldur-kapat emri ile, atıcı şarjörü takar ve sonra doldurur.

Daha sonra, eğiticinin ateş serbest emri ile atış ve atış sonu değerlendirme etkinliği başlar. Bu etkinliğin sorumlusu eğiticidir. Atış ve atış sonu değerlendirme etkinliği sırasında, mermi yolu, namlu hareketi ve geri tepme oluşur. Atış ve atış sonu değerlendirme etkinliği sırasında, tutukluk olmazsa, atış ve atış sonu değerlendirme etkinliğine devam edilir. Tutukluk olursa, atıcı tutukluk giderme etkinliğine geçer. Tutukluk giderilirse, atıcı doldur etkinliğine döner. Tutukluk giderilemezse, atış görevi sona erer. Atış ve atış sonu değerlendirme etkinliği biterse, atış eğitimi sona erer.

Atış ve Atış Sonu Değerlendirme İş Akışı Senaryosu

Atış ve atış sonu değerlendirme etkinliği, atıcının tetik düşürmesi ile başlar. Daha sonra silahın tutukluk durumu kontrol edilir. Tutukluk varsa, atıcı silahı boşaltır ve nezaretçi kırmızı flama kaldırır. Atıcı tutukluk giderme işine geçer ve daha sonra başa döner. Tetik düşürme sırasında tutukluk yoksa nezaretçi atıcıyı kontrol eder ve tüfek hakimiyet kartına işler. Atıcı hataları bariz ve devamlı ise atıcı silahı boşaltır ve atış görevi sona erdirilir.

Hatalar bariz ve devamlı değilse atış miktarı ve atış süresi kontrol edilir. Atıcı tahsis edilen miktarda atış yaptıysa nezaretçi yeşil flama kaldırır. Atıcı tahsis edilen miktarda atış yapmadıysa süre şartı varsa süresi dolup dolmadığı kontrol edilir. Süresi dolmadıysa veya süre şartı yoksa atıcı tetik düşürme işine devam eder. Süre şartı varsa ve dolduysa atıcı silahı boşaltır ve nezaretçi yeşil flama kaldırır. Eğiticinin hedef hattına marş emri ile atıcı hedef kağıdını getirir. Eğitici hedefi kontrol ederek hataları gözler ve hataları bildirir ve öneriler üretir. Nezaretçi sonuçları nezaretçi kontrol formuna, atış kayıt defterine ve atış karnesine işler. Atış ve atış sonu değerlendirme görevi sona erer.

Informed Consent Form

Bilgilendirme Onay Formu

Bu çalışmada KAMA projesi çerçevesinde geliştirilen gösterimlerin bilişsel incelemesi yapılacaktır. Katılımcıların göz hareketleri, sesleri ve video görüntüleri kaydedilecektir. Deney öncesinde ve sonrasında doldurulacak bilgi formlarındaki her türlü bilgi ile katılımcıların deney verileri kesinlikle gizli tutulacaktır ve yalnızca bu çalışma için kullanılacaktır.

Deney Sorumlusu: Özkan KILIÇ. Tel: 0536 266 92 47 e-mail: ozkankilic@gmail.com

Yukarıdaki açıklamayı okudum ve onaylıyorum.

İsim:

Tarih:

İmza:

Demographic Data Form

Kişisel Bilgi Formu

Katılımcı Numarası:

Yaşınız:

Cinsiyetiniz:

Mezun olduğunuz üniversite - mezuniyet yılı - alanınız:

<u>Lisans</u>	**Üniversite:**
	Yıl:
	Alan
<u>Y.Lisans</u>	**Üniversite:**
	Yıl:
	Alan:
<u>Doktora</u>	**Üniversite:**
	Yıl:
	Alan:

Asker iseniz kaç yıldır bu işi yapıyorsunuz:

Asker değilseniz ve askerliğinizi yaptıysanız kaç yıl önce askerliğinizi bitirdiniz:

Yazılımcı/Mühendis iseniz kaç yıldır bu işi yapıyorsunuz:

Daha önce yazılım geliştirme konusunda formal bir eğitim aldınız mı:

Kavramsal Modelleme ile ilgili bilginiz var mı:

Kavramsal modelleme ile ilgili şimdiye kadar katıldığınız tahmini proje sayısı:

Yazılım geliştirmede kullanılan notasyonlardan (ERD, DFD, UML ...)hangileri ile ilgili bilgi sahibisiniz ve seviyeniz (Az, Orta, İyi, Çok İyi) nedir?

*

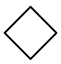
*

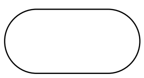
*

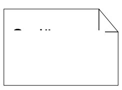
Simgeler ve Kısaltmalar

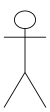
 **Başlangıç:** Durum ve iş akış diyagramlarında senaryo başlangıç simgesidir.

 **Bitiş:** Durum ve iş akış diyagramlarında senaryo bitiş simgesidir.

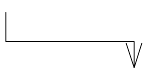
 **Karar:** Senaryo akışını etkileyen kararların alındığı yeri gösteren simgedir.

 **İş:** Bir aktör tarafından gerçekleştirilen veya sorumlusu olan, girdi-çıktı objeleri olabilen aktiviteleri gösteren simgedir.

 **Durum:** Gerçekleştireni olmayan ve kavramsal modelleme elemanlarının bulunduğu durumları gösteren simgedir.



Aktör: İş elemanlarını gerçekleştiren yada bunlardan sorumlu olan kişiyi gösteren simgedir. Durum elemanlarının aktörü yoktur.



Geçiş: Senaryoda akışların yönünü gösteren simgedir.

Girdi-Çıktı Nesnesi: İş elemanlarının gerçekleşmesi için gerekli girdi nesnelerini veya iş sonucunda üretilen çıktı nesnelerini gösteren simgedir. Nesnelerin ait oldukları sınıfları da olabilir. Örnek: "Hedef Hattına Marş" nesnesi, "Eğitici Emirleri" sınıfındandır. "Karga" nesnesi, "Kuşlar" sınıfındandır.

Çoklu bağlantı sağlama: İş akışında birden fazla dala ayrılmanın olduğu durumlarda, birden fazla iş dalından gelindiğinde veya birden fazla karar elemanının gerçekleşmesine bağlı olan durumlarda bağlantı gösteren simgedir.

Aktörleri ve girdi-çıktı nesnelerini, iş diyagramlarına bağlayan simgedir.

YGY Yerine getiren yayı. Silah mekanizmasının bir parçasıdır.

Mek.Niş.Tek. Kartı	Mekanik nişancılık tekamül kartı. Bir önceki görevdeki başarı durumunu gösteren atıcı kartıdır.
YDNY Al	Yatarak destekli nişan vaziyeti al.
Nez.Kont.Form	Nezaretçilerin atıcı bilgileri için tuttuğu formdur.
Tüf.Hak.Kart	Tüfek hakimiyet kartı. Atıcının tüfek hakimiyeti hakkındaki kartı.

Error Report From

HATA RAPOR FORMU

Hata No	Hatanın olduğu diyagram	Açıklama
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		
10		
11		

Interview Questions

Deney Sonrası Görüşme

KATILIMCI NUMARASI:

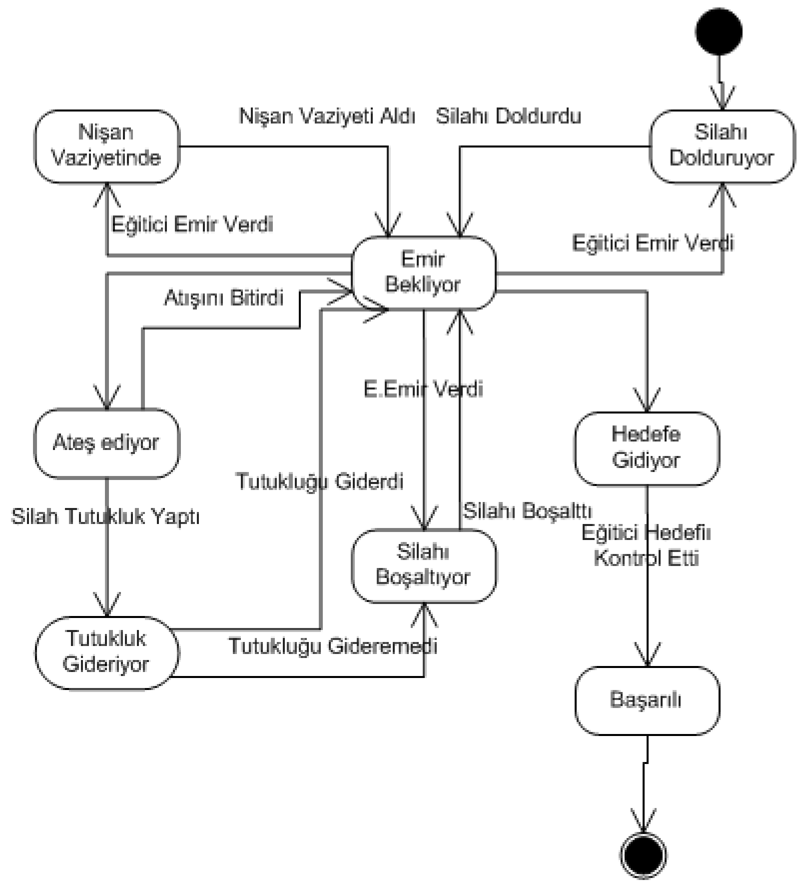
Aşağıdaki soruların tümünü cevaplandırmak zorunda değilsiniz. Doldurabildiğiniz kadarı yeterli olacaktır.

1. Hangi hataları kolaylıkla buldunuz?
   a. Durum diyagramındaki hataları (Atıcı durum)
   b. İş akış diyagramlarındaki hataları (Tutukluk giderme, atış, atış ve atış sonu değerlendirme iş akış diyagramları)
2. Hataları kolaylıkla bulmanızı sağlayan faktörler nelerdi?
3. Hataları bulmanızı zorlaştıran faktörler nelerdi?
4. Hataları bulurken neler yaptınız, genel olarak hangi yolu izlediniz?
5. Hata olduğundan emin olmadığınız bir durumla karşılaştınız mı? Neden emin olamadınız?
6. Bu gösterimle çalışırken istediğiniz herhangi bir sırayı izlediniz mi, yoksa sistem sizi belli bir sırayı takip etmek zorunda bıraktı mı?
7. Sizce Atıcı Durum Diyagramı ne kadar karışıktı? Anlaşılabilirliğini 0-5 arasında değerlendiriniz (5: çok kolay anladım)
8. Sizce Tutukluk Giderme İş Akış Diyagramı ne kadar karışıktı? Anlaşılabilirliğini 0-5 arasında değerlendiriniz (5: çok kolay anladım)
9. Sizce Atış İş Akış Diyagramı ne kadar karışıktı? Anlaşılabilirliğini 0-5 arasında değerlendiriniz (5: çok kolay anladım)
10. Sizce Atış ve Atış Sonu Değerlendirme İş Akış Diyagramı ne kadar karışıktı? Anlaşılabilirliğini 0-5 arasında değerlendiriniz (5: çok kolay anladım)
11. Çizimin anlaşılmasını zorlaştıran faktörler nelerdi? Siz olsaydınız neleri değiştirirdiniz?

