UPPSALA
UNIVERSITET

15043
Examensarbete 30 hp
Juni 2015

Investigating the Effects of Trends in an Interface to a Dynamic System

Sercan Caglarca

Institutionen för informationsteknologi
Department of Information Technology

Page intentionally left blank

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Investigating the Effects of Trends in an Interface to a Dynamic System

Sercan Caglarca

Uppsala University and Trafikverket (The Swedish Transport Administration) have been in collaboration in order to improve the train traffic control systems in Sweden for many years. As a result, a train traffic control system STEG (Swedish for ‘Control via an Electronic Graph') was built, and evaluated. Based on the evaluation results, need for more constrained experiments have been revealed.

The use of microworlds in such dynamic decision making research is a common approach. For that reason, a microworld (train traffic simulator) was built in Uppsala University. The purpose of the designed experiment was to explore the effects of absence or presence of trend lines on performance and perceived difficulty in an interface of a dynamic system for novice users. The study also answered whether instead of a generic goal, introduction of a target to the users affected their behavior. In the experiment, 32 participants, interacting with the microworld, tried to solve a logical problem and were given 40 trials to improve their performances. In order to test main and interaction effects between the proposed variables (performance, perceived difficulty), the experiment was based on a 2 x 2 factorial design (trend lines: present/absent, target: present/absent).

The results were analyzed by means of a mixed design ANOVA for repeated measures. In addition, Scheffé post-hoc analysis and regression analysis were conducted. The analysis results have shown that the trend lines did not improve performance and slowed down learning. The users who were subjected to trend lines and were introduced to a target perceived the task significantly harder.

Handledare: Mats Lind
Ämnesgranskare: Anders Jansson
Examinator: Lars Oestreicher
15043
Tryckt av: Reprocentralen ITC

Page intentionally left blank

Acknowledgements

I would like to thank the people who helped me with this thesis, along with the people who I worked with during the days I have been a project assistant in Uppsala University. It is simply because what I learned during those days constitutes the fundamentals of this research.

First, I would like to thank my supervisor Mats Lind not only for his supervision, but also for offering this thesis project to me in the first place. He is the source of profound wisdom and experience, and has been my role model both academically and professionally. I would like to thank my reviewer Anders Jansson for his discerning and invaluable supervision. His intellect and constructive attitude have played a key role. Without them, this thesis would never exist.

I would also like to thank Anton Axelsson for all his efforts to take his time to help me finish this thesis. He mentored, supported and encouraged me anytime I needed. I am very happy that you have been a part of this journey.

I would also like to thank Bengt Sandblad to whom I will eternally be grateful; not only for all his supervision but also for all the support he has shown during my time in Uppsala. Thanks to you for everything you taught me during the days I worked as a project assistant in the European research project ON-TIME (FP7-SCP0-GA-2011-265647). I learned a lot from you. I would like to mention Arne Andersson, who has also been very supportive during this period. Working with you has been enlightening. I would also like to thank Simon Tschirner for being an awesome colleague and a helpful, caring friend. I appreciate all your valuable feedback on the thesis.

I have to mention, Mikael Laaksoharju, who has been extremely helpful since the first day I came to Uppsala. Being an inspiring teacher, and a very good friend I really don’t know how to thank you enough.

Apart from my teachers and colleagues, I have had inspiring discussions with many of my classmates. Thanks to you all!

Page intentionally left blank

Sammanfattning

Uppsala universitet och Trafikverket (den del som tidigare hette Banverket) har under flera år samarbetat med målet att förbättra kontrollsystemen för den operativa tågtrafikstyrningen. Som ett resultat av det samarbetet utformades och byggdes STEG (Styrning av Tåg genom Elektronisk Graf). Arbetet med att utveckla STEG hade tre olika syften: att förbättra den kognitiva arbetsmiljön, att skapa ett verksamhetseffektivt arbetsredskap och på så sätt bidra till högre kapacitetsutnyttjande genom bättre och snabbare beslut i trafikplaneringen, samt att stimulera lärande och underlätta inlärning vid nyanställning och träning.

STEG utformades med hjälp av en användarcentrerad designprocess. Expertanvändare från den operativa trafikstyrningen deltog, under ledning av och i samarbete med forskare från Uppsala universitet, i både analys, design och successiv utvärdering av olika prototyper av STEG. STEG har hittills använts i skarp drift vid trafikövervakningen i Norrköping, och används idag på samma sätt i Boden. De preliminära analyserna från både Norrköping och Boden visar att STEG som designkoncept är mycket uppskattat och har stor potential ur ett verksamhetsperspektiv. När STEG används som det är tänkt, och implementeringen fungerar tekniskt, vill personalen inte återgå till tidigare arbetssätt. Det ligger därför nära till hands att dra slutsatsen att STEG uppfyller alla tre syften ovan, det vill säga STEG bidrar till en kognitivt enklare arbetsuppgift för trafikplanerarna, ett mer verksamhetseffektivt arbetsredskap och att det skapar en lägre inlärningströskel.

Exakt varför STEG är uppskattat, vilka förklaringar det finns till dess upplevda värde, har dock inte varit möjligt att studera tidigare. Någon systematisk och utförlig utvärdering av STEG har inte gjorts – det ligger i den användarcentrerade systemdesignens natur att det inte görs någon experimentell eller systematisk utvärdering av framtagna designkoncept. Ur ett långsiktigt verksamhetsperspektiv för Trafikverket, och ur ett vetenskapligt mer kontrollerat perspektiv, är det dock mycket intressant att klargöra varför STEG upplevs som enkelt, effektivt eller bättre. Tre alternativa hypoteser har identifierats i de preliminära analyserna: (1) att STEG grafiskt återger en relevant beskrivning av trafikplanerarens arbetsdomän och att hen därför enklare kan associera pågående aktiviteter med den semi-dynamiska representation som finns i STEG; (2) att den direktinteraktion med omedelbar återkoppling som finns i STEG medger ett feedback-baserat arbetssätt, vilket ur ett kognitivt belastningsperspektiv är att föredra framför ständig framförhållning (feed-forward); eller (3) att informationen som visas semi-dynamiskt i STEG gör det enklare att se vad som pågår, istället för att trafikplaneraren med hjälp av arbetsminnet ska behöva lägga ihop information från olika system för att skapa sig en helhetsbild av olika skeenden. I den aktuella studien är det den tredje och sista av dessa tre hypoteser

som ska undersökas experimentellt. Det som specifikt studeras är huruvida perceptuella beslutstöd, prognoser i form av visuella trendlinjer, leder till snabbare inlärning och bättre beslut än om sådana visuella prognoser saknas. Även effekten av specifikt målkriterium undersöks.

Tågtrafikplanerarens beslutsfattande och problemlösning kan karakteriseras som dynamiskt beslutsfattande. Inom dynamiskt beslutsfattande studeras beslutsprocesser med hjälp av mikroväldar. Med STEG som referenssystem byggdes därför en sådan mikrovärld, GridRail. Syftet med experimentets var att undersöka effekterna av närvaro respektive frånvaro av visuella trender på prestation och upplevd svårighetsgrad. För att jämföra objektiv prestation med subjektiv upplevelse ombads personerna som deltog i studien att skatta hur svår uppgiften var vid tre tillfällen.

Den aktuella studien är den första i en tänkt serie av experiment med den nya mikrovärlden GridRail. Tanken är att Grid Rail successivt ska byggas ut för att bli alltmer komplex och därmed i högre utsträckning än vad som nu är möjligt representera de arbetsuppgifter som finns i den operativa tågtrafikstyrningen. För den aktuella studien gjordes därför bedömningen att studenter kunde användas för att studera effekten av visuella trendlinjer. 32 deltagare interagerade med GridRail i en beslutsuppgift som har likheter med både klassiska problemlösningsuppgifter, dynamiska beslutsproblem, och arbetsminnesuppgifter. De fick 40 försök på sig för att förbättra sin prestation. För att testa såväl huvud- som interaktionseffekter grundade sig experimentet på en 2x2 faktoriell design, och resultaten analyserades med hjälp av en ANOVA för upprepade mätningar inom varje betingelse och med två mellanpersonsvariabler (trendlinjer och mål). Effekterna mättes som prestation (inlärningskurvor) och upplevd svårighet (subjektiva skattningar). Skattningarna av upplevd svårighet genomförde vid tre tillfällen under experimentet. Avslutningsvis gjordes en intervju med deltagarna.

Resultaten av studien visar att det inte fanns någon huvudeffekt av mål, däremot fanns en signifikant huvudeffekt av trendlinjer, men i strid med hypotesen om en förväntad positiv effekt av dessa – försökspersonerna utan trendlinjer presterade bättre! Den starkaste effekten utgjordes dock av en interaktionseffekt mellan mål och trendlinjer, där kombinationen trendlinjer och specifika mål utgjorde den betingelse där försökspersonerna fick den klart sämsta prestationen. Denna betingelse var också den där inlärningen gick långsammast sett över de 40 försöken. Intressant är också att konstatera att det är betingelsen med trendlinjer och specifika mål som upplevs som den signifikant svåraste.

Slutsatsen från studien är att det inte gick att påvisa några signifikanta effekter av varken mål eller trendlinjer, åtminstone inte i riktning med den inledande hypotesen. Istället verkar kombinationen av specifika mål och trendlinjer utgöra den svåraste betingelsen, både vad gäller prestation och upplevelse. Som konstaterades ovan är detta den allra första studien med GridRail, och vi kan därför inte dra några säkra slutsatser alls. Det faktum att mål och trendlinjer tillsammans skapar en uppgift som det tar längre tid att lära sig, och att samma betingelse dessutom upplevs som svårast indikerar möjligen att vi har skapat en komplexare och mer realistisk uppgift än vi hade tänkt oss. Fortsatta studier kommer att behövas för att utreda detta.

Page intentionally left blank

Contents

List of Figures

List of Tables

Chapter 1 Introduction

Humans' ability to solve problems greatly surpasses that of any other species, and thanks to the evolution of this ability we succeeded to survive and dominate for thousands of years. Nevertheless, we also created civilizations that constantly generate other novel problems for us to solve. We often find ourselves in situations in which we need to solve a problem. Imagine that you are at a job interview and asked to assemble *IKEA* furniture without being given the assembly instructions. Let us say that what you see on the ground are parts of a complex bookshelf. How would you start? You would, of course, immediately start *thinking*. As the famous philosopher Aristotle might say: we are rational animals. We will process any perceived information, make sense of things and apply logic to solve problems. Most of us take this further and keep believing that we are remarkably intelligent beings. However, this phenomenon could be approached from a different perspective; having cognitive capabilities, yes, we do reason, and yet, we often make mistakes. This is simply because our cognitive capacity is limited. As most cognitive scientists today would say, when we make decisions we are cognitively limited, and unfortunately most of the time we are highly biased (Kahneman, 2011). As prone as we are to intelligence and insight, we are equally liable to irrationality and false intuition. Essentially our cognitive skills are the main factors that determine who can assemble the bookshelf and who cannot in that particular instant.

However, we are quickly passing through the historical moment when people are confined only to their own cognition as they make decisions. With today's emerging technology, when we need to overcome a cognitively demanding task, especially in our work environments, we use computerized systems. Our environments are enriched with new possibilities of supporting our cognitively demanding tasks, e.g. through networked computers, ubiquitous systems or interactive devices. These digital artefacts thus enhance our ability to draw more correct conclusions from perceptual inferences (Hutchins, 2000). As a matter of fact, this was one of the core insights that became a reason for the system STEG, which was developed in Sweden to be used in train traffic control centers and became the inspiration of this thesis work.

For about twenty years Uppsala University and Trafikverket have been working together on research projects in order to create systems for train traffic control. After an analysis of the work of traffic controllers in Sweden (Andersson et al., 1997), a need for better control strategies has been identified for traffic controllers (Sandblad et al., 1997; Kauppi et al., 2006). Based on the ongoing research, a new operational traffic control system, called STEG, was developed (Sandblad et al., 2010). The system was deployed and tested in two different traffic control centers in Sweden with the support of the Swedish railway authority (Andersson et al., 2014).

Evaluations have shown that the new system contributed to an improved support to the dispatchers and a better planning of train traffic (Sandblad et al., 2007). Thereby, the system led to a radically improved performance (Sandblad et al., 2010). STEG supposedly reduces the unnecessary cognitive load by supporting train traffic controllers' mental models and increases the level of situational awareness among the users. Based on recent theoretical progress in the area of cognitive psychology, human-computer interaction and modern literature on problem solving and dynamic decision making, we would like to further investigate the reasons behind this improvement. For that reason, our research group has embarked upon a set of experiments to be conducted in the near future, and as a first step in that direction, our research group started designing a simulation game, called *GridRail*, which will serve as a microworld (see Chapter 4) to be used in our experiments and be further developed as we progress and find answers to the potential questions to be investigated through these studies. Eventually, this thesis project has been conducted as the very first of these studies.

STEG is a dynamic system with a complex user interface and it has many features which need to be tested. As a consequence, the experimental process we are proposing here is to use GridRail as a tool to assist us to test the effects of the major features STEG currently has in its interface. The findings of the first few experiments to be conducted, including this thesis project, are not supposed to be fully generalizable to STEG, but instead we are expecting them to reveal more general findings about how cognition works as people interact with a dynamic system. In addition, these first experiments have an extra role, and that is to assist us to further improve this game simulation and understand how we should design our future experiments.

In this first project, what we particularly would like to focus on is the elements being graphically presented in the interface. We believe that presentation of trend lines (see Chapter 4) is decreasing the cognitive load of the users in general, but what is more intriguing is to understand how things work in the minds of novice users who are also expending energy on learning. Despite the complexity of STEG interface, experienced train traffic controllers can immediately perceive any event, interpret and take further actions. However, inexperienced or untrained users would be overwhelmed by the number of available options offered in such complex systems (Tschirner, 2015). Consequently, in this thesis, the question of how the performance and learning of novice users are affected by the graphically presented predictions in the interface is investigated.

In the following sections I will introduce the reader to some of the basic concepts and the main aspects of train traffic control in Sweden, and give details about the decision support system STEG that is currently planned to be deployed at the train traffic centers located all around Sweden.

1.1 Train Traffic Control

Railway systems all around the world are controlled based on principles from past decades (Tschirner, 2015). When I worked in the On-Time project, I had the chance to analyze the differences in train traffic control processes throughout Europe and learned how differently it was organized in several different European countries, implying that historically the train traffic control organizations have matured quite diversely in different locations (Golightly et al., 2013). The main reason for this difference is grounded on the availability of the technology at different times and in different countries during the construction and upgrade of the infrastructures. This difference in railway systems in different countries inherently affects the train traffic organizations. In this thesis, the focus is on the Swedish organization.

The organization which is responsible for planning and controlling the road, air, sea and railway traffic around Sweden is the Swedish Transport Administration, Trafikverket. For railway traffic in particular, their responsibilities include train traffic control and its maintenance (Von Geijer, 2014). There are two distinct, unique features of the Swedish organization in terms of train traffic control processes; these are its centralization, and the role of the train traffic controller (Tschirner, 2015). After giving some details about the former, the latter will be clarified.

As shown in Figure 1, there are 8 train traffic control centers located in different parts of Sweden operating in their specific regions. In each of the 8 regional centers, the traffic is controlled as several isolated traffic segments (Sandblad et al., 2010).

Figure 1 Map of Sweden, indicating the eight train traffic control centers in Sweden

At the end of 2013, by introducing a set of extra regional and national control layers, Trafikverket restructured their train traffic control processes in order to achieve a better coordination of traffic control across the borders of different control areas and to provide a better communication between the peers. In each control center, there are a number of traffic controllers and at least one head controller, who is also in contact with other centers and is assigned to organize the collaboration of traffic controllers inside the train traffic control center (Tschirner, 2015).

Figure 2 The workplace of a train traffic controller at the train traffic control center in Stockholm.

Figure 2 exhibits the new appearance of the train traffic centers in Sweden after the redesign in 2013. A typical workspace of a train traffic controller, as can be seen, consists of regular computer screens, large wall panels, paper graphs and telephones. The computers nearby give access to the different control systems, while the large distant panels, located slightly in the background, show the track diagram, and display the blocks that are blocked by or reserved for certain trains. Paper based time-distance graphs placed on the desks are necessary in order to follow the daily traffic plan, and telephones with blue-tooth headsets are used for communication purposes such as informing train drivers or reporting anomalies (Tschirner, 2015).

It is a complex and dynamic work environment due to the high number of people who are affected, their communication and collaboration, as well as different support systems being interacted with by the controllers and the continuous development of the ongoing traffic. In addition, there are internal and external incidents, such as disruptions and disturbances on the railway tracks or the trains. These disruptions, varying from delayed departures from stations to infrastructure failure at busy junctions, could be small or large with consequences ranging from smaller delays to re-routing or the cancellation of scheduled trains. Moreover, it is known that even a short cumulative delay especially for freight trains on the Iron Ore Line, causes a loss of millions of kronor, forcing the train traffic controllers to act in a very short period of time and consequently generating a high level of stress within the work hours (Tschirner, 2015).

Additionally, in Sweden traffic planning and train signaling are integrated in one single role and it takes many years to become an expert train traffic controller. Unlike in most countries where the roles of dispatchers and signalers are discrete and performed by different individuals, in Sweden the train traffic controller works both as a signaler, who executes the plan and controls train paths and signals, and as a dispatcher who monitors the train movements and reschedules the current traffic plan with respect to perturbations and disruptions (Tschirner, 2015). This type of action, that is to only intervene when conflicts or disturbances occur, is called *control by exception* (Andersson et al., 1997). Acting only when a perturbation occurs obviously has many disadvantages. So the idea of changing this approach led those responsible to come up with a different control system. Since STEG's design is based on a real-time traffic plan, it is claimed that the developers could manage to change the control paradigm from *control by exception* to *control by re-planning* (Kauppi et al., 2006).

1.1.1 The Paper Graph

The paper graph (see Figure 3) that is being used by train traffic controllers is a printed time-distance graph reflecting the daily traffic plan with information about all the scheduled trains, their routes and the stations they are planned to stop at. The train traffic controllers have to check this paper during the whole shift in order to complete their tasks e.g. solving conflicts and simultaneously re-planning the traffic. The paper graph presents the routes of all the planned trains within the region or their arrival and departure times and the distances between stations. These are the kinds of information that the train traffic controllers cannot directly get from the systems they are interacting with. The paper graph helps them to receive such information (Tschirner, 2015).

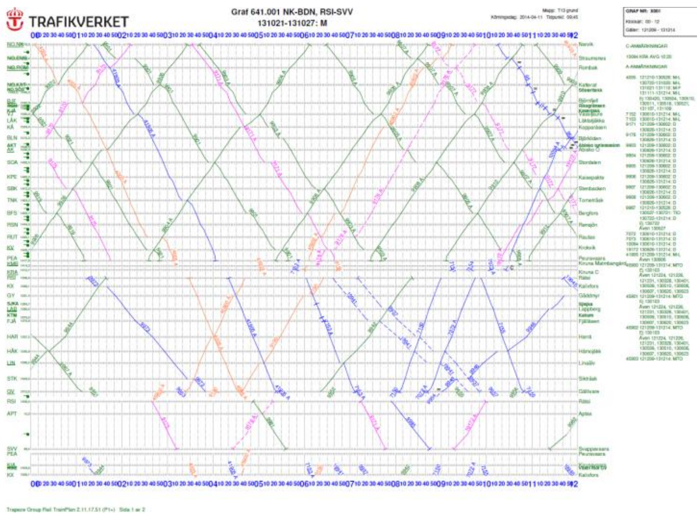


Figure 3 A typical paper graph used in train traffic control centers

The train traffic controllers also have to note things down on these papers. In other words, during their shifts, they use a pen to draw the changes on the daily traffic plan in order to solve and record their solutions to upcoming conflicts and delays in traffic. Undoubtedly, this method has a lot of disadvantages. For example, re-planning and accurate noting of a train's trajectory requires numerous redrawing. Since the data is drawn on these papers, it is also not possible to be shared quickly in digital platforms and instead all these changes have to be communicated via telephone (Tschirner, 2015). This can be considered as an outdated practice. Moreover, sometimes the shifts can be busy and require the traffic controllers to spend all their time on the phone. Recording an infrastructure failure, approving shunting or maintenance works could be potential reasons for such time consuming conversations. In such situations, the train traffic controllers might not have sufficient time to communicate noncritical information. Indeed, most of the changes in the plan are noncritical and they are not communicated due to this reason (Tschirner, 2015).

1.2 STEG: a Tool for Train Traffic Controllers

It was understood that to improve the process of controlling train traffic, the control paradigm had to be changed from low-level technical control tasks into higher-level traffic re-planning tasks, so that the train traffic controllers can spend most of their time thinking and testing how to re-plan a dynamically evolving time-plan.

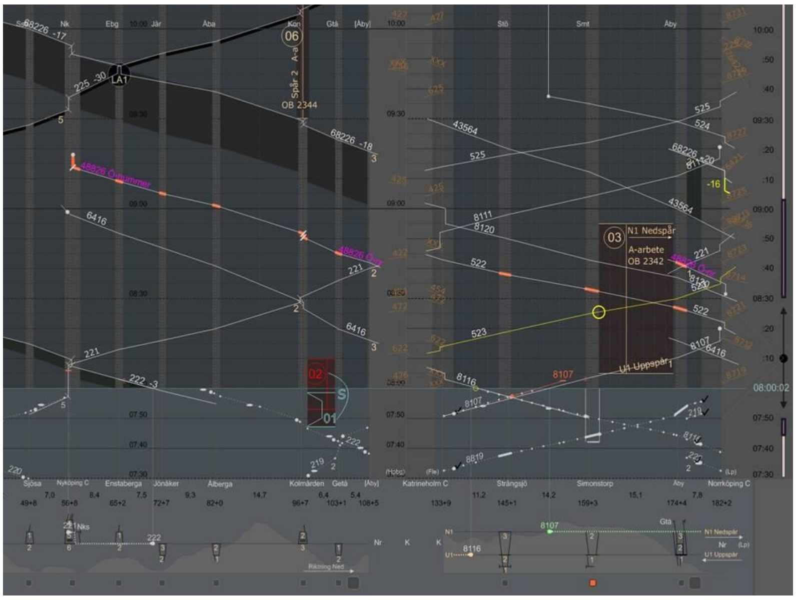


Figure 4 The STEG System User Interface - http://www.it.uu.se/research/project/ftts/steg

As a result, based on the ongoing research, a new operational traffic control system, called STEG was developed (Sandblad et al., 2010). STEG is used today in two train traffic control centers; Norrköping and Boden.

STEG was mainly designed to provide efficient user interfaces and better decision support in order to give the train traffic controllers the opportunity to be continuously updated and be able to examine the traffic. It is designed to support the users so that, by taking further actions, they can solve future potential traffic conflicts in advance, and re-plan the traffic situation whenever needed (Kauppi et al., 2006). For that reason, the developers of STEG employed a UCSD (User Centered Systems Design) approach which was defined and discussed by many researchers such as Norman & Draper (1986), and Karat (1997).

Figure 4 shows the user interface of STEG. When the main view in the interface covering most of the screen area was being developed, to be able to introduce the users with a familiar design, the developers were inspired by the paper graph (a.k.a. time-distance graph) that was already being used by the train traffic controllers in order to complete their duties. The x-dimension representing the distance and the y-dimension representing the time, the traffic controllers can continuously observe the dynamic development of the traffic. The current timeline is indicated by a horizontal line. The main view, showing both the history and the future of the trains, automatically scrolls downwards as time evolves (Sandblad et al., 2010). The time scale is adjustable and the user is able to scroll back and forth in time, e.g. the user can compare the current plan with situations that occurred before. It is also possible to see the other plans belonging to other traffic controllers who perform in the adjacent areas. In this main view, the lines represent the train routes and by clicking on or dragging them via mouse, traffic controllers can change the trajectories of the trains in the plan. Using the scroll wheel, a trajectory can be put forward or backward in order to reschedule a train's stop for an earlier time, or for instance to postpone one of its meetings. The track usage at a station can be configured or additional stops can be added to a train's route. The train trajectories are drawn on a time-distance graph and their slopes indicate the speeds of the trains. The interface thus allows the users to adjust the speeds of the trains by changing the slopes of their trajectories (Tschirner, 2015).

As the users spend time working on the plan, the system identifies conflicts with respect to track usage on the train lines or in the stations and automatically indicates them in the interface. The interface also visualizes the results of all re-planning actions and the effects of the valid traffic plans (Sandblad et al., 2010).

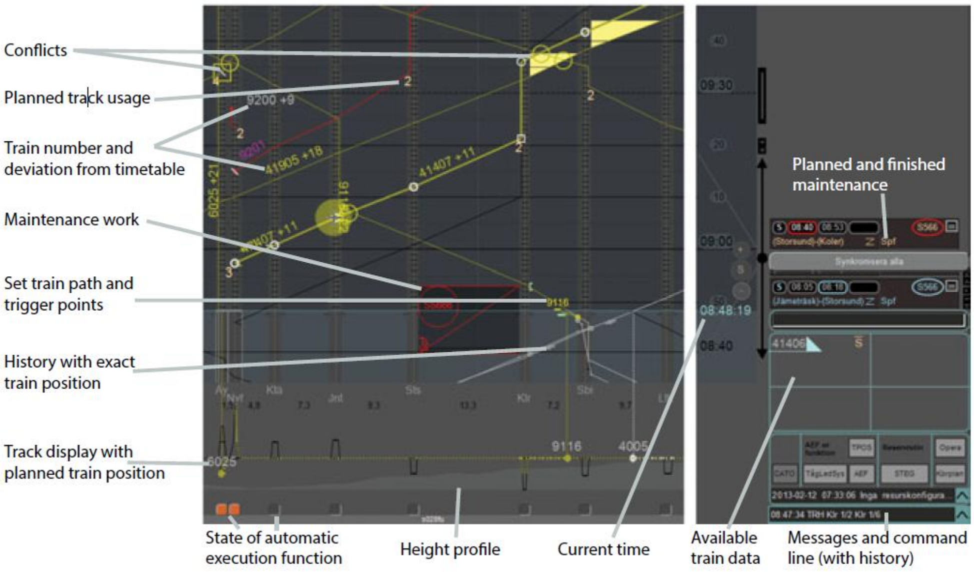


Figure 5 STEG's user interface, close-up (adopted from Tschirner, 2015)

For this thesis, most functions and features in the interface are out of scope, but to give an idea on how STEG interface works, some of the basic elements are briefly described. Figure 5 is a close-up view from STEG's user interface. Given the descriptions of different elements in the figure, here it shows a part of the interface including the future and the history of the train routes, track structure, train and station information and planned maintenance work (Tschirner, 2015). For example, in order to direct the operators' attention to what is important, track or line conflicts are visualized with a high contrast to the background as yellow shapes or frames. Also, the orange boxes seen at the bottom of the screen represent whether the automation function is enabled or disabled. The user can also see the track usage at a planned stop, such that it is indicated via numbers over the stops in the interface (Tschirner, 2015).