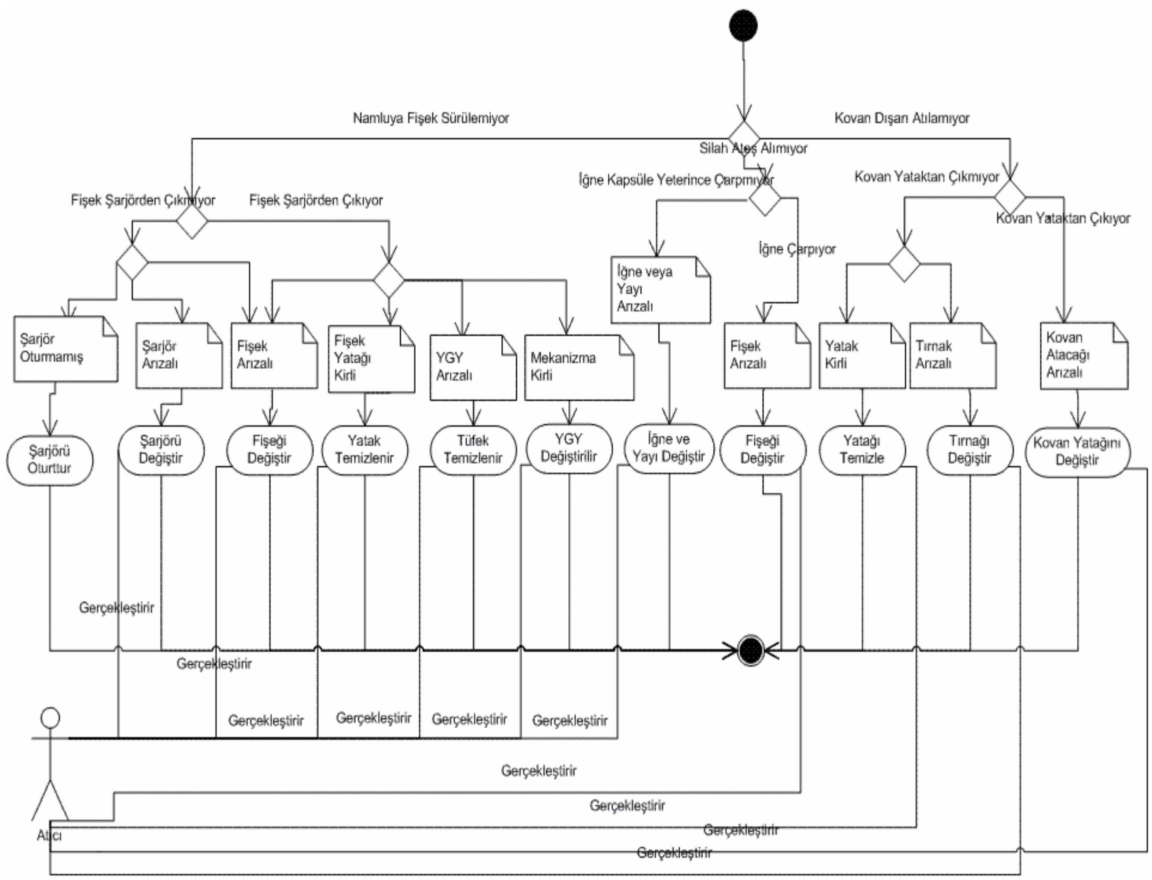
12. Senaryonun anlaşılmasını zorlaştıran faktörler nelerdi? Siz olsaydınız neleri değiştirirdiniz?

APPENDIX B. Diagrams of the Main Experiment

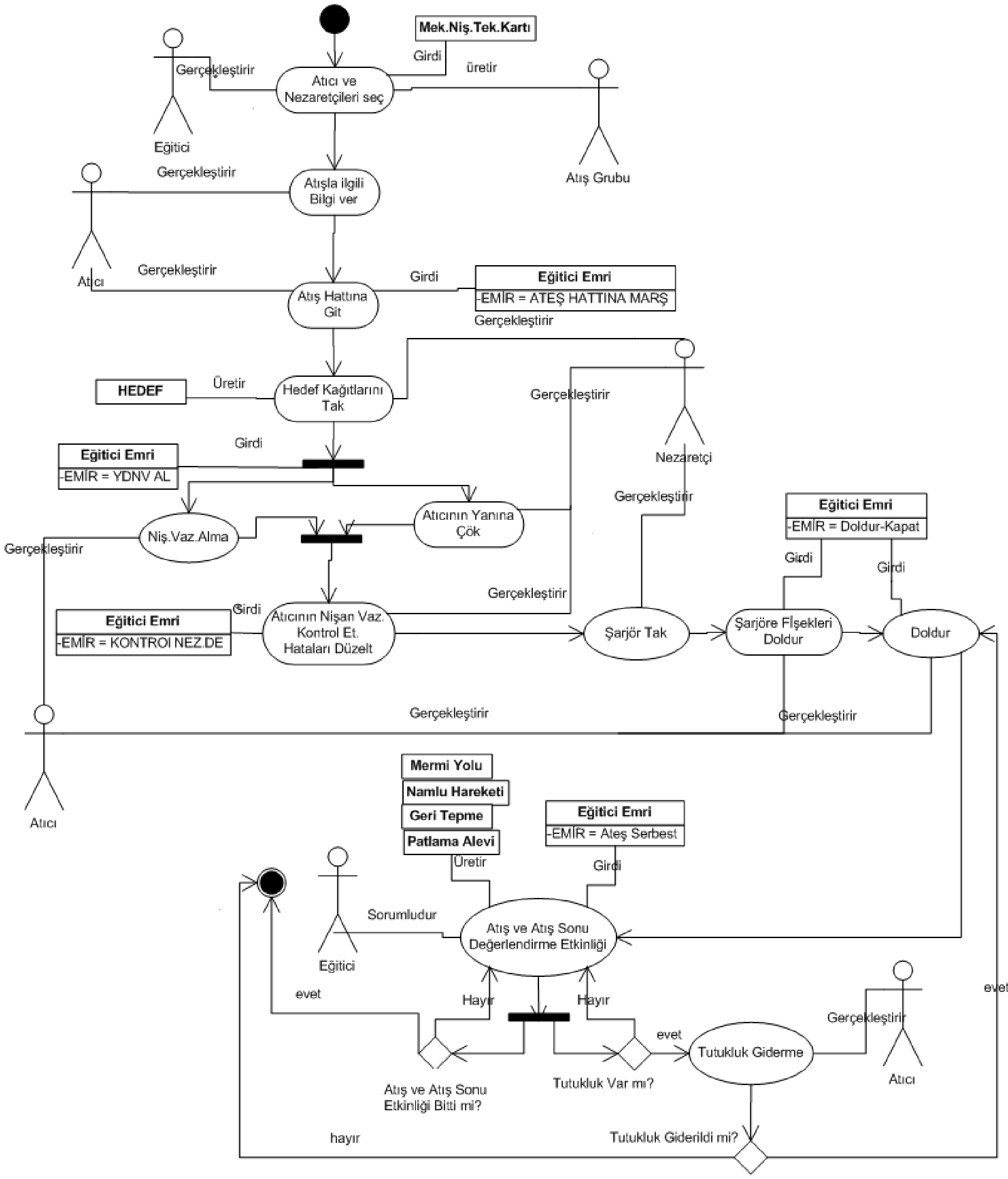
The Marksman State Diagram



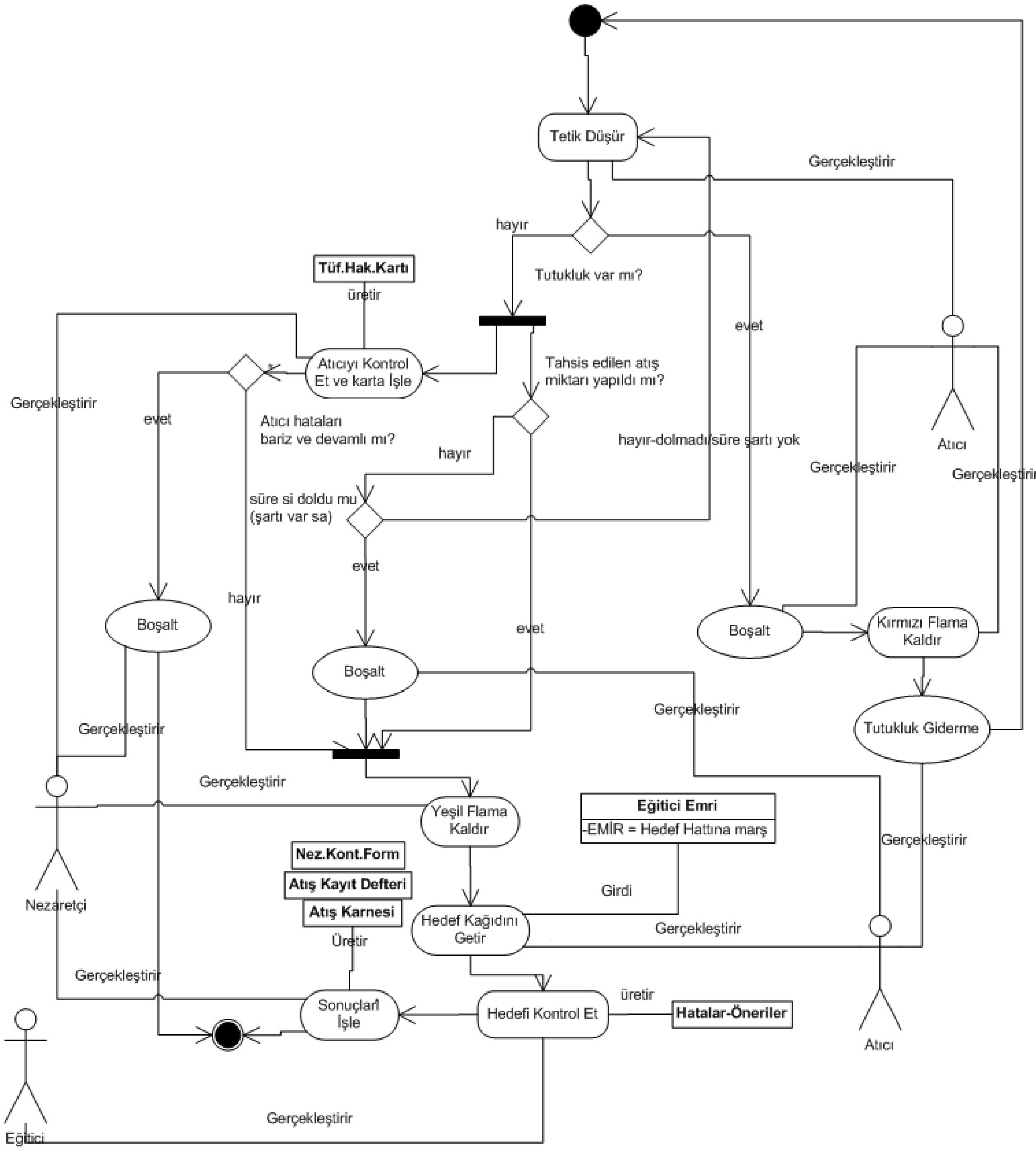
The Stoppage Removal Activity Diagram



The Marksman Mission Training Activity Diagram



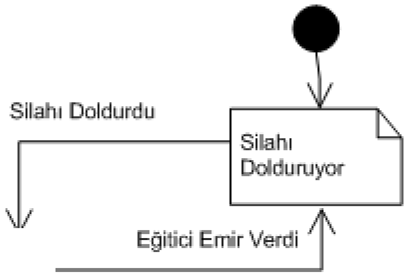
The Marksman Mission Ending and Evaluating Activity Diagram



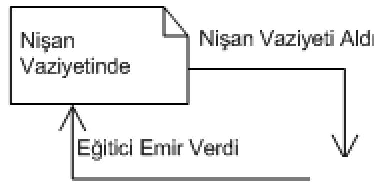
Sub-parts of the Marksman State Diagram

p1_a1

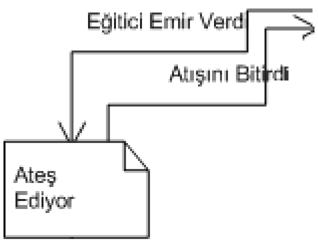
p1_a2



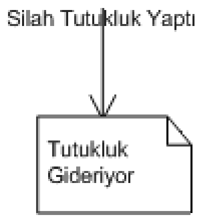
p1_a3



p1_a4

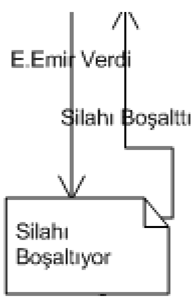


p1_a5

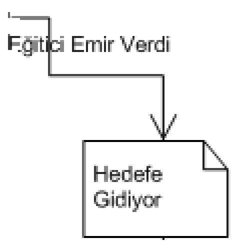


p1_a6

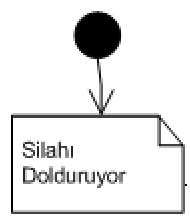
p1_a7



p1_a8

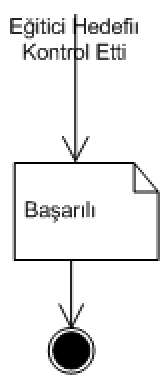


p1_e1



p1_e2

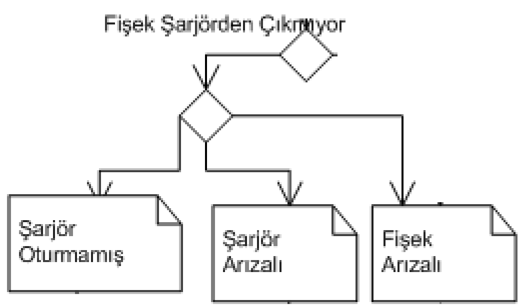
p1_e3



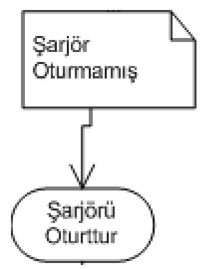
Sub-parts of the Stoppage Removal Activity Diagram