1.2.1 Evaluation Results of STEG

After evaluations with case studies which were conducted using a functioinal prototype (Kauppi et al., 2006), STEG was developed, deployed and tested at two traffic control centers in Sweden; Norrköping (center 1) and Boden (center 2) respectively (see Figure 6). The first evaluation of STEG performed in the Spring of 2008 in train traffic control center 1 and was conducted through semi-structured interviews, observations on the workplace and questionnaires. Likewise, the evaluation in train traffic control center 2 was performed with semi-structured interviews, but with both non-STEG users and STEG-users. The interviews were conducted before and after the deployment at both centers. (Tschirner, 2015)

The evaluation process and its results were structured according to a model called *GMOC (*an acronym used for *goals*, *models*, *observability* and *controllability)* by researchers who conducted the test in order to formulate and explain their results in relation to the theories. In this thesis, we are basing our studies on their explanations. The GMOC-model will only be shortly mentioned so that the relation between the evaluation results and the explanations made by the researchers is clear. GMOC is closely related to *control theory* and the model describes human work in complex dynamic environments. According to the related literature, for human beings to achieve control over a task and a system, these four elements are considered as necessary prerequisites (Brehmer, 1992).

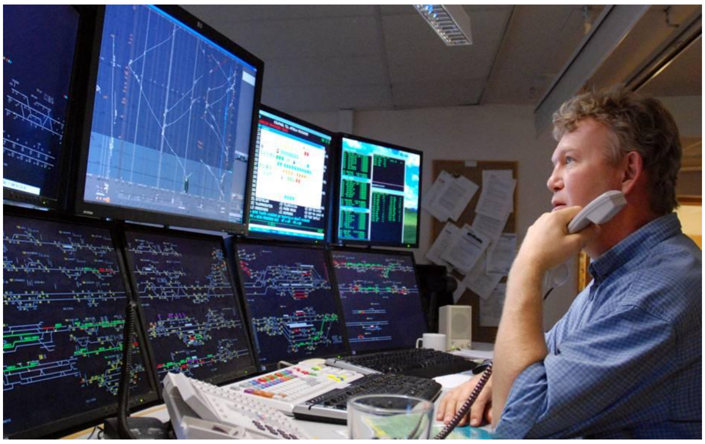


Figure 6 The workplace in Norrköping - http://www.it.uu.se/research/project/ftts/steg

Although some problems were encountered during the deployment of STEG in center 2, the deployment of STEG in center 1 has been successfully completed. In center 1, since the train traffic controllers were so satisfied with the results it was decided that the system would be kept in operation, while in center 2 some problems in the way the new system was understood and used were indicated (Tschirner, 2015). In order to read more about the evaluations, and problems encountered during the work and what might have affected the results, please see the works of Sandblad et al. (2010), Andersson et al. (2014), Tschirner et al. (2014) and Tschirner (2015).

Goal: It is hard for a system to evaluate events based on their critical influence. As a consequence of the evaluations, it has been understood that although STEG supports prioritization of the goal through its interface, it might direct the operators' attention towards parts in the plan which would affect goal achievement in a negative way and mislead the operators (Tschirner, 2015).

Mental Models: According to the evaluations, it has been revealed that STEG interface design supported the users' mental models. However, the use of time-distance graph in the interface had some limitations such as displaying lines with several tracks. As a result, it has been revealed that with the existing design managing larger stations with several platforms and complex layouts would be hard (Tschirner, 2015).

Moreover, except the fact that two traffic controllers expressed their concern that it would be much harder in reality, the case study results have shown that the train traffic controllers thought that it was easy to learn how to operate STEG (Kauppi et al., 2006). One reason behind this is considered to be that since STEG supported users mental models they do not have to change their planning strategies. In addition, the evaluations in center 1 and 2 have shown that since STEG took care of the plan execution in real time, it was considered to be reducing the unnecessary cognitive load on the train traffic controllers and they could focus more on the future plan (Tschirner, 2015).

Observability: According to the researchers the train traffic controllers experienced that STEG gave them a better overview and situation awareness, as well as improving their communication and collaboration. Thus, it is thought that STEG improved observability and controllability. However, through their discussions and observations the researchers concluded that the actors still had deficient observability which led them to construct insufficient models in forms of prejudices about their colleagues (Tschirner, 2015).

Controllability: The results of the case study have shown that traffic controllers felt more in control and able to plan more accurately. It is believed that the main reason for this was that "they were able to see the results of their re-planning decisions, identify conflicts, and observe a train's position and dynamics" (Tschirner, 2015). The new control strategy that came with STEG made it easier for the controllers to handle the traffic process and made them feel more in charge (Tschirner, 2015).

In general, the positive comments from the traffic controllers led the researchers to conclude that "STEG and control by awareness improve the traffic controllers' work environment" (Tschirner, 2015). Despite STEG's lack of efficiency in certain kinds of activities, the traffic controllers evaluated it very positively (Tschirner, 2015) and it improved their performance (Sandblad et al., 2010). The results thus verify what the known HCI researcher Don Norman (1993) says: "Cognitive artifacts are the things that make us smart".

According to the findings of the above evaluations which took place in the real work environments, we can say that STEG, apparently, improved the user experience of train traffic controllers. There is no doubt that STEG is a product of cognitive activity, and it is claimed that cognitive artefacts do seem to amplify human abilities or transform the difficult cognitive tasks into relatively simpler ones (Hutchins, 1990). We believe that these positive findings are mostly based on one tenet of human-computer interaction research field, that is, by looking into the previous research on distributed cognition we see that cognitive artifacts are involved in a process of organizing functional skills into cognitive functional systems, thus they decrease some of the cognitive load the users have to deal with (Hutchins, 2000).

Furthermore, as Hutchins (2000) states in his paper: “While the study of cognition in the wild can answer many kinds of questions about the nature of human cognition in real workplaces, the richness of real-world settings places limits on the power of observational methods. This is where well-motivated experiments come in”. It is apparent that the evaluations in the natural settings tell us a lot about the work environment and the users’ perception and behavior. However, having observed this in the real world environments we can set about designing more constrained experiments which test specific aspects of the systems and their effects on human behavior. Therefore, we believe that these evaluation results raise a number of important questions that can only be resolved by experimental investigation.

Chapter 2 Purpose

2.1 Purpose and Research Questions

The main research question behind this first particular study and all the remaining planned work – including the studies that are supposed to follow – could be narrowed down to one general question we had:

What aspects of STEG improved the user experience of train traffic control?

So in the long run we will try to understand what exact features of STEG made the user experience of train traffic control processes in Sweden improve. In collaboration with Trafikverket, after many years of evaluation and investigation of how to improve the train traffic control in Sweden, the designers of STEG who employed a user centered systems design approach, developed solid design heuristics and had a clear idea of what is unique with it and how it improved the experience. However, the research group who took over would like to conduct studies on STEG today from a different perspective mostly based on human reasoning and decision-making theories.

Based on our previous experiences and existing theories, in order to investigate why STEG worked well and what is good about it, we concluded a number of possible reasons: Is it minimizing the cognitive load by changing a cognitive task to a perceptual task? Does the design of the interface which was based on a traditionally used paper graph (by train traffic controllers) make things easier? Could the immediate feedback be another possible reason for why the users felt more comfortable or was it because of the visualization of the history or the future prognosis of potential conflicts? As mentioned in Chapter 1, in order to answer our general research question a series of studies must be conducted. However, this thesis project, being the very first of our forthcoming studies, will only explore one research question derived from the potential answers to this main question and two explorative sub-questions regarding how we must design our potential future studies.

Thus, we wanted to start our studies by investigating the design of the interface regarding what is being visualized to the users. From a designer's point of view, it is claimed that for the users of complex systems, visualizing a lot of information at a time might be crucial, and could be preferred instead of hiding some of the necessary information in order to make sure that the users can see both the overall picture and the details (Andersson et al., 2014). This approach is considered as helpful for expert users. However, deriving from the aforementioned details about novice STEG users (see Chapter 1), in this study for the case of dynamic systems, the potential effects of showing the novice users a prognosis of their certain actions are investigated. The aim of the present study thus has been to shed light on the importance of graphically presented predictions, referred to as *trend lines* (see Chapter 4) in this thesis. The study primarily focused on investigating the effects of the absence or presence of trend lines in an interface of a dynamic system and was especially designed to look at the novice users' performance and learning. With that said, for this thesis we formulated the following research question:

Research Question: How is the performance and learning of novice users affected by the absence or presence of trends in an interface of a dynamic system?

Hypothesis: The presence of trend lines in a simple dynamic system will accelerate learning and improve performance.

Additionally, a more exploratory aspect of this study is to look at how to define the goal for our future experiments. With an explorative point of view and for methodological reasons, deriving from the related discussions on how we should design our experiments, what methods we should use, and how we should approach these problems in our future studies we also wanted to investigate the right way of defining the goal of the tasks to be introduced to the participants in the microworld being implemented for our studies. Therefore we composed the following question:

Sub-Research Question – 1: How does the introduction of a target affect the user behavior?

Moreover, we are interested in the subjective opinions of the users and their perception of the experience.

Sub-Research Question – 2: How is perceived difficulty affected by the absence or presence of trend lines and the introduction of a target?

Therefore, the long term goal of our study is to understand STEG better through experiments and aid further development of our future studies, with its potential for investigating dynamic systems.

2.2 Delimitations

The domain of train traffic control offers a broad field for research, as well as the use of microworld applications. This thesis study is limited by a number of factors.

Firstly, having based our starting point to evaluations conducted in real work places, our findings in this experiment are not yet generalizable to STEG. Yet it is the final goal of these planned studies, this first one does not serve this purpose. It will only be possible when we complete all the planned studies and transform the microworld we developed into a complete simulation. GridRail currently simulates execution of train traffic. However, as was explained, the use of STEG interface is mainly based on re-planning, and not executing. This is the main reason why our findings are not generalizable to the use of STEG yet. In our future studies the game's interface will be introduced to perturbations and disruptions, followed by real time planning activities. We believe our findings will only be generalizable to STEG by then.

Secondly, the use of microworlds has gained an important place as educational tools in the field of computer aided instruction. However, this experiment is not designed in the context of education. The designed microworld is not an educational game, that is, the learning outcomes achieved through the microworld are not designed to teach a specific subject, but instead the microworld is supposed to help us find answers to our experimental questions.

Additionally, the study will be focused entirely on novice users. There will not be any comparisons between novice and expert performances, and no such long-term training will be given to the novice users. How they develop in complex environments over long periods of practice is outside of the scope of this study. This is one of the topics that is planned to be covered in our future studies.

Chapter 3 Related Work

The form of decision making that is relevant when interacting with STEG includes a number of different issues, for example problem solving, working memory, and learning. As novice operators solve conflicts and re-plan traffic by interacting with STEG they make decisions, and this cognitive process combined with perceptual inferences requires them to use their working memory as they approach to the problems. GridRail is designed to imitate this process and accordingly to evaluate the user behavior. It is therefore necessary to introduce and include literature on this as a background to the study conducted and is of great importance to understand how to compose the related future studies.

Thus, the related studies about problem solving are explained and especially how novice users approach these problems or how the user representations of problems differ is presented. Moreover, how perception and cognition works when people make decisions as they solve problems in dynamic environments are introduced in addition to the related types of learning that takes place when novice users interact with dynamic systems such as STEG and GridRail.

3.1 Human Cognitive Processing

The field of cognitive science is devoted to exploring the nature of human cognitive processes such as reasoning, decision making, problem solving, attention, perception, memory and learning etc. (Hutchins, 2000). It is known that human cognition is well adapted to its natural ecology, and for many years, researchers have been explaining its reasons from highly contradictory perspectives. Although Daniel Kahneman, Amos Tversky and other cognitive psychologists tried long to disprove the belief that humans are rational decision makers (Tversky & Kahneman, 1974, 1983); based on his fieldwork studies Gary Klein (1999) claimed that humans are excellent problem solvers and viewed people as inherently skilled and experienced. However, today, in most cognitive science literature there are two fundamentally different cognitive processes; and these are referred to as *System 1* and *System 2* (Kahneman, 2011). Daniel Kahneman (2011), when describing these two systems in his book “*Thinking, Fast and Slow*” refers to the terms as two fictitious characters, and describes the workings of the mind as an uneasy interaction between the two. System 1, which is also referred to as intuitive judgement, is known to be the simplest cognitive process we have. Relieving us from mental computations, it is rapid and automatically responding to stimuli with low level processing and efficient pattern recognition. If for instance, we need to answer a question, it simultaneously generates the answers to related questions and may substitute a response that more easily comes to mind for the one that was requested, meaning that it is highly error-prone and comes with a number of consequences in terms of biases (Kahneman, 2011). As Evans (1989) thoroughly demonstrates and explains the different kinds of biases in human reasoning in his book, it is known that many fallacies in our judgments and inferences are the results of this phenomenon known as ‘cognitive heuristics‘, which basically belongs to the System 1 processes. So, System 1 is not constrained by capacity limits but its answers are mostly only approximately correct and it sometimes makes mistakes. So using the analogy of a minefield as Kahneman states: “The way to block errors that originate from System 1 is simple in principle: recognize the signs that you are in a cognitive minefield, slow down, and ask for reinforcement from System 2“(Kahneman, 2011). System 2, allocating a lot of attention to the task at hand, takes its time to think just like the times when we are asked to answer the problem 17x24=?. Using the working memory it solves the problems. However, on the down side, System 2 is limited in capacity and is easily disturbed.

3.1.1 Working memory

In 1968, the theory of short-term memory was developed by Atkinson and Shiffrin (as cited in Anderson, 2010). The theory proposed that the received information first went into a limited short-term memory and for the information to go to a relatively permanent long-term memory, it had to be rehearsed. Otherwise, it would be lost forever (Anderson, 2010).