p2_a1

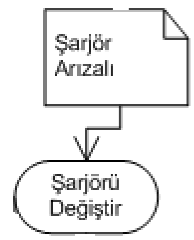
p2_a2



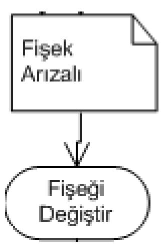
p2_a3



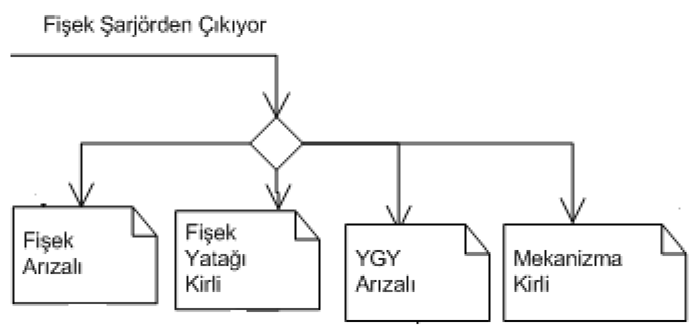
p2_a4



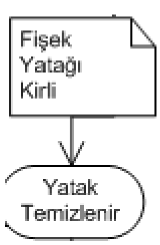
p2_a5



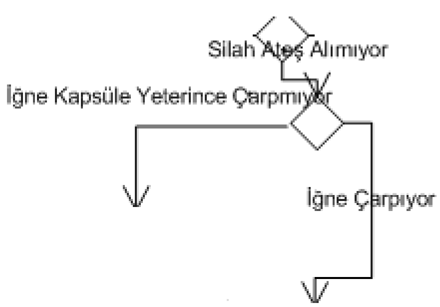
p2_a6



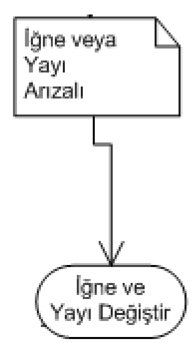
p2_a7



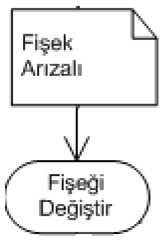
p2_a8



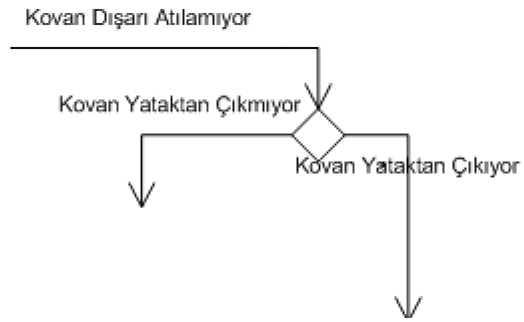
p2_a9



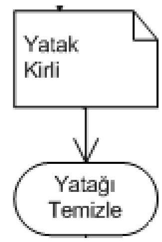
p2_a10



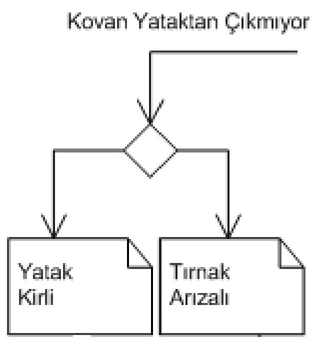
p2_a11



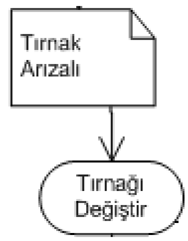
p2_a13



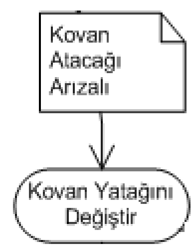
p2_a12



p2_a14

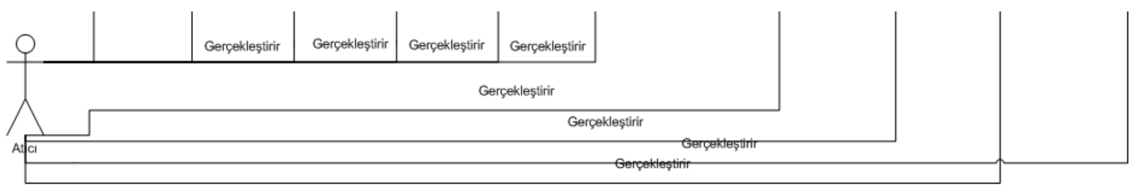


p2_a15

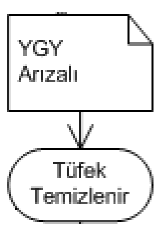


p2_a16

p2_a17

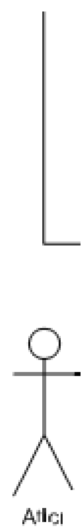


p2_e1



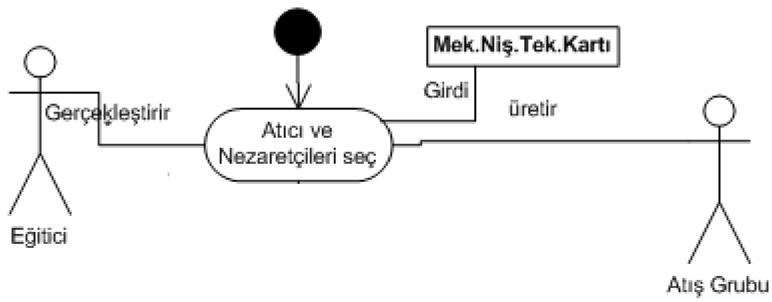
p2_e2

p2_e3

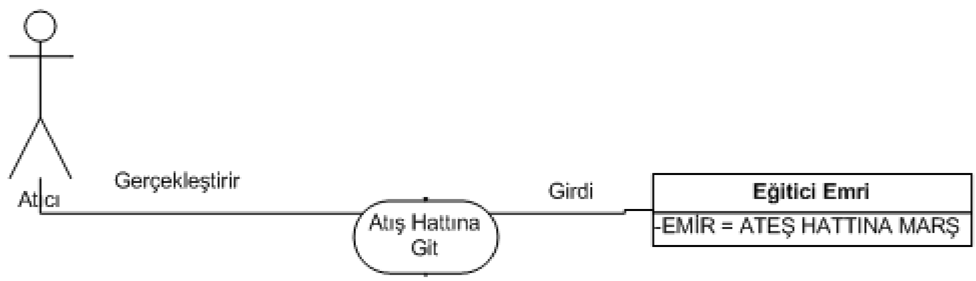


Sub-parts of the Marksman Mission Training Activity Diagram

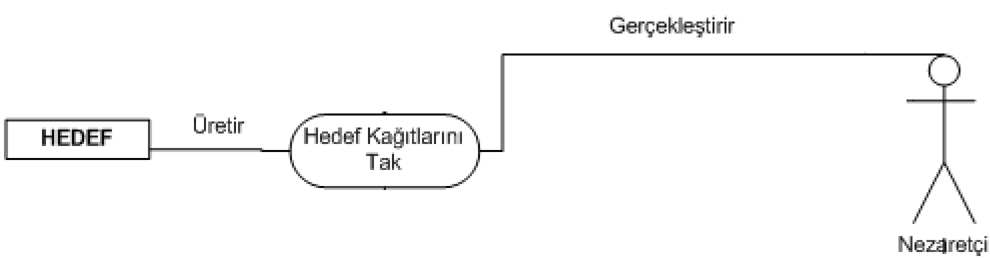
p3_a1



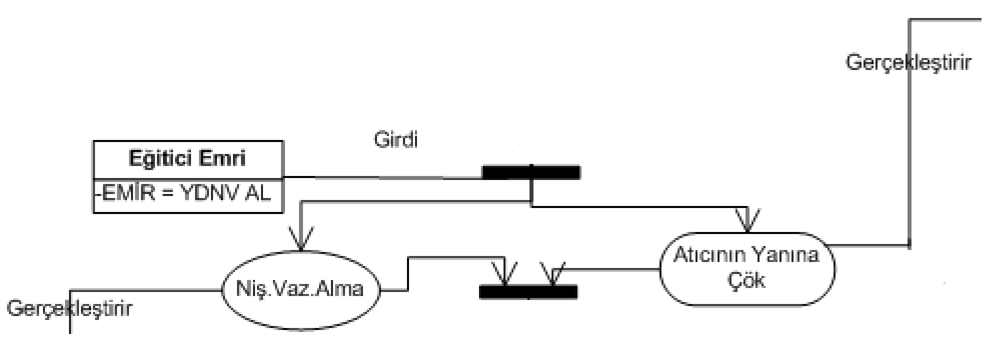
p3_a2



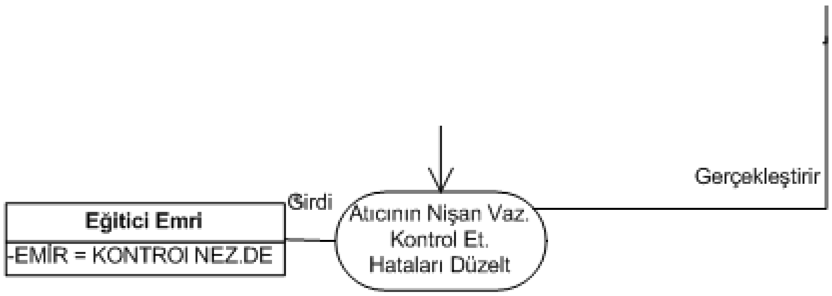
p3_a3



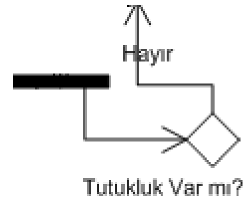
p3_a4



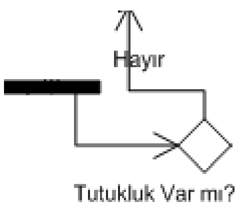
p3_a5



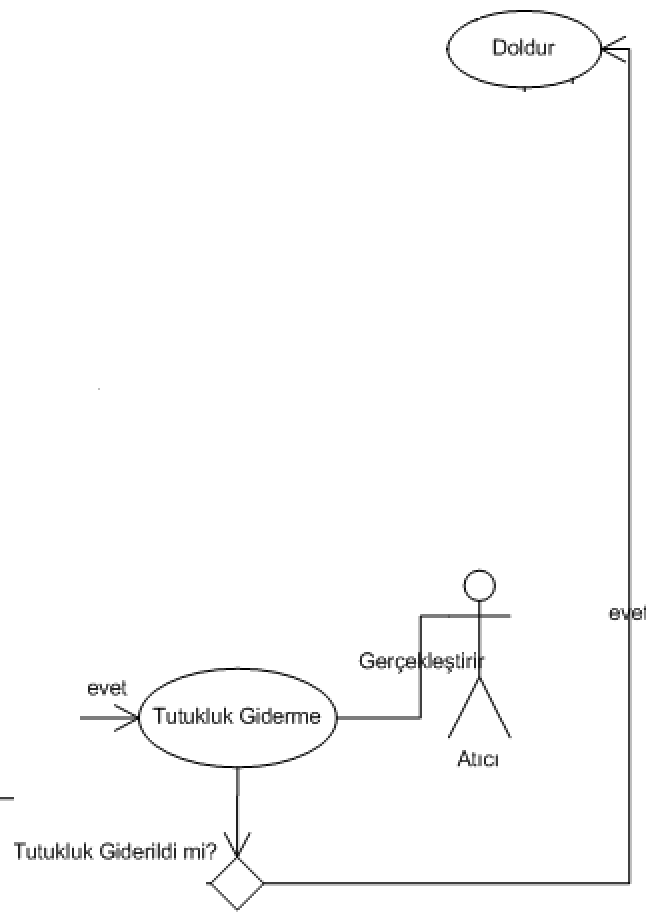
p3_a6



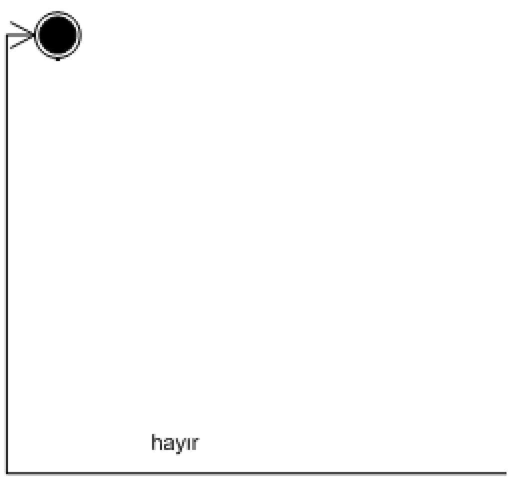
p3_a7



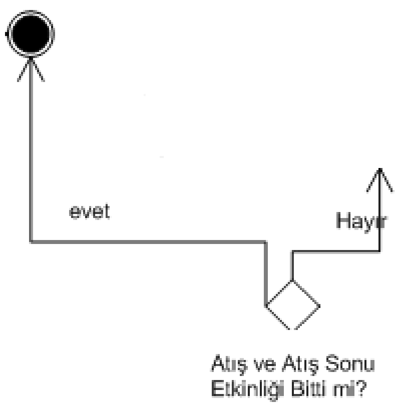
p3_a8



p3_a9

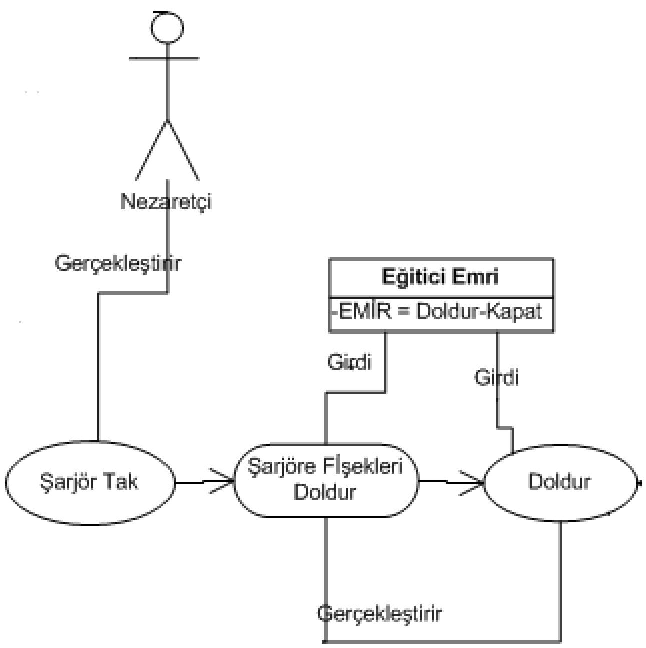


p3_a10



p3_e1

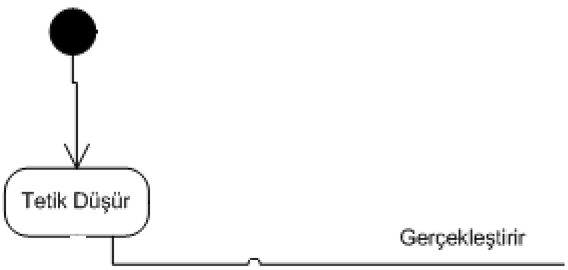
p3_e2



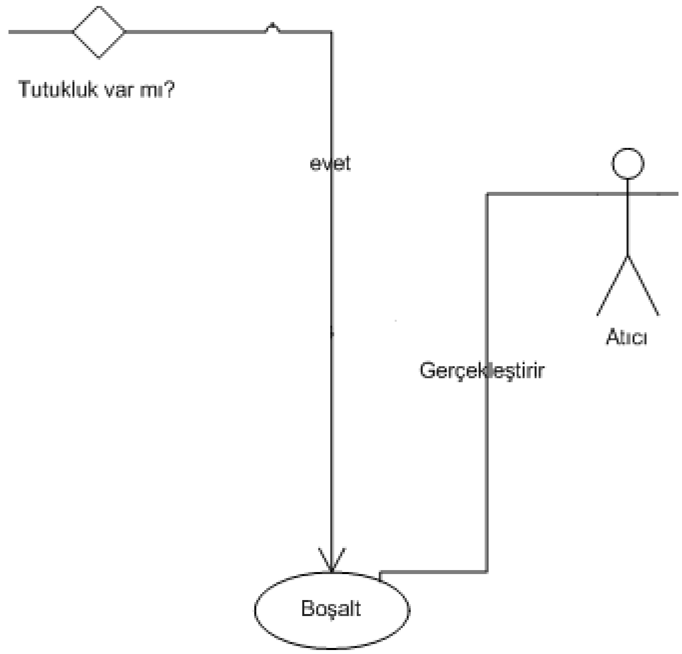
p3_e3

Sub-parts of the Marksman Mission Ending and Evaluating Activity Diagram

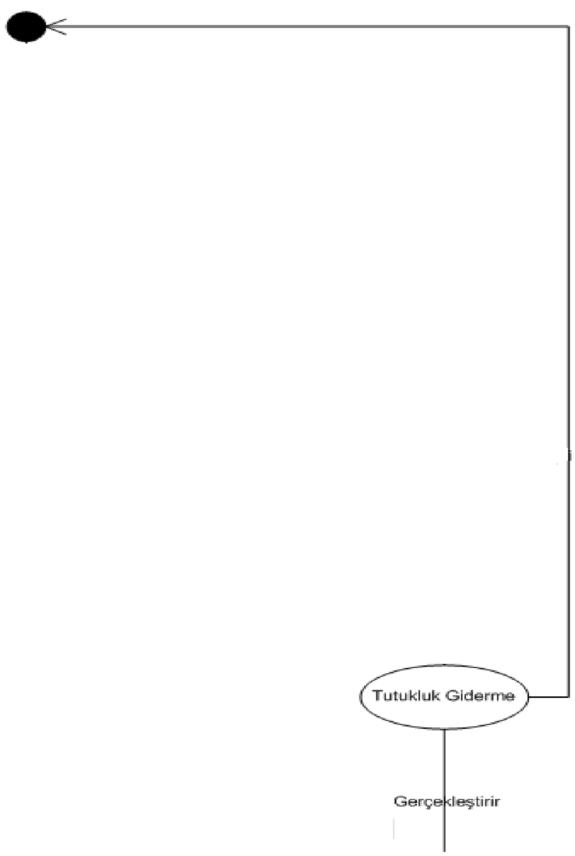
p4_a1



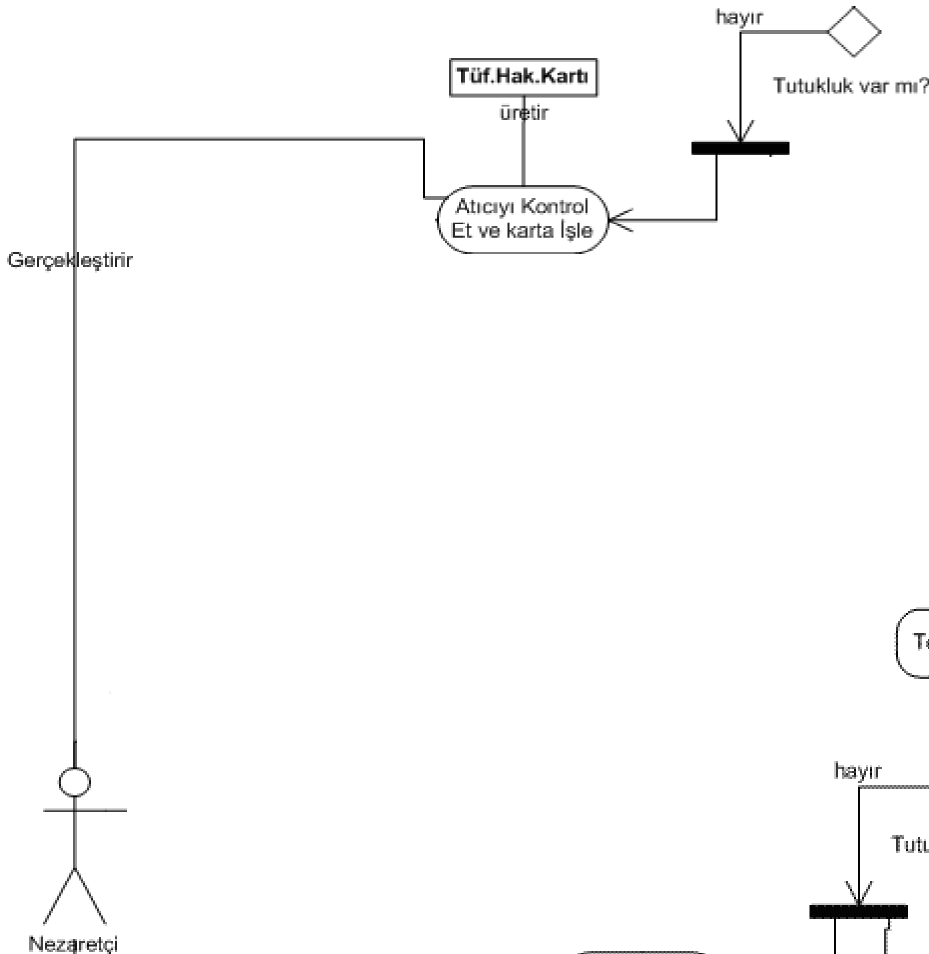
p4_a2



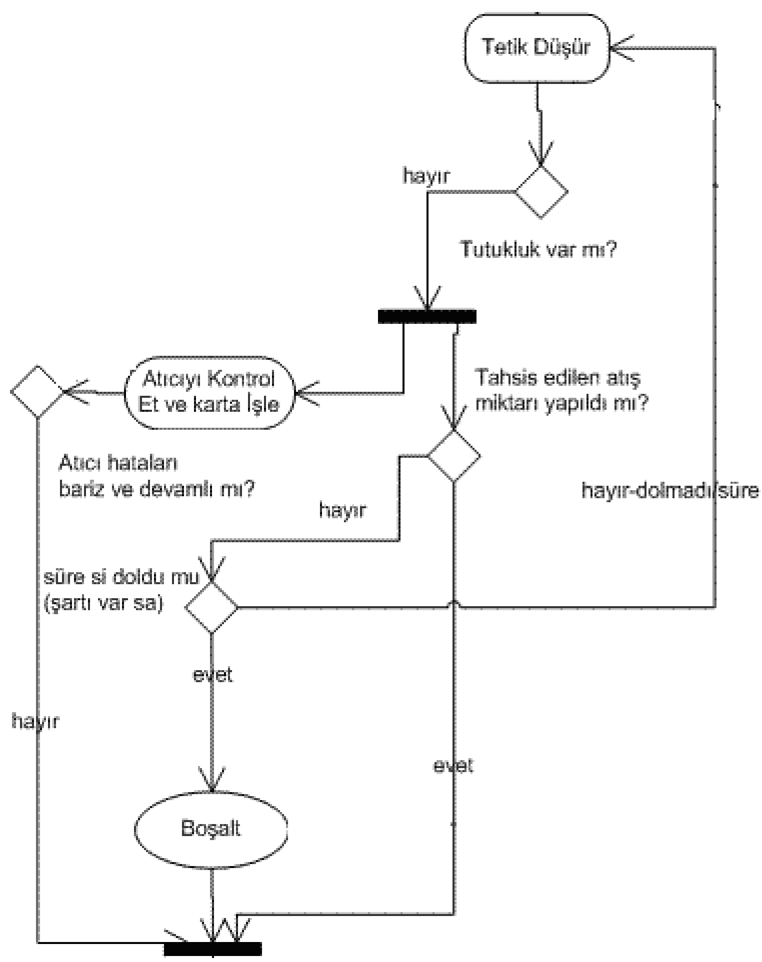
p4_a3



p4_a4

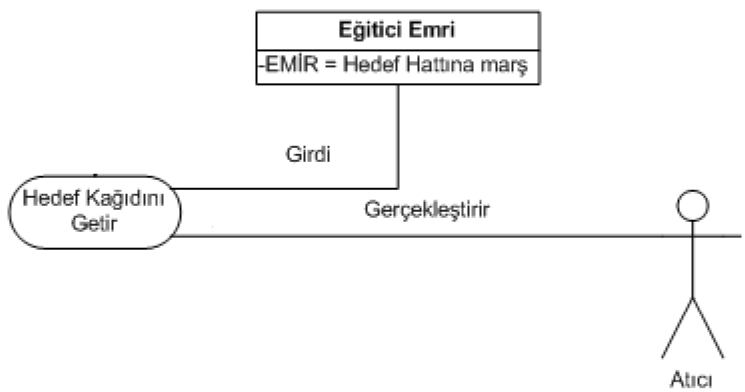

hayır
Tutukluk var mı?
Tüf.Hak.Kartı
üretir
Atıcıyı Kontrol Et ve karta İşle
Gerçekleştirir
Nezaretçi

p4_a5

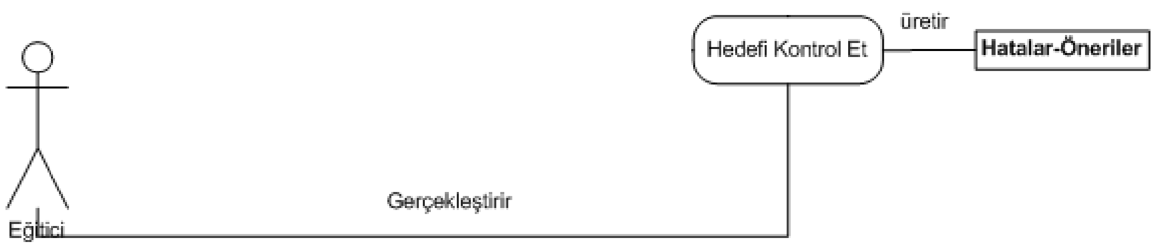

Tetik Düşür
hayır
Tutukluk var mı?
Atıcıyı Kontrol Et ve karta İşle
Tahsis edilen atış miktarı yapıldı mı?
Atıcı hataları bariz ve devamlı mı?
hayır
hayır-dolmadı/süre
süre si doldu mu (şartı var sa)
evet
hayır
evet
Boşalt

p4_a6

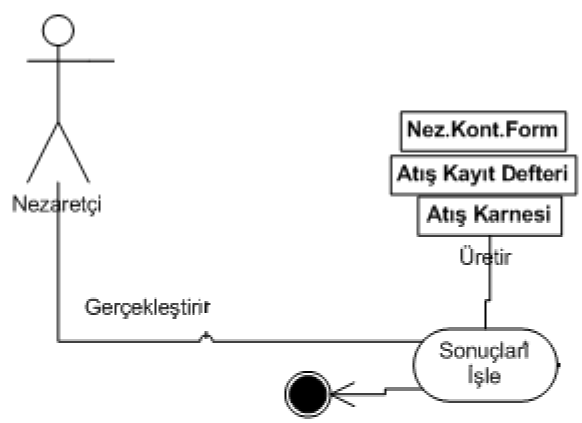
p4_a7



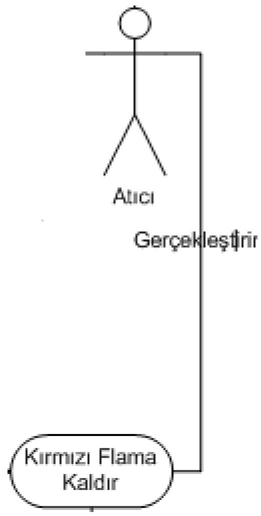
p4_a8



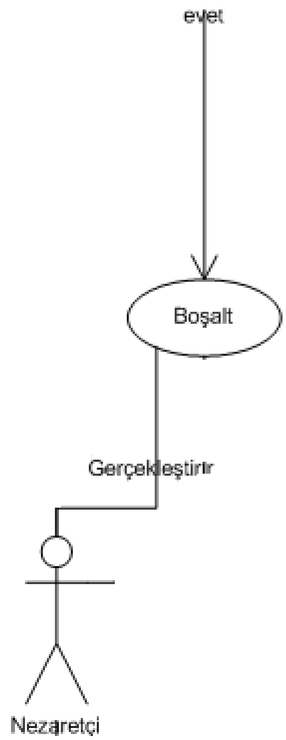
p4_a9



p4_e1


Atıcı
Gerçekleştirir
Kırmızı Flama
Kaldır

p4_e2


evet
Boşalt
Gerçekleştirir
Nezaretçi

APPENDIX C. Scenario and Diagrams of the Follow-up Experiment

Scenario of the Follow-up Experiment

Motorlu Taşıtlar Vergisi Ödeme İş Akış Diyagramı

Motorlu taşıtlar vergisi 3 farklı şekilde ödenir: Otomobil için, toplu taşıma ve yük aracı için ve motorlu deniz aracı icin. Otomobil 6 yaşından küçük ise, spor araç olup olmadığı kontrol edilir. Spor araç ise araç sahibi en az 329 YTL öder ve iş sona erer. Spor araç değil ise araç sahibi en az 229 YTL öder ve iş sona erer. Araç 6 yaşından küçük değil ise, araç sahibi 38YTL öder ve iş sona erer.

Toplu taşıma ve yük aracı icin, 1500kg'a kadar ise, araç sahibi kapasiteye göre düşük vergi öder ve iş sona erer. 1500-10000kg arasında ise araç, araç sahibi kapasiteye göre orta seviye vergi öder ve iş sona erer. 10000kg'dan ağır ise araç sahibi kapasiteye göre yüksek vergi öder. Daha sonra aracın 15 yaşından büyük olup olmadığı kontrol edilir. 15 yaşından büyük ise, araç sahibi satış işlemi başlatır ve bu işlemden sonra gelir vergisi ödeyer ve iş sona erer. 15 yaşındaın büyük değil ise iş sonra erer.

Son olarak motorlu deniz aracı için beygir gücü (BG) kontrol edilir. 50BG'ne kadar ise, araç sahibi en fazla 27YTL öder. Daha sonra aracın turizmde kullanılıp kullanılmadığı kontrol edilir. Araç turizmde kullanılıyor ise, araç sahibi hizmet yenilemek icin Turizm Bakanlığı'na hizmet yenileme dilekçesi ile başvurur. Araç turizmde kullanılmıyor ise iş sona erer. Araç 50BG'nden fazla ise araç sahibi en az 60 YTL öder. Daha sonra araç sahibi kayıt yenilemek için Denizcilik Müsteşarlığı'na hizmet yenileme dilekçesi ile başvurur. Daha sonra iş sona erer

İcra Takip İş Akış Diyagramı

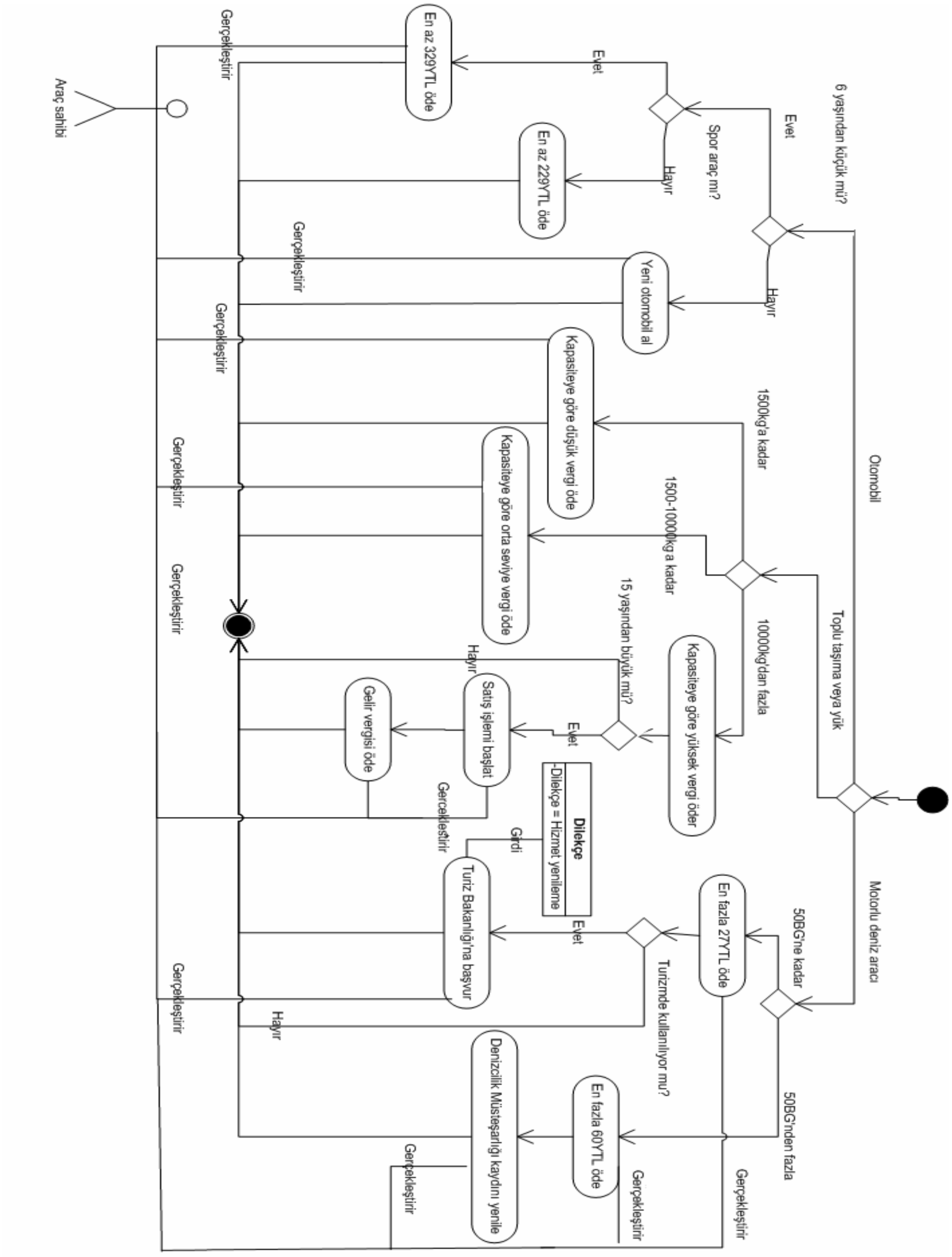
İcra takibi alacaklının borcun cinsine karar vermesi ile başlar. Para ve teminat alacağı değil ise, alacaklı ilamlı icra işlemini başlatır ve icra takip işi sona erer. Para ve teminat alacağı ise, alacaklı ilamsız icra işlemi için başvurur. Alacaklı daha sonra haciz yolunu seçer. Alacaklıda çek, bono, poliçe olup olmadığı kontrol edilir. Bunlar varsa alacaklı KSMHY (Kambiyo Senetlerine Mahsus Haciz Yolu) başlatır ve icra takip işi sona erer.