In 1974 the concept of short-term memory was replaced with that of working memory by Baddeley and Hitch (as cited in Anderson, 2010). According to the theory, the working memory system has four components: (1) a modality-free *central executive*, (2) a *phonological loop*, (3) a *visio-spatial sketchpad*, and (4) an *episodic buffer* (Baddeley, 2001). The episodic buffer is a temporary storage system that holds information (Eysenck & Keane, 2005), and the phonological loop and visio-spatial sketchpad are what he called slave systems. In order to understand these terms let us remember the multiplication problem above; when we are asked to multiply 17 by 24, what we do is to develop a visual image of the written format of the problem “17x24” by our visio-spatial sketchpad, and as we proceed with the multiplication we find ourselves rehearsing the stages of the solution through our phonological loop. The central executive, resembling attention, is the key component of working memory, and it is the one that puts or retrieves the information into the slaves, as well as controlling the slave systems (Anderson, 2010).

3.1.2 Problem Solving

Problem solving is defined as,”cognitive processing directed at transforming a given situation into a goal situation when no obvious method of solution is available to the problem solver” (Eysenck & Keane, 2005). When having to come up with a solution, what people must do is to look for operators, and select one that takes them to the solution from multiple other choices (Lovett & Anderson, 1996). However, due to the fact that only few paths take the problem solver from the initial state to the goal state, according to Newell and Simon (1972), we rely highly on heuristics or rules of thumb. Their theoretical approach is consistent with our knowledge of human information processing.

For example, we have limited working capacity and that helps to explain why we typically have a tendency to choose the shortest path or use heuristics such as *means-ends analysis* and *difference reduction methods* rather than algorithms.

One of the most common methods particularly being used in unfamiliar domains is the *difference reduction method*. That is, simply to reduce the difference between the current state and the goal state. Köhler (as cited in Anderson, 2010), who conducted experiments for understanding the mentality of apes in 1927, gives an example with a chicken that plainly demonstrates this method. Imagine a chicken which "will move directly toward desired food without going around a fence that is blocking it" (Anderson, 2010). In other words, the difference-reduction method assists the problem solver based on the evaluations of the similarity between the current state and the goal state. Although it is claimed that the difference reduction method mostly works, one might end up in a position just like in the chicken-fence case. This is simply because of the fact that in order to solve some problems one might sometimes need to go against a step of similarity. The solution to the *hobbits and orcs* problem (a.k.a. *missionaries and cannibals*) simply illustrates this issue. The problem is given as follows:

> [...] On one side of a river are three hobbits and three orcs. They have a boat on their side that is capable of carrying two creatures at a time across river. The goal is to transport all six creatures across to the other side of the river. At no point on either side of the river can orcs outnumber hobbits (or the orcs would eat the outnumbered hobbits) (Anderson, 2010).

Greeno (1974), based on his research on such river-crossing problems, points out that humans often do plan small sequences of moves, and the solution to the hobbits and orcs problem necessitates the subjects to make a series of moves or transfers of hobbits and orcs back and forth across the river (Thomas, 1974). There are 12 successive states that lead to the goal state in the typical solution to the problem, and state-6 requires the users to move two creatures back to the wrong side of the river. The move seems to be going further from the goal. For that reason, in their study Jeffries et al. (1977) found out that about one third of the participants, instead of executing state-6, chose to back up to a previous state. Therefore, interestingly, they preferred to undo a move instead of taking a step that moves them to a state that in the first sight appears to be further from the goal.

These experiments on problem solving revealed some findings about human behavior. For example, the studies of Thomas (1974) has shown that when the problem solvers had to make a move that temporarily increased the distance between the current state and the goal state they experienced severe difficulties. He has also observed that the participants divided up the problems and formed sub-goals (Eysenck & Keane, 2005).

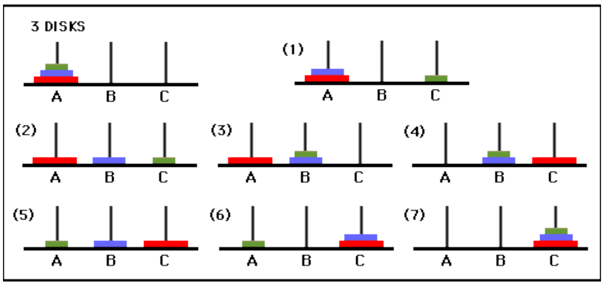


Figure 7 The solution to the three disk version of the Tower of Hanoi Problem. (http://mathforum.org/dr.math/faq/faq.tower.hanoi.html)

Means-ends analysis is similar to the difference reduction method; however it is a more sophisticated version. One firstly notes the difference between the current state and the goal state; secondly, sets a sub-goal that will reduce the difference between the current and the goal state; and lastly, selects an operator to attain the sub-goal (Eysenck & Keane, 2005).

A nice example study can be given from Köhler's (as cited in Anderson, 2010) experiment on chimpanzees. In 1927 the chimpanzee Sultan was posed a problem which was to get some bananas that were placed outside his cage and he was given two poles, neither of which he could reach the bananas with. So after trying for some time, he suddenly came up with a sub-goal which is to put one pole inside the other, creating a long enough pole to reach the bananas. The chimpanzee Sultan did not randomly try to reach the bananas, instead at some point he created a tool that assisted him to get the food. He generated a sub-goal to be achieved, as a means of achieving his actual goal (Anderson, 2010).

The Tower of Hanoi problem (see Figure 7) is another good example to illustrate the means-ends analysis. In the problem there are three pegs and three discs of different sizes. There are holes in the pegs for the problem solvers to stack them on the pegs. In the initial state the discs are all placed on the 1st peg and the goal is to move them all to the 3rd peg on which they must stand in the same order. The constraint is to do this by moving one disc at a time. A reasonable sub-goal of the problem is to try to place the largest disc on the last peg in the early stages of the solution. Otherwise achieving the actual goal state becomes more or less impossible.

Anzai and Simon (1979) used the five disk version of the Tower of Hanoi problem in the context of a learning-by-doing experiment. They recruited one single participant and the experiment took 1.5 hours divided in 4 episodes. They asked the participant to think aloud as she tried to solve the problem. The results they obtained are intriguing. I will try to summarize their findings under five points. (1) First of all, their findings have shown that after working for 1.5 hours on the problem the subject found at least four distinct successive strategies, however, with varying degrees of awareness. (2) They observed a gradual transformation in her strategies. Except in the last case, where according to their findings the participant used the information she had stored in her long-term memory about the previous trials as a cue for her strategy transformation. They believe in the last case, the perceptual cues which were received from external elements and affected her decisions. (3) As she tried to solve the problem she used her short-term memory to store the sub-goals she generated during the course of problem solution. On the other hand, as well as perceiving relevant information from the problem situation, she used some related information she retrieved from her long-term memory. They also showed that rehearsal occurred in a trial only when the participant needed some information that she revealed in her earlier trials. (4) They observed four main processes as she approached the problem: "(a) applying the current strategy, (b) gathering information that will later be used to modify the strategy, (c) using information gathered in previous episodes, and (d) deciding to terminate the solution attempt (successfully or unsuccessfully)" (Anzai & Simon, 1979). Their fifth finding is about when she changed her strategy. (5) During the first episode the strategy changes were less conscious. Between the first and second episodes she explicitly verbalized the changes in her strategy. The same happened between the second and third episodes, but during the fourth episode the strategy changes were again less conscious. The researchers believe the reason behind this finding might be the plurality of her learning processes, meaning that the learning that took place was both cognitive and perceptual. (Anzai & Simon, 1979)

Their findings have also shown that the general strategies she used, such as avoiding from returning to the formerly resided states, or in order to achieve a goal- preferring shorter sequences of moves over longer ones etc; led the participant to learn better sequences of moves in time. And to solve the problem in the later stages of the experiment, these sequences were used during the trials (Eysenck & Keane, 2005).

3.1.3 Skill Acquisition

It usually takes people years to acquire knowledge and skills in a particular area. Ultimately, this long-term learning allows them to develop expertise. In this thesis, our focus is on novice users and we look into the processes involved on the road to improving cognitive, perceptual or physical skills. In his paper about acquisition of intellectual and perceptual skills, Rosenbaum et al. (2001) defines skill acquisition as follows:

> […] When we speak of a "skill" we mean an ability that allows a goal to be achieved within some domain with increasing likelihood as a result of practice. When we speak of "acquisition of a skill" we refer to the attainment of those practice-related capabilities that contribute to the increased likelihood of goal achievement.

Additionally, even if a task requires cognitive effort, by practice one reduces the thought required for a task to be completed by actually reducing the central cognitive component of information processing (Anderson, 2010). For example, when in traffic, driving a car requires a lot of things to be considered, such as traffic rules, behavior of other cars, bikes, pedestrians, routes, speed, weather conditions etc; yet an experienced driver can do it without expending much cognitive effort. However, in the case of a novice driver, who probably gets some instructions before the actual driving experience, will spend much effort both cognitively and perceptually.

The development of a skill typically comprises three stages; the *cognitive stage*, the *associative stage*, and the *autonomous stage* (Aderson, 2010). The cognitive stage is the first stage when only a declarative encoding of the skill is developed and the use of knowledge is noticeably slow. This is the stage where most novices are categorized in. The knowledge used is still in the declarative form in this stage. The second stage is the associative stage. In this stage, the errors in the initial representation are slowly distinguished and avoided. It is mostly procedural knowledge that carries out the skill. The autonomous stage, referred to as the final stage, is the one when the skill requires fewer cognitive processing resources and becomes relatively automatic (Anderson, 2010).

3.1.3.1 Tactical Learning

Tactic means a method that provides goal achievement. When we practice problems, we usually learn the sequence of actions required to solve a problem or parts of it (Anderson, 2010). Tactical learning is defined as "a process by which people learn specific procedures for solving specific problems" (Anderson, 2010). This implies that when tactical learning occurs we simply learn how to execute sequences of actions that lead us to the goal state.

For instance, in his experiment with the hobbits and orcs problem Greeno (1974) observed that the participants learned the sequence of moves to get the creatures across the river. He found that once the participants had learned a sequence, what they did was to simply recall it in their future trials. They did not have to re-explore the successive solutions. In terms of tactical learning he discovered that it took only about four repetitions for the participants to perfectly solve the problem (Anderson, 2010).

Another example is the study of Jenkins et al. (1994) about learning motor sequences. Using a positron emission tomography they investigated participants' learning of the various sequences of finger presses such as "ring, index, middle, little, middle, index, ring, index". Comparing the participants' initial learning of the sequences with the learning of the participants who previously practiced these sequences, their findings have revealed that, early in a task, the part that deals with rational cognitive functions in the brain takes charge in organizing the behavior. However, late in learning, participants are just recalling the answers from memory. (Jenkins et al., 1994)

3.1.3.2 Pattern Learning

In 1965 de Groot conducted a study about pattern recognition. The participants were selected from chess players and a short-term recall task was given both to master and novice chess players. The pieces were meaningfully placed on the chessboard and several configurations were presented to the participants. Their task was to recall the positions of the pieces located on the chessboard after seeing it for 5 seconds. The masters could reconstruct the configurations even with 20 pieces. However the novice chess players could only reconstruct 4 or 5 pieces (Newel & Simon, 1972). The findings for the novice users were in line with the capacity of working memory. Later, when the configurations of pieces were changed to random formations even the masters could not perform well. Thus, the results were explained by pattern learning. As a result of experience with chess, the masters could remember these common patterns, but not the individual places of the pieces (Chase & Simon, 1973).

3.2 Distributed Cognition

When distributed cognition was first developed as a new way of examining cognition in late 1980s, Hutchins explained that their approach was derived from the theories and methodologies of different disciplines such as cognitive science, cognitive anthropology and the social sciences. In contrast to the traditional means of studying cognition, distributed cognition covers and examines a wider spectrum of actors within a system. The traditional view of cognition was focused on explaining cognitive phenomena at an individual level. However, a distributed cognition approach puts the emphasis on the distributed nature of cognition across different individuals, artefacts, and their internal and external representations (Rogers, 1997). Moreover, Hutchins (2000) claimed that when human activity is observed, one encounters three kinds of distribution of the cognitive processes; it could be distributed among the members of the social group, on the operational coordination between internal and external material or environmental structure, and it could be distributed through time.

Today, distributed cognition is typically applied in areas such as human-computer interaction, computer-supported cooperative work, and computer supported collaborative learning. Before distributed cognition gained a seat as a foundation for the field of human-computer interaction, most research was confined to investigating one desktop and its user (Hutchins, 2000). The main standing point in a distributed cognition approach, at the work setting level of analysis, lies behind the belief that the cognitive processes in human activity took place not only inside, but also outside of an individual actor. Therefore, for a given activity the method applies the same concepts not only to one individual, but also to the interactions among several human actors and technological devices (Rogers, 1997).

One other important aspect to be noted about distributed cognition is how, in this domain, the nature of representations is taken into account. The nature of representations and how they are used in work related activities has been one of the most important factors within the field of distributed cognition. In terms of traditional information processing, psychology symbols are considered as referring to things other than what they actually are as objects. As illustrated in the examples below, in the field of distributed cognition unlike the traditional approach, not only what they resemble but also the strategies people may develop through the properties of these representations are being investigated (Hutchins, 2000)

In 1990s a number of studies were conducted with a focus on cognitive systems of work practices such as ship navigation (Hutchins, 1990) cockpit automation (Hutchins, 1995), air traffic control (Halverson, 1994) or other engineering practices (Rogers, 1997).