Alacaklıda çek, bono, poliçe yok ise genel icra için başvurur. İcra dairesi, icra başlatma dilekçesi ile icra takip talebini başlatır. İcra dairesi borçluya ödeme emri gönderir. Daha sonra da ödeme süresini başlatır.

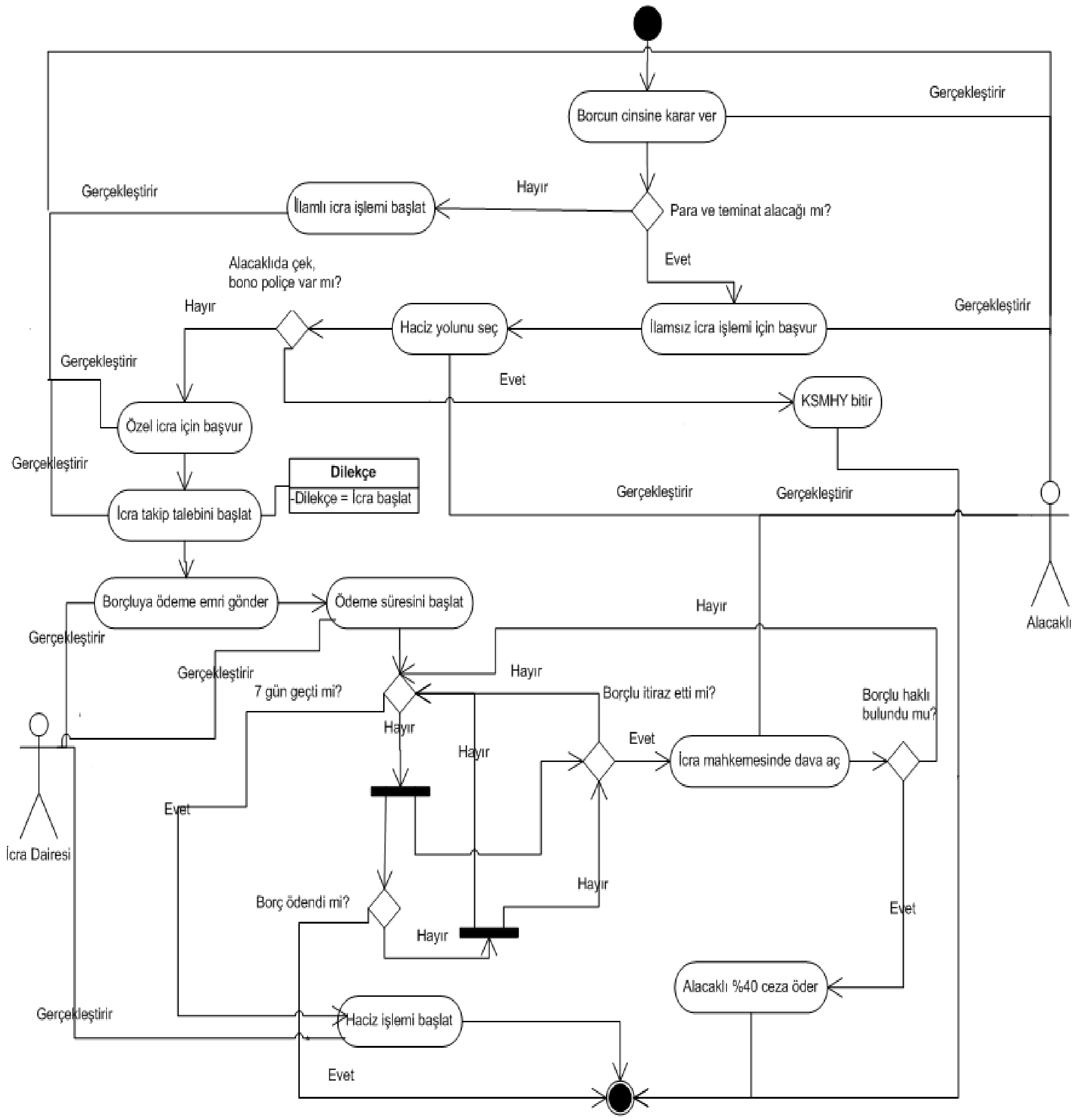
Ödeme süresi başladıktan sonra 7 günün geçip geçmediği kontrol edilir. 7 gün geçti ise icra dairesi haciz işlemi başlatır ve icra takip işlemi sona erer. 7 gün geçmedi ise hem borcun ödenip ödenmediği hem de borçlunun itiraz edip etmediği kontrol edilir. Borç ödendi ise iş sona erer. Borç ödenmediyse hem 7 gün geçip geçmediği hem de borçlunun itiraz edip etmediği kontrol edilir. Borçlu itiraz etmediyse yine 7 gün geçip geçmediği. Borçlu itiraz ederse, alacaklı icra mahkemesinde dava açar. Bu dava sonucunda borçlu haklı bulunmazsa 7 gün sürenin geçip geçmediği kontrolüne geri dönülür. Mahkeme sonucunda borçlu haklı bulunursa icra takip işi sona erer.