3.2.1.1 Ship Navigation

After his cognitive ethnography studies about ship navigations, Hutchins realized that the cognitive properties of one single navigator did not really affect any important consequences for the ship, but instead what mattered were the interactions among several navigators and how they interacted with a set of tools (Hutchins, 2000). This study focused on the cultural-cognitive processes that take place when steering a ship into harbour (Rogers, 1997). In particular, the study explored navigation task and the nature of the tools used in the task, as well as investigating the division of labor among the team members and how they coordinated their activities. The study has shown that in this context the technological devices used are actually perceived as media for representation. In other words, it is claimed that the devices did not actually amplify the cognitive abilities of the team members, what they did instead was to transform the difficult cognitive tasks into easy ones (Hutchins, 1990). The study has also pointed out that learning happened during task execution both at an individual and organizational level. (Hollan et al., 2000)

3.2.1.2 Airline Cockpit Automation

In 1995 Huthins studied an airplane cockpit and how it remembers its speed through the interaction between the pilot and the cockpit devices and tools. In this study, the primary unit of analysis was considered to be this socio-technical system as a whole rather than one individual (Hutchins, 1995). The study focused on the distribution of cognitive activity among the crew, as well as analyzing the interactions of internal and external representational structures. Hutchins observed the pilots in flight, studied the operations manuals, conducted interviews with pilots and participated in the training programs (Hollan et al., 2000). He was highly interested in understanding how information is represented and how representations were transformed and affected the task performances. His analysis has shown that the cognitive properties of the system as a whole differed greatly from the cognitive properties of the individuals who inhabit them (Hutchins, 1995). For example, Hutchins found that when using an airspeed indicator dial, most of the time the pilots did not think of speed as a number, but instead using the spatial structure of the display they tried to perceive the relations among the actual and desired speeds. (Hollan et al., 2000)

3.2.1.3 Engineering Practice

Rogers (1997) conducted a study in an engineering company to find out the effects of the networking technology on their working practices. After a distributed cognition analysis was conducted, the breakdowns occurring in the work activities and various mechanisms were revealed and based on these results the solutions to overcome these breakdowns were documented in order for the engineers to adapt their working practices (Rogers, 1997).

3.2.1.4 Air Traffic Control

Halverson (1994) carried out a study in order to exhibit how distributed cognition theory could aid human-computer interaction. By applying the theory of distributed cognition to the analysis of an automation tool for air traffic control (Halverson, 1994), Halverson revealed important aspects about what should be retained in the existing design and how future automated decision-making tools could be developed in order to improve the work quality of the air traffic controllers (Rogers, 1997).

3.3 Dynamic Decision Making

In 1962 three defining characteristics have been identified in Edward's (as cited in Brehmer, 1992) classic description of *dynamic decision making*. Firstly, a set of decisions has to reach the goal, meaning that many decisions have to be made in order to achieve and maintain control. Secondly, the decisions made are dependent on each other, that is, later decisions are affected by the previous decisions, and finally, the state of the problem changes in time, either with respect to, or regardless of, the decision maker's actions (Brehmer, 1992).

Brehmer and Dörner's framework for studying dynamic decision making is most relevant for the domain of this thesis, because unlike, Tversky and Kahneman (1981) or Gigerenzer (2008); in their experiments they could control the dynamics of the environment. For that reason in their studies they could neither go to the field like the researchers of *naturalistic decision making*, nor they could use pen and paper methods like most researchers did when studying general decision making theories. Their solution thus involved an experimental paradigm that allowed them to study dynamic decision making empirically. With the help of emerging technology in 1980s, they had the chance to create computer simulations of real world environments.

Some examples of microworlds used in dynamic decision making research are explained below.

3.3.1.1 Lohhausen

In 1983, a computer simulation that simulates the dynamics of a small German town, Lohhausen, was developed by Dörner and his colleagues (as cited in Brehmer & Dörner, 1993). The subjects had to act as the mayor and rule the town for a period of ten years with their dictatorial powers. The simulation presented the subjects with a set of action possibilities that varied over time. The course of the events changed based on the interaction history (Brehmer & Dörner, 1993).

Expecting to see better performances, Dörner et al. (as cited in Brehmer & Dörner, 1993) gave one group of subjects a short course about some methods, whereas the control group didn't receive any such training. The results have shown that, the subjects did not perform any better than the subjects in the control group. (Stuhler & DeTombe, 1999)

3.3.1.2 Moro

The Moro is structurally highly similar to Lohhausen, and was also developed by Dörner and his colleagues. Both microworlds were designed to present complex problems that engaged a variety of cognitive processes (Brehmer & Dörner, 1993). In Moro, the subjects' task was to act as an advisor to a nomadic tribe in the South Sahara and the goal of the game was to increase their welfare by balancing different variables such as cattle, food production and water supplies and avoiding tsetse flies and diseases. The subjects, by adjusting these critical relations should avert results such as starvation or drought, potentially caused by lack of cattle or insufficient water supplies (Stuhler & DeTombe, 1999).

In 1995 Brehmer and Jansson conducted an experiment using the microworld Moro as a tool. The six most important variables of the microworld were told to the subjects in one group before the experiment while the subjects in the other group were not given any such information. Although the participants who were informed about the six most important variables paid more attention to these variables, their performance was no better than the participants in the other group (Stuhler & DeTombe, 1999).

3.3.1.3 Dessy

DESSY (Dynamic Environment Simulation System) is another example microworld which was developed by Brehmer and Allard (as cited in Brehmer & Dörner, 1993) in Uppsala University in 1990 and used to study how subjects control forest fires. Unlike Klein's (1999) studies about fire fighters in real world environments, the microworld only supported very limited scenarios. Experiments in order to investigate human behavior were conducted with the system.

In Chapter 1 and 2, some of the alternative explanations there is to the success of STEG were mentioned. We argue that three of these are: (1) the analogy in the interface with the real domain. Supporting the users' mental models, the design of STEG was inspired by the traditional methods – this may explain why it is easy for the operators to understand and use STEG; (2) the interactive relationship in the form of immediate feedback when interacting with STEG. Task representations are supported in the interface design by providing relevant information immediately for gaining situation awareness; (3) the fact that information presented in the STEG interface is mainly picked directly with perceptual processes and that this relieves some of the cognitive burden placed on train traffic controllers when have to do this on a cognitive level instead. It is this third explanation that we explore in this first study. For that reason within our research group a browser based logical problem solving game GridRail was developed to serve as a microworld. The details of GridRail and the general characteristics of microworlds are discussed below in Chapter 4.

Chapter 4 Methods

4.1 Microworlds

We used the microworld paradigm in order to answer our research questions. For investigation of different innovative features and related elements within complex systems, the use of discrete game simulations is known as a common approach, especially within the railway sector (Lo et al., 2013). Before explaining the experimental design and the details of the game, I would like to reason why we used this approach.

There has always been a huge gap between field research and laboratory experiments, and that has been hindering psychological research due to their distinctive disadvantages (Brehmer & Dörner, 1993; Omodei & Wearing, 1995). Especially, when doing field research with a focus on decision support systems (e.g. STEG) and their users, it is considered a necessity to take lots of factors into account. This kind of research within complex, dynamic environments (e.g. train traffic control centers) requires analyzing these environments where there are a high number of interconnected variables that repeatedly keep influencing each other. In such environments the parameters and variables keep changing and evolving even if the users do not interact with the system. For that reason, field studies involve a lot of challenges for researchers who want to analyze or influence such complex, dynamic environments (Brehmer & Dörner, 1993).

Although the importance of identifying and explicating environmental constraints is stressed, many would also support the claim that a comprehensive analysis of a complete environment is often not feasible for practical reasons (Vicente, 1999). So the researchers would tend to focus on the parts of the environment which is most relevant to their study. However, this approach risks ignoring the overall picture and might lead to sub-optimization by a lack of the best possible coordination between different components, elements or variables within the field. Hence, we need methods that keep our holistic view and help us focus on and control the important components, elements or variables. (Tschirner, 2015)

In the early 1990s, the advances in computer technology led Brehmer and Dörner to carry such decision making studies from the field to the laboratory environments, where they created computer simulations of real-world environments with complex, dynamic scenarios (Brehmer, 1992, 2005; Brehmer & Dörner, 1993). Within these microworlds the experimenter, being the constructor of the so-called field, knows every aspect of the field in a way that the field researcher cannot. So he or she could control the conditions, dynamics and variables of the field, further finding a chance to recruit and test more participants in such settings and generating more results in shorter periods of time.

Eventually, with microworlds understanding and resolving complex human behavior became feasible in a way that was not possible in field studies as Brehmer & Dörner (1993) explains below:

> […] to provide a tool for doing very complex experiments, hopefully enabling us to bridge the all too great gap between laboratory and the field that has plagued psychological research since the beginning. Thus, it is now possible to create very complex systems, such as a small town or an industrial process in the laboratory, and study how subjects interact with this simulation.

As a matter of fact, these computer simulations, a.k.a. microworlds, yet giving only a very abstract and limited representation of the real-life situations, can be effectively used to examine people's behavior. Microworlds by nature have three general characteristics. They are considered to be complex, dynamic, and opaque (Brehmer & Dörner, 1993). Actually, these are the characteristics that make microworlds represent the cognitive tasks people usually encounter in their work environments (Brehmer & Dörner, 1993). For that reason, the microworlds are also inherently stressful environments. They simulate this complex and the dynamic character by forcing the participants to consider a number of things at a time and to make trade-offs as they try to act in real time, not knowing at which pace they have to make decisions (Brehmer & Dörner, 1993).

When studying with microworlds it is essential to keep the norms of microworlds in mind. For that reason, as I explain the microworld we used in our study below, I would like to briefly refer to the general characteristics of microworlds and try to explain what the participants are expected to do in a standard microworld. In order to diminish the potential questions that might arise in the minds of readers about our analysis and discussion of our results on the participants' performance, difficulty perception and behavior, I will try to explain these general characteristics as I introduce the microworld we used.

4.1.1 The Microworld Used in Our Experiment: GridRail

Our microworld presented the participants with a highly dynamic river-crossing type problem that engaged a variety of cognitive and perceptual processes. The microworld, which could also be considered as a browser based game, was a basic simulation of a railway track with six trains; three at each end of the track (see Figure 8).

4.1.1.1 Task and the Goal

The participants' task was to play the browser-based game in 40 trials and try to improve their performance. In the experiment, before a participant was allowed to start interacting with the game, he or she was handed an instructions sheet (see Appendix C) that was manipulated in accordance with the conditions he or she was being subjected to. The goal of the game was written on this instructions sheet, as it is to solve a logical problem as fast as possible. The problem that the participants had to solve in the game is that the three trains standing on the left end of the track (red, pink, yellow) have to be moved to the opposite end of the track, and the other three trains standing on the right end of the track (black, grey, white) have to be moved to the left end of the track, changing places with the aforementioned three trains.

Taking all the information written in the instructions sheet into consideration, this problem could be referred to as a well-defined, knowledge-lean problem, meaning that all the aspects of the problem are clearly specified; including the initial state or situation, the range of possible moves, and the goal, in addition to the fact that the problem does not require the possession of any specific knowledge and that most of the information required is given in the problem statement.

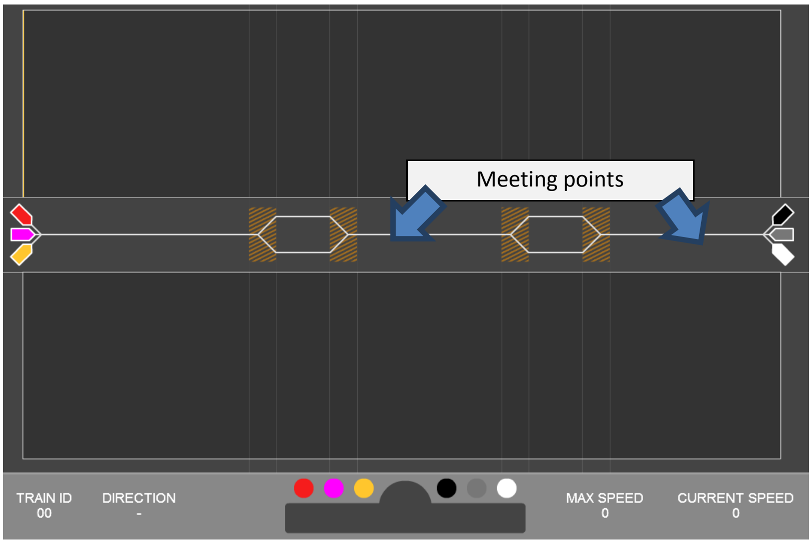


Figure 8 Screenshot from the game

4.1.1.2 The Rules and Instructions

Microworlds usually have their own set of rules (Brehmer & Dörner, 1993). The instructions sheet involved the goal of the game, the core features and rules of the game and information on how to use the controllers. A trial ended when the participant managed to move the three grey-scale trains to the left end of the track and the colored trains to the right end of the track. When a trial ended, the participants could see the time they spent and were given a possibility to start a new trial by clicking on the start button as can be seen in Figure 9. It should be noted that the timer started counting once they selected the first train.

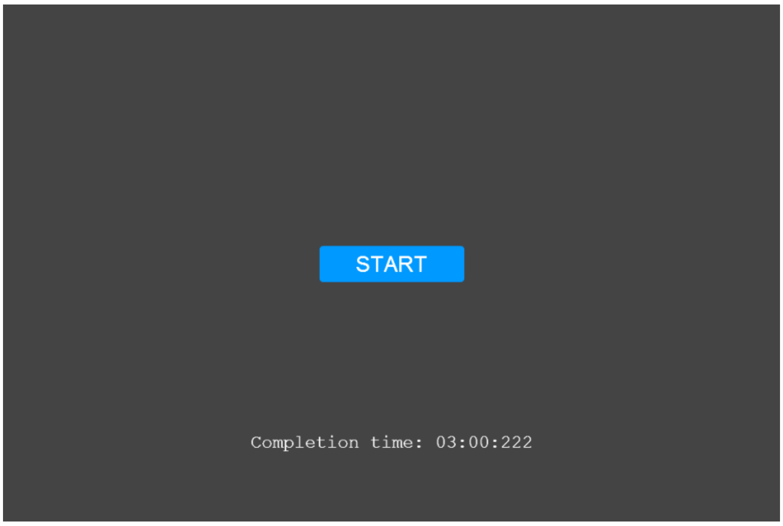


Figure 9 End of Game Screen

The set of rules written in the instructions sheet included information about the features, obstacles or limitations of the game. There are two designated meeting points along the track, as depicted in Figure 8, and trains can only meet where the two tracks of the meeting points are parallel.

Players control the trains in the lower panel of the window. A train is selected by clicking on the correspondingly colored circle in the lower panel. The selected color is highlighted and also displayed in the speed control. The speed of the selected train is set by clicking on the speed control, as can be seen in Figure 10 below.

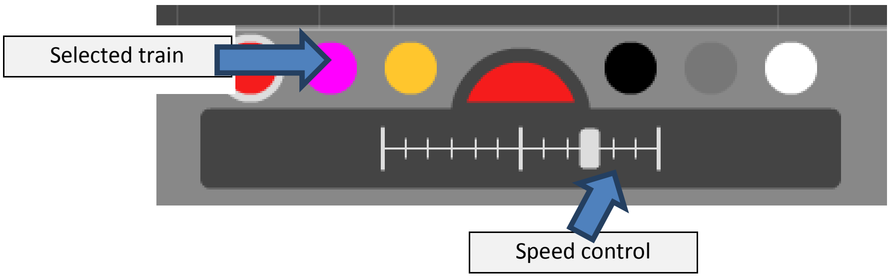


Figure 10 Game controllers, close-up view

The instructions sheet explains the microworld to some extent, but leaves out some features that the participants are supposed to learn themselves as they interact with it. For example how many trains could fit on a side track or if the trains had different speeds (see Table 1) were not written in the instructions sheet. These features were expected from the users to be learned during the gameplay.

> "[…] The subjects must *learn* about the microworld. In these experiments the subjects are usually confronted with a system that is, at least in part, unknown to them. They may know the general form of relations among the variables in the microworld, […] but they are usually not familiar with the time constants or precise nature of other quantitative relations in the microworld. This means that to act, they must form a picture of the specific nature of the microworld they

face; they must collect and integrate information and form hypotheses about hidden structure of the microworld." (Brehmer & Dörner, 1993)

Table 1 Vehicles, their properties, starting and ending stations

Train ID	Color	Max. Speed	Starting Station	Ending Station
01	Red	60 km/h	Station - 1	Station - 2
02	Pink	100 km/h	Station - 1	Station - 2
03	Yellow	120 km/h	Station - 1	Station - 2
04	Black	60 km/h	Station - 2	Station - 1
05	Grey	100 km/h	Station - 2	Station - 1
06	White	120 km/h	Station - 2	Station - 1

Microworlds are opaque in this sense, referring to the fact that not everything is visible and that some aspects of the system have to be inferred, just like the dynamics of the microworld which are mostly learned by the participants as they interact with the system. However, since most microworlds represent actual systems, in our case the railway track, it is unavoidable that the participants will bring their pre-experimental knowledge about these real systems to the lab. So it is expected that the participants know the general form of the relations among the variables in the system such as the general properties of a train or a railway track, but what they do not know and will have to learn during the experiment will be the precise quantitative values such as the time constants or maximum speeds of the trains.

4.1.1.3 Trend Lines

In the top of the game screen, participants see lines representing a prognosis of the trains' future horizontal positions. These are referred to as *trend lines* in this thesis. In Figure 11, they are illustrated in pink and white over the two correspondingly colored trains. The trend lines are derived from STEG interface, where the main screen was designed based on the traditional, printed paper graphs which are used by train traffic controllers to plan the daily traffic. It is basically a time-distance graph, illustrating the distance each train travels, their relative speeds, and at which stations they stop. Inspired by this paper-graph, STEG interface was designed to visualize the history, present situation and the planned traffic. As explained in Chapter 1, due to this, STEG interface supposedly supports the development of the users' mental models.

The trend lines designed for this game appear over every train that is initiated and keep moving horizontally along the track as the trains travel. Using the controllers in the lower panel, if the participants change the speed of the selected train, the trend line correspondingly changes its angle of inclination. This, for instance, helps the participants to predict the future positions of the trains and understand where the trains would meet by looking at the intersection points of two trend lines.

In a standard microworld the participants are expected to make prognoses concerning the future development of the microworld and form expectations concerning the future (Brehmer & Dörner, 1993). On the basis of that, in the game the trend lines are designed in a way that we could hide or visualize them so that we could see whether they have any effects on the participants' performance, behavior, or difficulty perception.

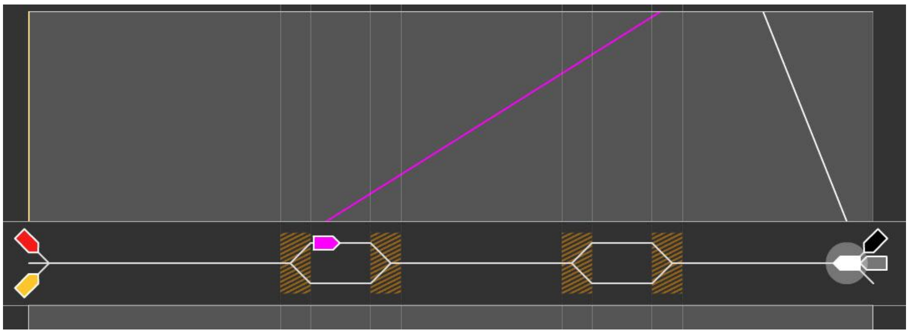


Figure 11 Game Screenshot with Trend Lines

The participants acting within a microworld are expected to constantly be aware of the developments of important variables and the effects of their own actions. Consequently, they are supposed to take corrective actions based on the hypotheses they simultaneously keep forming and testing. So, we wanted to investigate whether visualizing such information will affect any of these processes; because convincingly, as Brehmer and Dörner (1993) also suggest in their studies: "subjects are supposed to be able to organize all these different activities into some coherent whole and that the microworld experiments thus do not only require 'thinking,' or 'problem solving,' or 'planning' or 'decision making,' but all of these together, as well as the coordination of these activities".

4.2 Pilot Study

A pilot study was conducted in order to test the feasibility and the structure of the research project and to see what the results are likely to be. It was conducted in a controlled setting in an isolated room where the participants used a laptop to play the game. We recruited 4 participants (3 males, 1 female) with an average age of 28.5, and they were selected from either university staff or PhD students in the IT (Information Technologies) department. The experiment took place in different times for each participant.

The game interface had two versions. It was either generated by the software; with the trend lines or without them. Also, we introduced the goal to the participants either with a target or without one. Without a target the goal was defined as "to solve the problem as fast as possible", and with a target the goal was defined as "to solve the problem in less than 32 seconds". After filling in the background questionnaire, each participant played the game under a different condition. The first participant played the game with no trend lines being visualized in the game interface and was not introduced to a target in the definition of the goal. The second participant played the game under the condition in which the game presented the trend lines, and again the goal was introduced without a target. Yet the third and the fourth participants played the game in which a target was introduced, and the trend lines were visible to one, but not to the other. The participants played the game in three blocks, 25 minutes each. After the first and the second blocks, they had to take 5 minutes of breaks. At the end, an interview with each participant was conducted. The results obtained in this pilot study, were used in the planning of the main research project.

The experiment had to be designed to manifest a learning experiment layout. Hence, we believed that the users would need to interact with the system for at least 1 to 1.5 hours in order to understand and learn the dynamics of the game, find the best strategy to solve the posed problem and to improve their motor skills so that they can put their thoughts into practice. Taking this into account, we first designed the experiment to the extent in which the players would play the game for 75 minutes in total, with 5 minutes of breaks after the 25th and the 55th minutes. However, after the pilot study we realized that this approach was generating some practical problems in comparing the data collected. Since the participants were allowed to think as long as they want in between each trial before they restarted the game for a new trial, the number of trials the game was played by each participant exhibited a high variance. In other words, the trials were self-paced, meaning that the participants could decide for themselves the tempo between the trials and it seemed to vary a lot. For that matter, in the actual experiment we decided not to use time as a variable but instead the number of trials. And since, according to the data collected, the learning took place mostly in the first block and seemed to remain steady afterwards, the number of blocks was decided to be set to 2 instead of 3.

4.3 Experimental Design

The experiment is designed to test our hypothesis to research question 1 and to explore research questions 2 and 3. Therefore, a between subjects study, which is considered to be a 2 by 2 factorial design, was configured, such that our independent variables were defined as the trend lines (absent or present) and the target (absent or present), whereas our dependent variables are set to be the performance and the perceived difficulty.

As can be seen from Table 2 below, we determined 4 conditions to test our hypotheses; (1) without trend lines, without target, (2) without trend lines, with target, (3) with trend lines, without target, (4) with trend lines, with target. To answer our first research question regarding the importance of the absence or presence of graphically presented predictions, in two of the conditions, differing from the other two conditions, the participants could see the trend lines. In order to answer our second research question about the effects of introducing a target as a specific goal, in two of the conditions in the instructions sheet, the task definition did not include a target of completion time in seconds, but only a statement that they will be playing a game with the aim of solving a logistic problem as fast as possible; while in the other two condition, the instructions sheet stated that it was possible to solve the problem in less than 32 seconds and told them to aim at a score that is at least as slow as 32 seconds. This number was calculated by adding a margin of 4 seconds to the number 28, which was achieved only via the ultimate solution pattern identified by the developer of the software.

Table 2 Conditions and Number of Participants for each Condition

	Without Trend Lines	With Trend Lines
With Target	8 participants	8 participants
Without Target	8 participants	8 participants

In Table 2 the way we used the word *target* should not be misinterpreted. As was also explained in the previous sections, in our study, since we are using microworld paradigm all our participants are, of course, introduced to a goal. When we say “without a target”, instead of “with a target”, we mean that the participants in this condition were not introduced to a certain number of seconds to achieve, but instead, they were only told to solve the problem as fast as they can. The word *target* here refers to the intention of solving the problem in less than 32 seconds.

4.4 Participants

32 participants (17 male and 15 female) in total were recruited for the experiment. Their age varied between 21 and 34 years, with an average age of 24.75 (SD = 2.94 years). They were selected from university students, all studying in technical programs, such as computer science, electrical engineering, or physics etc. either for their bachelor’s, masters or PhD degrees. The average programming experience was 2.06 years, and on average the participants reported spending 7.37 hours each week on playing games or solving puzzles, including computer games, mobile phone games, board games, and newspaper puzzles like Sudoku etc. Each participant received a movie ticket as a small compensation for participating in the experiment. None of the participants had any prior knowledge about the purpose of the experiment or of the specific hypotheses being tested.

4.5 Environment and Materials

The experiment took place in a controlled setting, an isolated room at ITC (Information Technology Center), Polacksbacken and the tests were carried out over 32 separate sessions under the supervision of the test leader. The participants used a commercial off-the-shelf laptop (13’’ screen-size), running Ubuntu as the operating system, and the assignments were performed using a browser-based game running on a Firefox (v14.0.1) browser in full screen. The users were not allowed to use the touch-pad, meaning that the only way to interact with the software was to use the mouse. The game was built using *Adobe Flash CS*.

4.6 Procedure

Upon arrival at the room, the participants were briefly informed about the study and were asked to read and sign an informed consent form (see Appendix B). After signing the consent form they were told to fill in a background questionnaire (see Appendix A) that appeared on the computer screen. When they completed the background questionnaire an instructions sheet was given to the participants. The instructions sheet briefly introduced the problem to be solved and the goal to be achieved, as well as informing the users about some of the basic rules of the game and how to use the controls with illustrations from the game interface. The instructions sheet was read aloud by the test leader as the participants followed the same text on the copy they were given. The game was set to run in 2 blocks of 20 trials each, where the participants had to take 5 minutes break in between the two blocks. After the completion of each trial they could see their times and could click on the start button in order to start a new trial. They were allowed to wait as much as they wanted before they started a new trial. In order for the participants to rate the difficulty of the game a slider appeared on the screen at three different points; after the first trial and at the end of each block. After the completion of 40 trials, a semi-structured interview took place. Depending individually on the participant, the study took approximately 1 to 1.5 hours per participant in total. After the experiment, the participants received their compensation and were able to ask questions about the experiment and its purpose.

4.7 Measurements

A script was used to collect, save and export all the quantitative data we needed. All participants were subjected to a pre-test background questionnaire which provided the descriptive data about the participants. The dependent variables measured were performance and perceived difficulty. For performance, we measured the completion time for each trial. To measure perceived difficulty, the participants were asked to rate the difficulty of the game three times during the experiment by the help of a difficulty scale slider which had labels on each side; *very easy* and *very hard*. It appeared on the screen after the first trial and at the end of each block. The settings of the slider were adjusted to have an interval from 0.0 to 1.0. The interviews were recorded and later transcribed for a consecutive analysis of the received information about the related issues.

The quantitative and qualitative data depicted in Table 3 were collected:

Table 3 Collected Data

	Collected Data
1	Trial Completion Time
2	Perceived Difficulty
3	Post-Test Interview Results

Although not used in our analysis, the mean time spent between trials was also measured and calculated in order to understand how long the participants waited before starting a new trial. Additionally, we recorded the gameplay of the participants but have not used this data in the thesis either. When we write a script to quantify this data set, we will be able to categorize and have a deeper understanding of the strategies used.

Chapter 5 Results

5.1 Quantitative Results

5.1.1 Performance

Figure 12 is an illustration of the mean performance values with respect to the number of trials for different conditions.