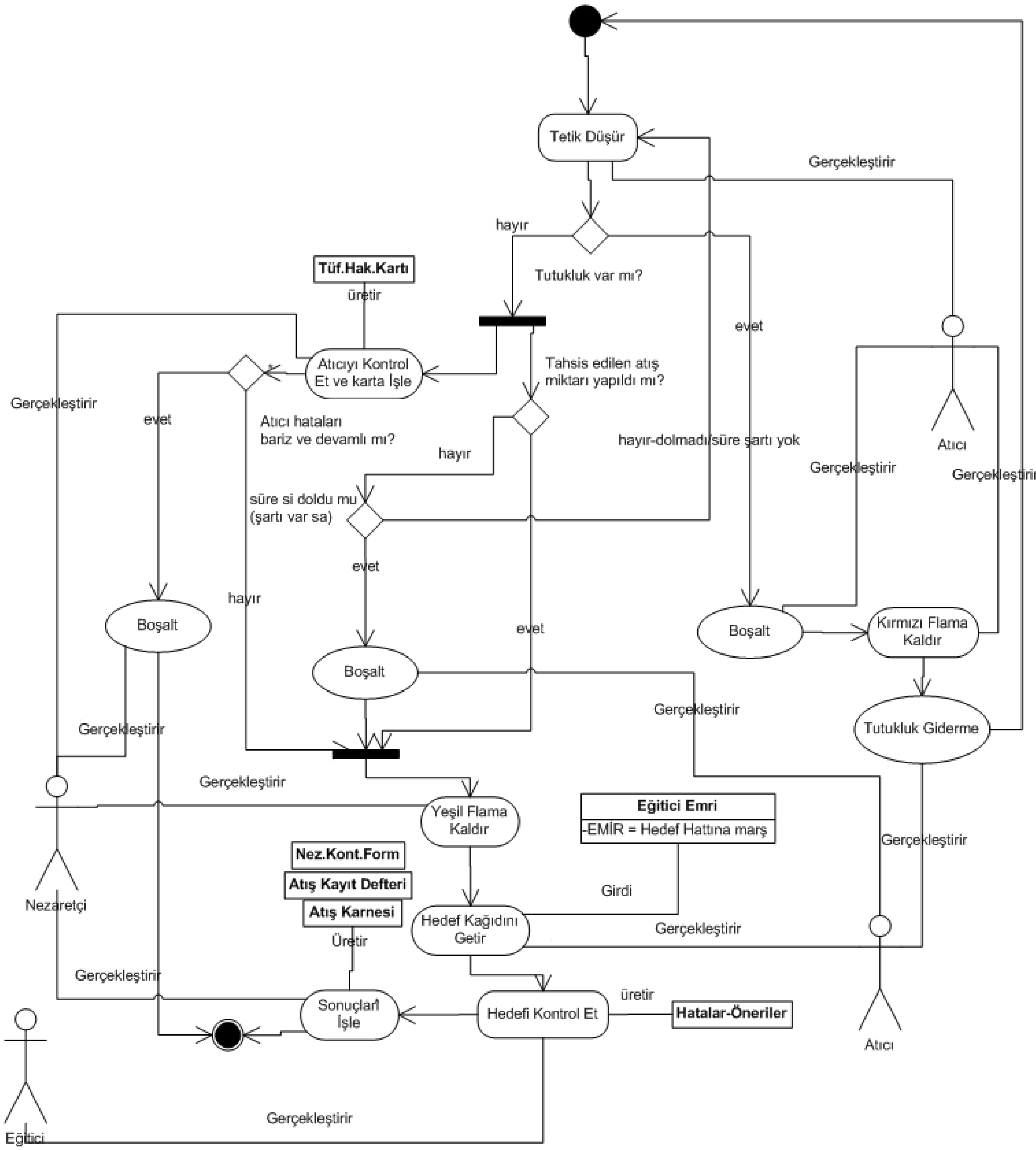
Diagrams of the Follow-up Experiment

The Motor Vehicle Tax Law Activity Diagram



The Debt collection and Bankruptcy Activity Diagram

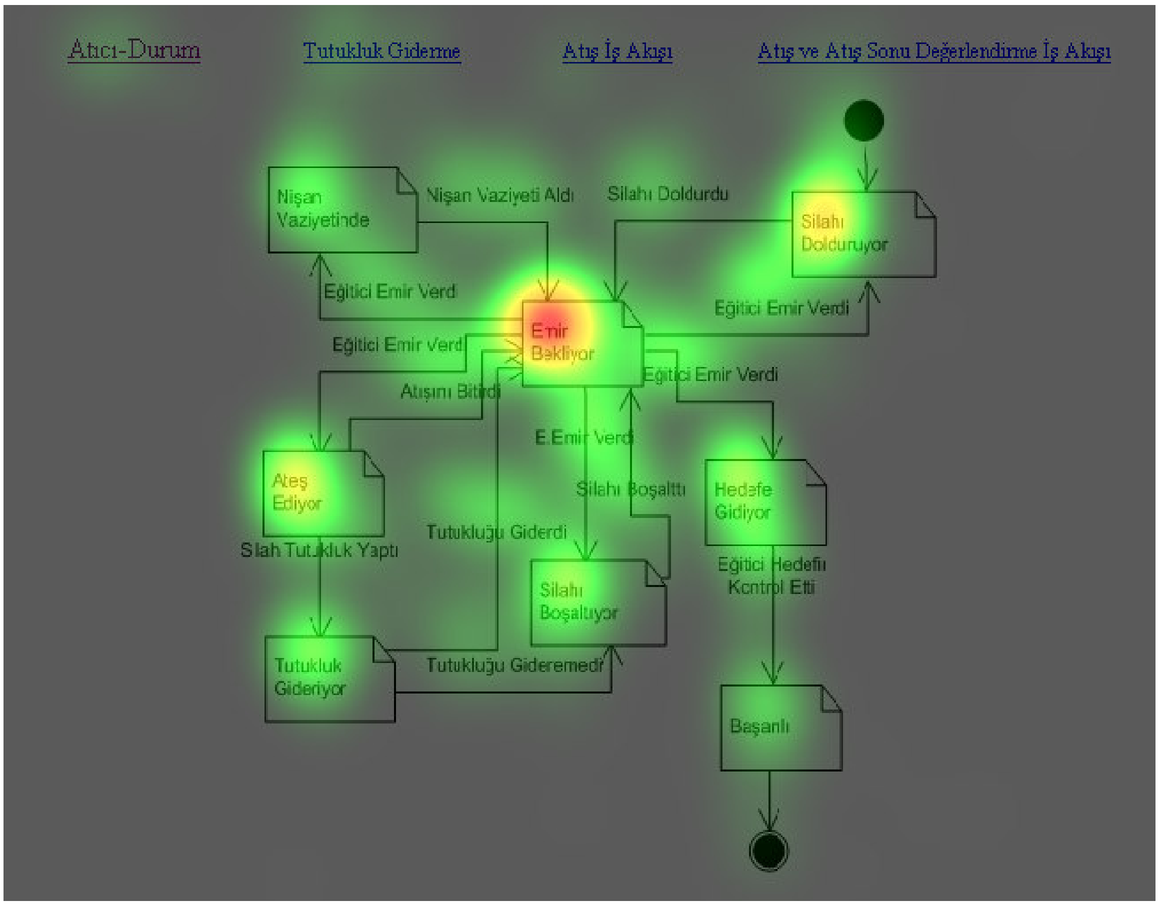




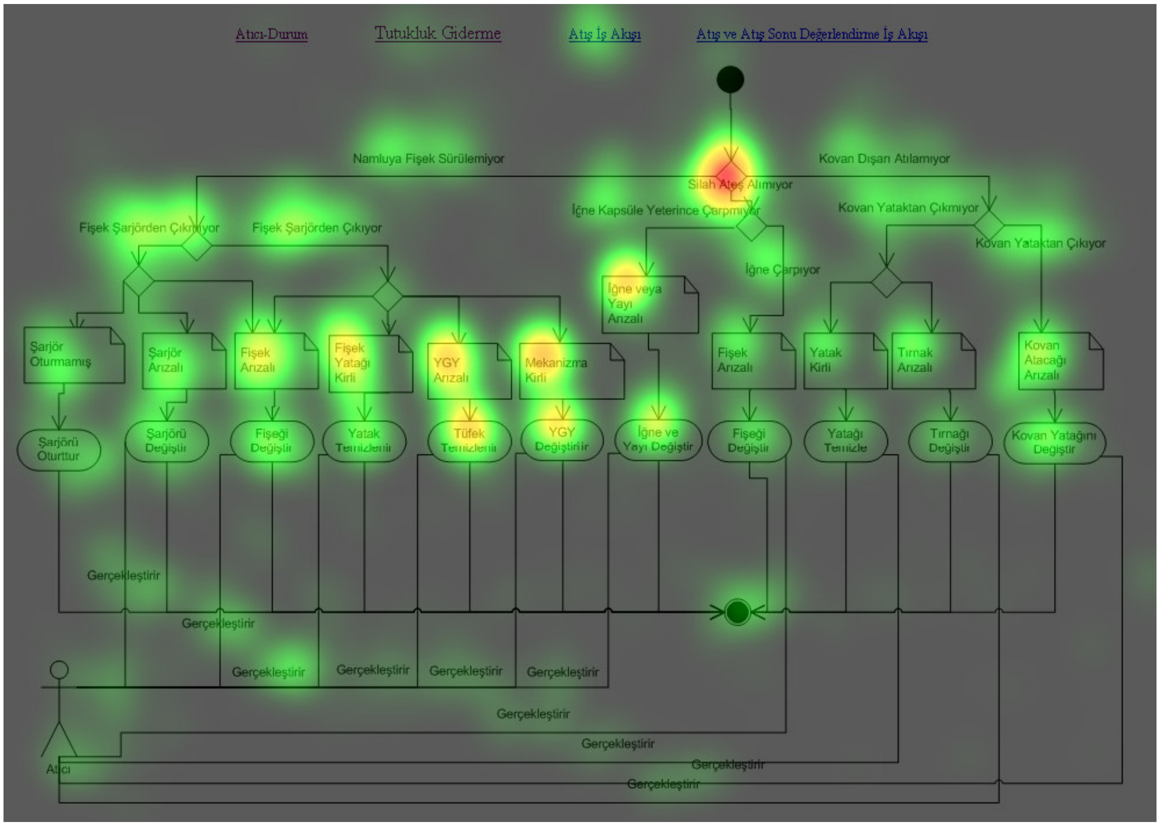
Tetik Düşür
Gerçekleştirir
hayır
Tutukluk var mı?
Tüf.Hak.Kartı
üretir
evet
Atıcı
Tahsis edilen atış miktarı yapıldı mı?
Atıcıyı Kontrol Et ve karta İşle
Gerçekleştirir
evet
Atıcı hataları bariz ve devamlı mı?
hayır
hayır-dolmadı/süre şartı yok
Gerçekleştirir
Gerçekleştirir
süre si doldu mu (şartı var sa)
evet
hayır
Boşalt
evet
Boşalt
Kırmızı Flama Kaldır
Boşalt
Gerçekleştirir
Tutukluk Giderme
Gerçekleştirir
Gerçekleştirir
Yeşil Flama Kaldır
Eğitici Emri
-EMİR = Hedef Hattına marş
Nez.Kont.Form
Atış Kayıt Defteri
Gerçekleştirir
Nezaretçi
Girdi
Atış Karnesi
Hedef Kağıdını Getir
Gerçekleştirir
Üretir
Gerçekleştirir
Sonuçları İşle
Hedefi Kontrol Et
üretir
Hatalar-Öneriler
Atıcı
Eğitici
Gerçekleştirir

APPENDIX D. Hot Spots of Eye Movement Data

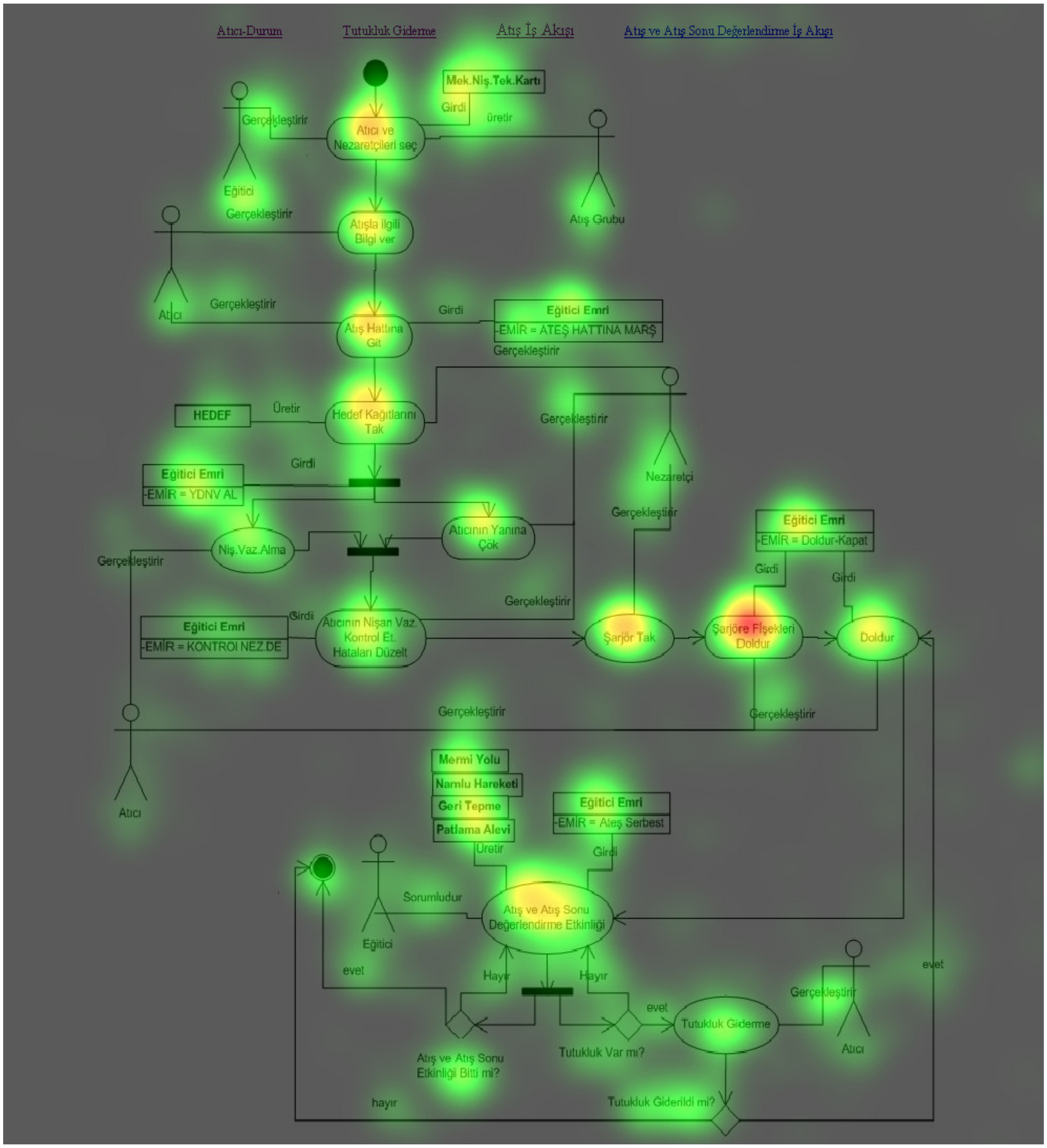
Hot Spot of the Marksman State Diagram for Domain Familiar Group



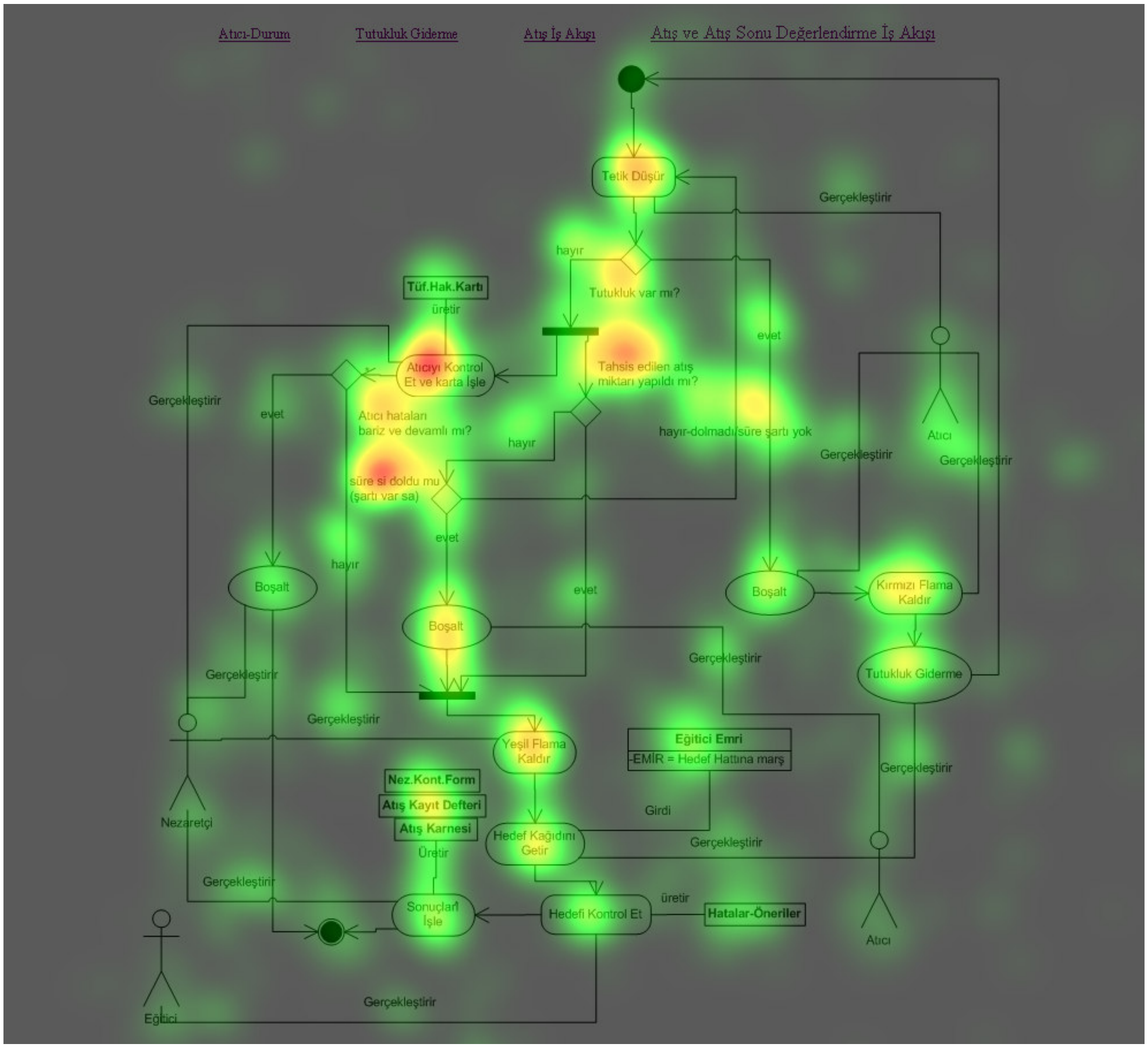
Hot Spot of the Stoppage Removal Activity Diagram for Domain Familiar Group