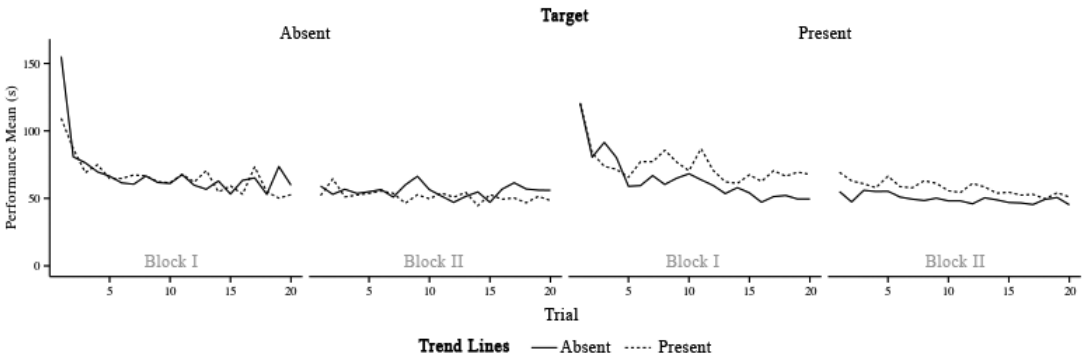


Figure 12 The graphs show the mean performance in seconds for the four different conditions.

In order to obtain a normal distribution we employed a logarithmic transformation of the data before the statistical testing. Results were first analyzed by means of a mixed-design ANOVA for repeated measures in order to see the significance of main and interaction effects.

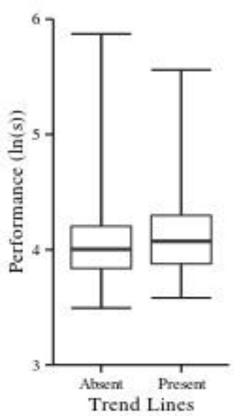


Figure 13 Mean performance in logarithmic scale for *trend lines condition*.

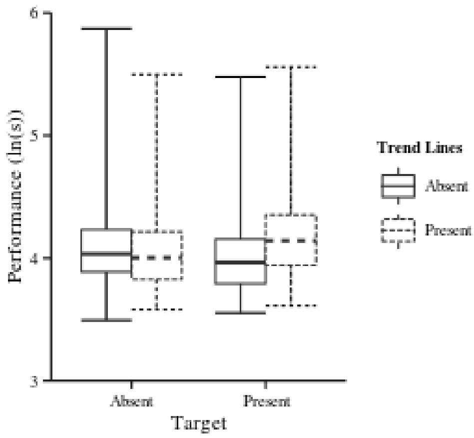


Figure 14 Mean performance in logarithmic scale for *trend lines* and *target conditions*

Table 4 Mixed-design ANOVA Results for Performance

	Df	F value	Pr(>F)
Trial	1	45.966	1.84e-11 ***
Block	1	36.003	2.57e-09 ***
Trend Lines Condition	1	11.496	0.000719 ***
Target Condition	1	0.842	0.359065
Trial: Block	1	11.330	0.000785 ***
Trial : Trend Lines Condition	1	0.209	0.648013
Block: Trend Lines Condition	1	0.712	0.399067
Trial: Target Condition	1	2.516	0.112922
Block: Target Condition	1	0.655	0.418351
Trend Lines Condition: Target Condition	1	38.163	8.77e-10 ***
Trial: Block: Trend Lines Condition	1	6.056	0.013994 *
Trial: Block: Target Condition	1	0.452	0.501298
Trial: Trend Lines Condition: Target Condition	1	3.455	0.063301 .
Block: Trend Lines Condition: Target Condition	1	0.302	0.582559
Trial: Block: Trend Lines Condition: Target Condition	1	2.895	0.089128 .
Residuals	1260		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The analysis was conducted with *performance time* as the dependent variable and *trend lines* and *target* conditions as between subject independent variables, as well as *trial* and *block* as within-subject independent variables. A decision criterion of 5% was used in the analysis.

As shown in Table 4, the results obtained from the mixed design ANOVA for repeated measures revealed a statistically significant main effect on *trial* (F=45.966, $p < .001$) and a statistically significant main effect on *block* (F=36.003, $p < .001$). There is also a statistically significant interaction effect between *trial* and *block* (F=11.330, $p < .001$). These effects are indeed considered as trivial, simply because it is obvious that the participants learn more in block 1 than in block 2.

There is a significant main effect of *trend lines condition* (F=11.496, $p < .0001$). This is illustrated in Figure 13. There is no significant main effect on the *target condition*.

The analysis revealed a statistically significant interaction effect between trend lines condition and target condition (F=38.163, p<0.0001). This effect can be observed in Figure 14.

Scheffé post-hoc comparisons showed that performance times were larger in the group *with trend lines* and *with target* (M=69.12s, SD=28.34) compared to the group *with trend lines* and *without target* (M=61.66s, SD=26.44) with a statistical significance at the .001 level. *Performance time means* were also different between the groups *without trend lines* and *with target* (M=58.86s, SD=23.52) and *without trend lines* and *without target* (M=65.33s, SD=33.05).

Table 5 Scheffé Analysis Results

	With Trend Lines : Without Target (TL: ¬T)	With Trend Lines : With Target (TL:T)	Without Trend Lines : Without Target (¬TL: ¬T)
With Trend Lines : With Target (TL:T)	.115***		
Without Trend Lines : Without Target (¬TL: ¬T)	.045	-.070	
Without Trend Lines : With Target (¬TL:T)	-.040	-.154***	-.085*

* .05 ** .01 *** .001

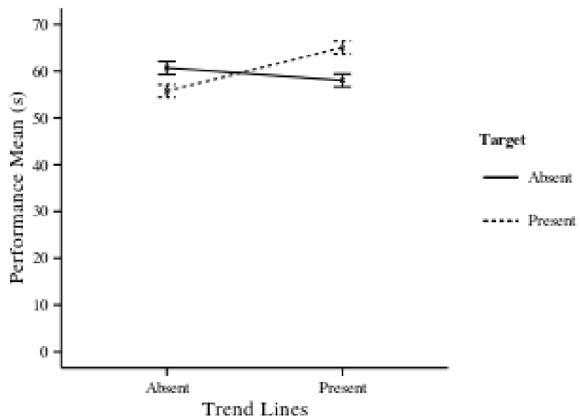


Figure 15 The changes in mean performance in seconds for *trend lines* and *target conditions*

The statistically significant interaction effect between *trend lines condition* and *target condition* is illustrated in Figure 15 where *performance time means* in seconds are visible.

Moreover, as shown in Table 4, there is a statistically significant interaction effect among *trial*, *block* and *trend lines condition* (F=6.056, p<0.01). Two regression analyses were carried out in order to focus on the relation between these variables. One for each of the two *trend lines conditions*. The regression analyses were carried out for *block 2* data with *performance time* as dependent variable and *trial* as independent variable. Results of the regression analyses are presented in Table 6 and 7.

Table 6 Without Trend Lines Condition Regression Analysis

Coefficients	Estimate Std.	Error	t value	Pr(>|t|)
(Intercept)	3.999265	0.032810	121.890	<2e-16 ***
trial	-0.003994	0.002739	-1.458	0.146

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2825 on 318 degrees of freedom
Multiple R-squared: 0.006641, Adjusted R-squared: 0.003517
F-statistic: 2.126 on 1 and 318 DF, p-value: 0.1458

Table 7 With Trend Lines Condition Regression Analysis

Coefficients	Estimate Std.	Error	t value	Pr(>|t|)
(Intercept)	4.100529	0.030271	135.460	< 2e-16 ***
trial	-0.009691	0.002527	-3.835	0.000151 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.11 on 318 degrees of freedom
Multiple R-squared: 0.03924, Adjusted R-squared: 0.03622
F-statistic: 12.99 on 1 and 318 DF, p-value: 0.0003638

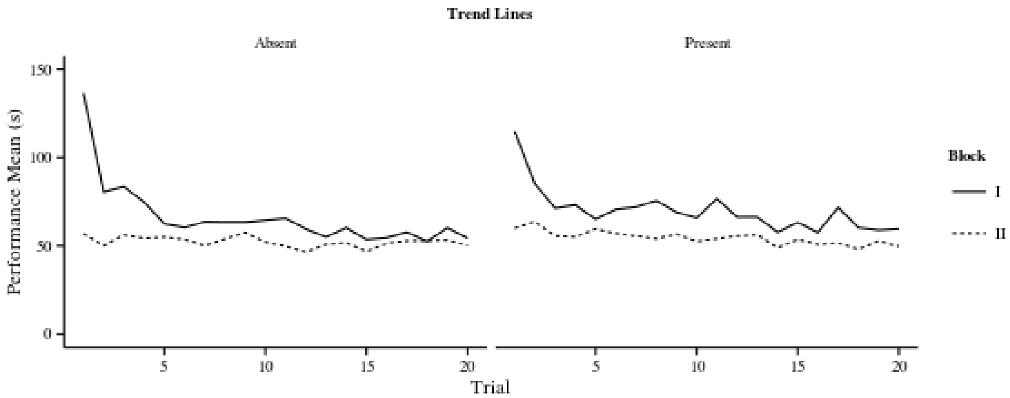


Figure 16 Mean performance in seconds for *trend lines condition*

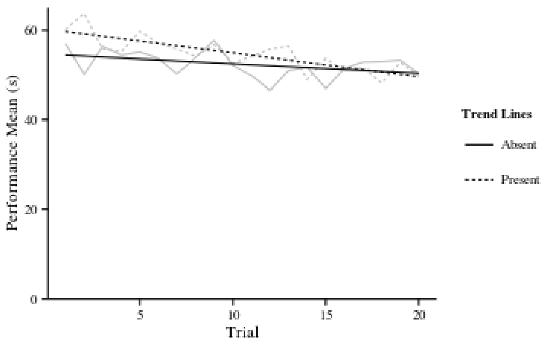


Figure 17 Mean performance for *trend lines condition* in *block 2* regardless of the *target condition*

5.1.2 Perceived Difficulty

The experimental results on the subjective data are shown below. As mentioned in Section 4.7, the perceived difficulty data was collected at three different points during the experiment.

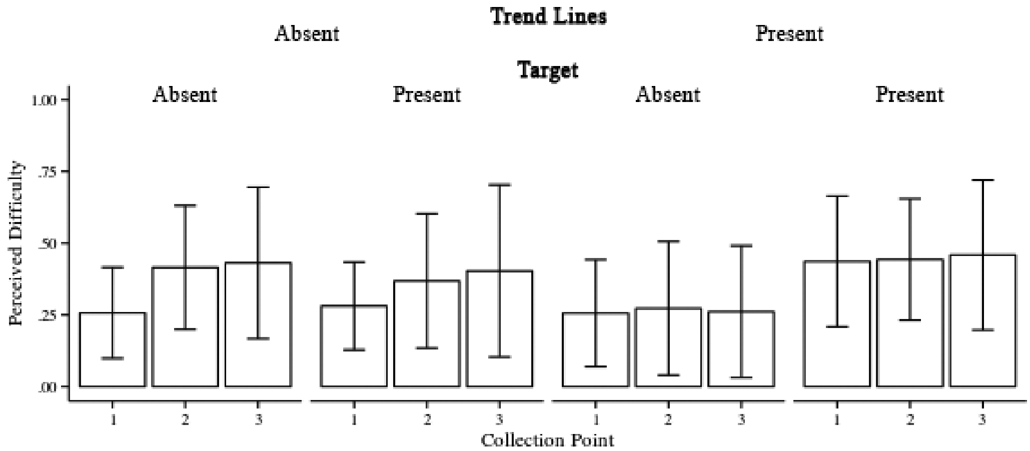


Figure 18 Mean perceived difficulty for the four different conditions.

Figure 18 depicts the mean perceived difficulty ratings collected for four different conditions. The perceived difficulty was analyzed by means of a mixed-design ANOVA for repeated measures.

Table 8 illustrates that for the perceived difficulty there is a significant interaction between the trend line condition and the target condition (F=5.440, p<0.05). This effect can be seen in Figure 19 and Figure 20. There are no significant main effects or interaction effects between other variables.

Table 8 Mixed-design ANOVA Results for Perceived Difficulty

	Df	F value	Pr(>F)
Collection Point	1	0.563	0.4552
Trend Lines Condition	1	0.001	0.9786
Target Condition	1	3.321	0.0719 .
Collection Point: Trend Lines Condition	1	1.504	0.2234
Collection Point: Target Condition	1	0.050	0.8244
Trend Lines Condition: Target Condition	1	5.440	0.0220 *
Collection Point: Trend Lines Condition: Target Condition	1	0.115	0.7359
Residuals	86		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

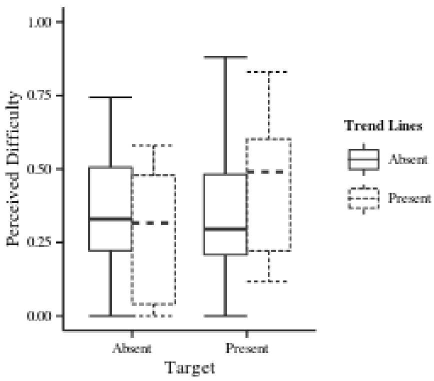


Figure 19 Perceived difficulty values for four different conditions.

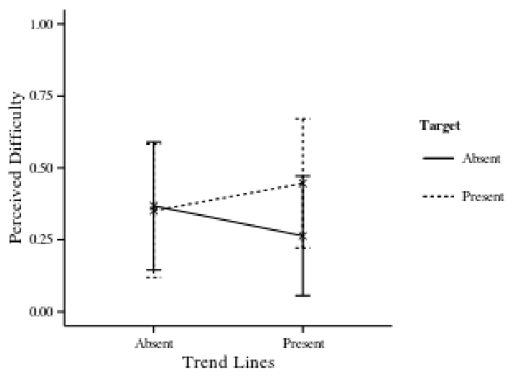


Figure 20 The Relation between perceived difficulty, *trend lines condition* and *target condition*.