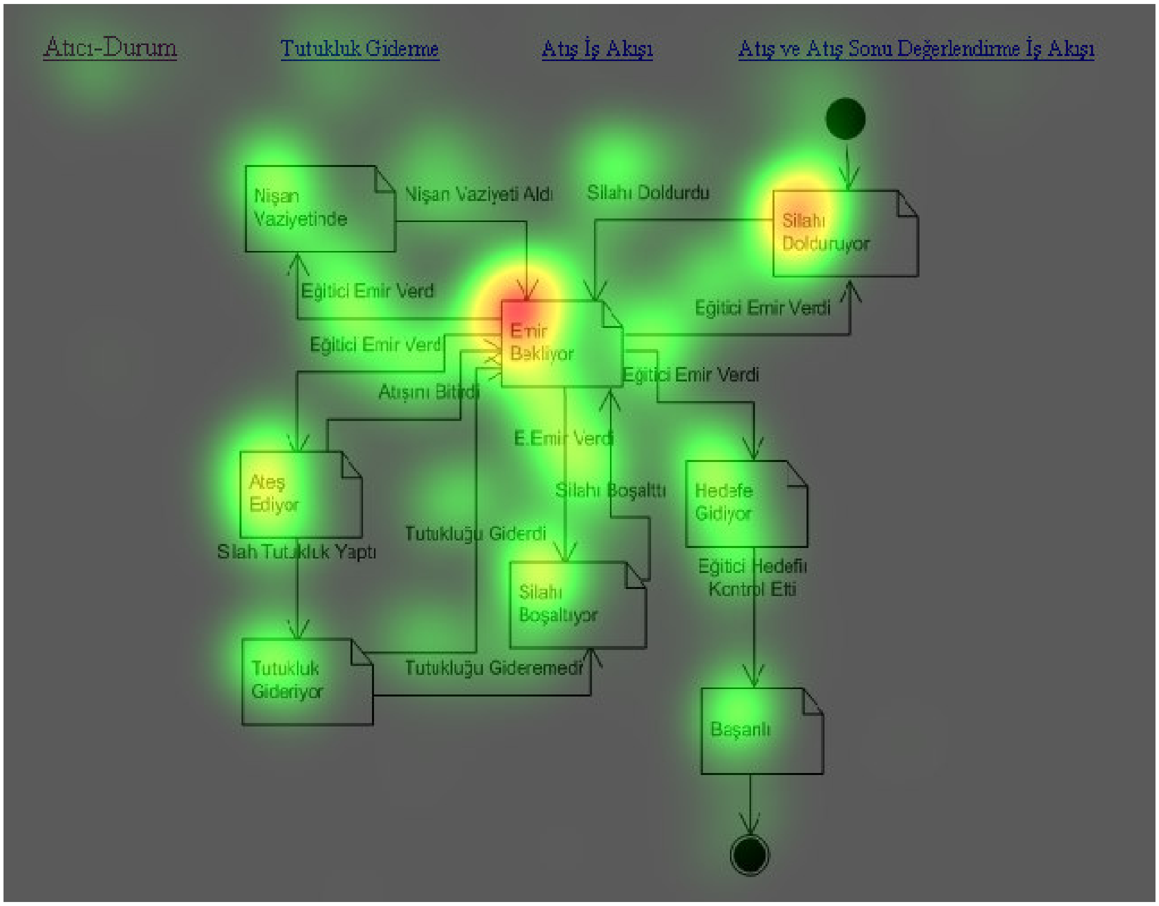
Hot Spot of the Marksman Mission Training Activity Diagram for Domain Familiar Group



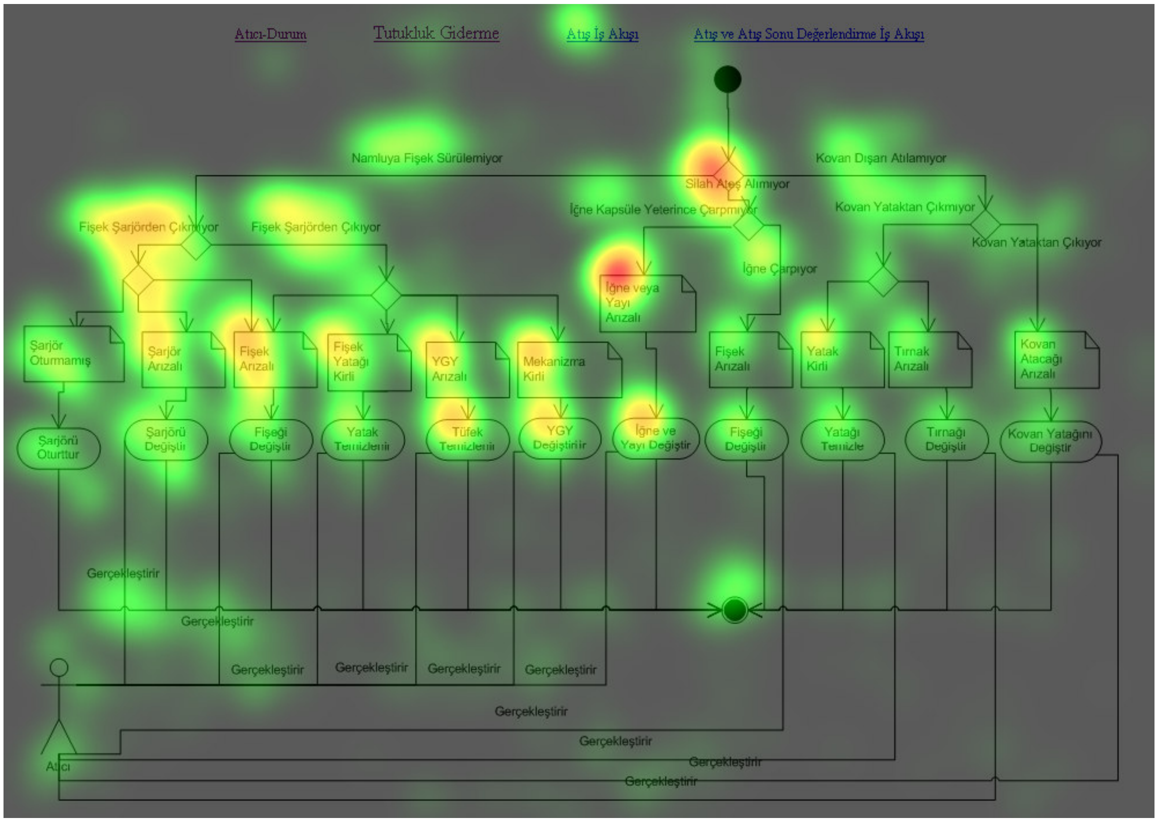
Hot Spot of the Marksman Mission Ending and Evaluating Activity Diagram for Domain Familiar Group



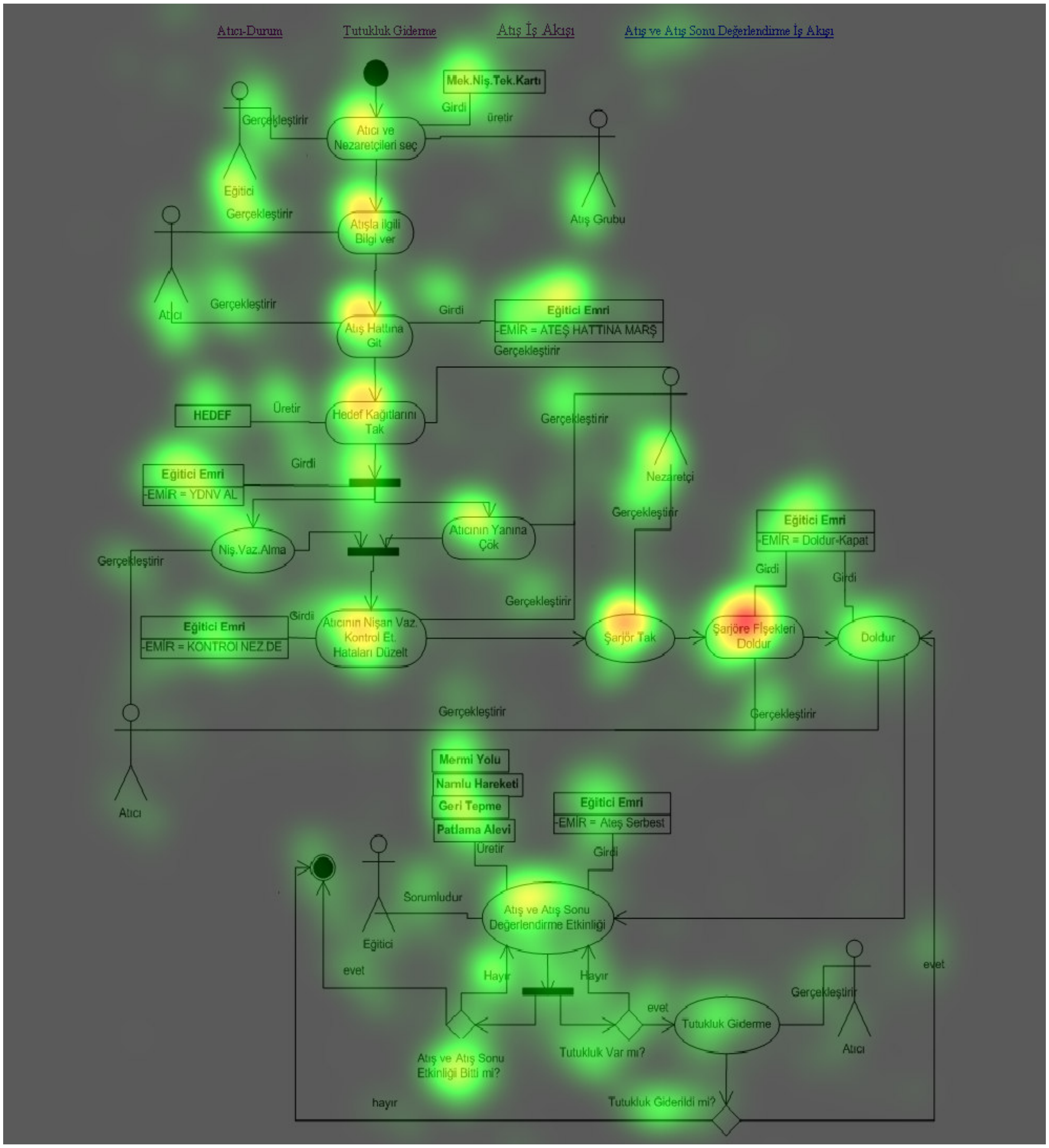
Hot Spot of the Marksman State Diagram for Notation-Familiar Group



Hot Spot of the Stoppage Removal Activity Diagram for Notation-Familiar Group



Hot Spot of the Marksman Mission Training Activity Diagram for Notation-Familiar Group



Hot Spot of the Marksman Mission Ending and Evaluating Activity Diagram for Notation-Familiar Group