5.2 Interviews

After the test, the participants were asked to attend a short post-experimental interview with six questions, conducted in order to understand their representation of the problem, their problem solving approaches, the number of different strategies they tried and whether they can articulate what they have been thinking or whether their reasoning were biased. The interview investigated the perceptive and behavioral aspects among the participants as was directed in our second and third research questions, as well as giving us an insight for our future studies. Additionally, we tried to understand the usage of the trend lines by the participants in order to make sense of our statistical testing results.

After the warm up question, when asked how they would describe this task to someone else, all of the respondents could verbalize the task they were supposed to complete and the goal of the game in accordance with the condition they were subjected to. This ensures that the problem was well-defined and could have been transferred to the participants successfully.

In order to understand more about the general thoughts about their experience with the experiment and their approaches to the game, the respondents were then asked to indicate how they would recommend a novice player to approach the game. In response to this question, the majority of the respondents commented on their own strategies that they thought worked well, as well as criticizing themselves. Explaining what they did right or what they should have done instead, almost half of the participants reported that they figured out the importance of the slowest trains, though some of them said that they did not know how to make use of this idea.

In response to the questions whether they changed strategies at any point during the game, and if they did, how many times they changed strategies approximately, most of the participants expressed a number between 3 and 5 regardless of the condition they took the test under. I believe that this is because of the misinterpretation of the word *strategy*. As one interviewee put it, he did not really have a detailed strategy from beginning to the end, but rather came to conclusions like: “the more the trains on the track at a time the fastest the problem is solved”. However, most other participants explained their strategies in detail, pointing out in which order they initiated a particular train and/or how fast he moved them. None, however, came close to identifying the ultimate solution pattern, previously calculated to define the target number 32.

One of the most important questions directed only to the respondents who had seen the trend lines in the interface required them to give information on whether they made use of the trend lines somehow, when performing the task, and in that case, how they used them. 95% of the respondents reported that they used the trend lines and the trend lines had been really helpful to see where the trains would meet by looking at the intersection points of the trend lines. The respondents also reported that they used the trend lines in order to avoid collision by checking the alignment of two trend lines to see whether the trains had the same speed when they were going in the same direction. One interviewee even said: “I actually don't think that the game would have been possible without the lines, because it would have been too frustrating and inaccurate to have to estimate the meeting point without any additional help. “ However, the participants who claimed they did not use the trend lines or that they used it only for some time in the beginning or for instance, only during the last 30 minutes, referred to the necessity of acting quickly in the game. As one interviewee commented: “for me the lines introduced a bit too much when I had to do something quick so I decided to just check at which location two trains would reach each other and after that I ignored the lines” or again about this issue another participant said: “Yep, I used it, but not so much, because it is impossible to pause the process. I used them in order to predict the first 2~3 trains“.

Chapter 6 Analysis

In this experiment, our hypothesis based on the main research question was tested and the answers to our sub-questions were explored. Based on our statistical analysis, the general finding in the experiment is, perhaps surprisingly, that our hypothesis was falsified. The regression analysis revealed that regardless of the target condition, the presence of trend lines did not really accelerate learning nor improved the performance; on the contrary, although the trend lines maintained some kind of learning among the users for 40 trials, we can say that in a way the lines slowed down their learning. Thus, our results point to that the trend lines do have an effect on performance in conjunction with time, yet we cannot really conclude whether they would affect positively or a negatively in the long run. (See Figure 16 and 17)

As can be seen in Figure 16, where *block 1* stands for the first 20 trials and *block 2* stands for the last 20 trials, it is clearly illustrated that the participants who were under the *without trend lines* condition almost stopped improving their performances in *block 2*, whereas the participants who were under the *with trend lines* condition kept slightly improving their performances. The performances have improved as a power function of trial and have shown modest declines over the two blocks of intervals. However, as depicted in Figure 17, the participants who played the game under the condition *with trend lines* start *block 2* with a higher *mean performance time* than the participants who did not see the *trend lines*, implying that their performances were still worse than the other group, though until the end by improving their performances at a slower rate, they finally managed to converge the other group's performance levels.

As also shown in Figure 13, according to the graph the data has shown a similar distribution for both conditions (*with trend lines* or *without trend lines*), indicating that the samples of participants are roughly comparable with respect to performance distribution. On the other hand, the same graph depicts that the box plot drawn for the condition *with trend lines*, has a wider upper quartile and a slightly higher median than the box plot drawn for the condition *without trend lines*. Therefore, it is noticeable that the participants under the condition *with trend lines* exhibited a slightly higher *mean performance time*.

Our second research question was whether the introduction of a target would affect the user behavior or not. As illustrated in Figure 12, for the *without target* condition the collected data has shown that there has been a steep decline in the mean time spent for each trial during the first quarter of *block 1*, meaning that the participants' performance has shown a sharp start and a rapid progress in the first few trials. On the other hand, in contrast to *block 1*, during *block 2* the *mean performance time* in seconds seemed to remain steady. Also, again for the participants under the *without target* condition, the absence or presence of trend lines exhibited a slightly similar trajectory during the experiment including both *block 1* and *2*.

For the *with target condition*, Figure 12 reveals that compared to the *without target* condition the participants showed lower initial data points in seconds in the first quarter of *block 1*, meaning that they found faster solutions in their first few trials than the participants under *with target condition*. Unlike the participants who performed under the *without target condition*, there has only been a gradual improvement in their performances, and this continued during both blocks. However, in terms of trajectories for the *mean performance time*, there has been a slight difference in slope between *block 1* and *block 2*. In *block 2* the improvement in performance slowed down, but didn't stop. Interestingly, this trajectory of slight improvement remained and the mean performance has been observed to keep improving in time regardless of the existence of trend lines. On the top of it, surprisingly but coherently, for the *with target condition*, the participants who have seen the trend lines in the interface have generated lower performance values during both blocks than the participant who have not seen the trend lines in the interface.

Furthermore, as can be seen in Figure 14, the target condition seems to have had an impact on performance. Although for the *without target* condition there is no major change in performance within different trend line conditions, for the *with target* condition, where the participants were aware of an exact number of seconds (32 sec.) to aim at, they have performed significantly worse under the condition in which the trend lines were visible in the interface. Thus, the presence of trend lines has had a slightly negative impact on performance for the participants who played the game under the *with target* condition.

Scheffé post-hoc comparison has shown that, as also can be seen from Table 5, the mean time spent in a trial is remarkably longer in condition with trend lines and with target than in condition without trend lines and with target. This indicates that when the participants are introduced with a target, counterintuitively, the trend lines do not really seem to help the participants to solve the puzzle quicker, but instead they seem to be affecting their results in a negative way and causes the mean time spent on a trial to be higher.

Figure 15 presents the significant differences among the conditions with respect to mean performance. When a target is introduced to the participants, the visualization of trend lines led to significantly lower performance results, than the condition in which the trend lines were not visualized. For *without target* condition the trend lines had no significant effect. Thus, the single most striking observation to emerge from the data comparison was that when the participants were told to aim at a specific number, hiding the trend lines assisted the participants to perform significantly higher than the participants under the *without target* condition in which they were only told to solve the problem as fast as possible.

Figure 19 provides a revealing summary of the data collected to investigate how perceived difficulty was affected. As expected, participants hold quite different opinions about difficulty. Under the *without target* condition, although the medians are almost the same, the distributions are different. The ranges indicate that, overall, there is more variation in the condition with trend lines than the condition without trend lines. Comprehensibly, for the *with target* condition the medians indicate that a typical perceived difficulty value for the participants under the condition with trend lines is noticeably larger than a typical value for participants who were under the condition *without trend lines*. As also was depicted in Figure 20, the difficulty was significantly perceived harder than all the other three conditions when the trend lines are visualized and the target was introduced.

Our third research question was how the absence or presence of trend lines and the introduction of a target affected the perceived difficulty. Figure 20 presents the mean results for perceived difficulty for the four different conditions. A significant difference in terms of perceived difficulty appears between condition *with trend lines* and *without target* and condition *with trend lines* and *with target*. When the trend lines are visible to the participants, the determining factor seems to be the *target condition*. In coherence with the mean performance results, the participants under the condition, in which the trend lines were visible and a target was introduced, perceived the task relatively harder than the participants under the condition in which the trend lines were visible but the goal introduced to the participants was only defined in general as solving the problem as fast as possible.

Chapter 7 Discussion and Conclusion

7.1 Discussion

Based on the earlier discussion about distributed cognition, we thought that the presence of trend lines might transform the participants' representation of the problem and inspire them to develop novel solutions by transforming their cognitive tasks into relatively easier ones. Thereby, we believed that the presence of trend lines would have improved the task performance and accelerated learning. However, at least for the first 40 trials, our findings have revealed that the presence of trend lines does not improve the performance and slows down learning. It has been shown that in *block 2* the performance only gradually improves. This effect is also salient when the users are introduced to a target and our results on perceived difficulty are consistent with this. On the other hand, in the last few trials the participants who were subjected to the condition *with trend lines* appeared to slightly outperform the participants who were subjected to the condition *without trend lines* (see Figure 17). The collected data is not sufficient to conclude that this trend would continue, but these results thus further support the idea of a new hypothesis: *In the long run, the presence of trend lines will increase maximum performance.*

Larkin, McDermott, Simon and Simon (1980) using physics problems, found that the strategies used by expert and novice problem solvers differed. Since novices are not able to recognize and memorize problem configurations and are forced to use general problem-solving strategies such as means-ends analysis (Sweller, 1988), unlike experts for novice problem solvers problem solving requires large cognitive processes. Sweller (1988) states that some forms of problem-solving search such as means-ends analysis interfere with learning. He claims that there are two potential reasons; these are cognitive processing capacity and selective attention.

There might be more than one reason why the presence of trend lines did not improve the performance and slowed down learning for the first 40 trials. First of all, we know that the participants take their time to generate at least one solution to the problem in their first attempts. As they keep playing they try the same strategy more than once. During the first few trials they learn these sequences of moves that take them to the goal state. Once they learned the sequence of moves, they recall it in their subsequent trials. As they perform they probably keep thinking about other potentially successive states to modify the strategy or try to think about totally different solutions that would take them to the goal state faster. This activity brings a high cognitive load and it is probably the same reason why some participants take their time in between the trials to think more about other potential strategies.

Moreover, we know that learning takes place as the participants play the game. First they learn the dynamics of the microworld. Then, tactical and strategic learning takes place. The participants start to memorize the sequences for already found strategies and try to apply them. However in order to apply their strategies they must also improve their motor skills. Some participants, who were immediately warned, were observed in a tendency to try out the touchpad instead of the mouse. And some participants verbalized the difficulty of the controllers. The participants couldn't make their plans come to life immediately since they had to improve their motor skills to solve the problem. At the same time, they keep thinking about potential other solutions. So their attention is divided into two that creates a negative effect on learning.

Another explanation for the reason why participants who were subjected to trend lines have shown continuing but rather modest improvement in terms of performance might be an indirect result of the cognitive effort that they have been unconsciously directing towards *pattern learning*. For instance, imagine an instant with 4-5 trains on the track. The trend lines on the screen would generate a grid shaped colored pattern. Some parts of the participants' working memory might have been occupied by this activity that brings extra cognitive load and divides their attention. Since they learn the patterns, they would perhaps present a better maximum performance in the long run.

Sweller and Levine (1982) using maze problems did some experiments about the effects of giving participants a specific or a non-specific goal. In the non-specific goal case, "it is not possible to reduce differences between a given problem state and the goal state because the goal state is not known until it is attained" (Sweller, 1988). Therefore, for the novice users, difference reduction method or means-ends analysis becomes remarkably harder to apply. During the pilot study, one participant who took the test under the *without target* condition was observed to initiate one slowest train from one side, and wait until it reaches the opposite end as he counted in the mean time to see how much the speed of the slowest trains did limit the trial completion time in the best case scenario. The participant probably thought of calculating a reasonable target to achieve for himself. So the participants who didn't have a target number in mind might have kept thinking whether they were doing well or bad. However, despite this might have hindered finding novel solutions to the problem, its effects on performance are not visible. In contrast, our findings have shown that when the trend lines were presented the participants who were introduced to a target have shown noticeably slower performances than the participants under the three other conditions. The introduction to a target might have led them to put more effort on discovering successive solutions to the problem, and this might have brought extra cognitive load along with the presence of trend lines.

Furthermore, Sweller's results were based on physics or maze problems. In our experiment, I believe the dynamicity of the microworld brought extra complexity, which might have negatively affected the performances. The interview results have also shown that some participants found it hard to keep up with the game pace. Additionally, one other interesting finding was about the constrained representations of the problem that the dynamics of the microworld might have caused among the participants. Participants, supposedly using their deductive reasoning, might have generated a wrong mental model, leading them to conclude that it was impossible to move 3 trains through one side track at a time, despite it was possible. This feature of the game that hinders one of the best solutions, in which 3 trains can move in one side track but not stand on it, is a good example of what Ohlsson (as cited in Eysenck & Keane, 2005) showed in 1999 in the nine-dot problem. “He claimed that individuals who fail to solve the nine-dot problem do so because their initial representation of the problem is too constrained (i.e., is based on the assumption that they shouldn’t go outside the boundaries of the square)” (Eysenck & Keane, 2005). Because of the participants’ first few trials to place the trains in the same track failed, their representation of the problem became constrained.

7.2 Conclusion

The purpose of this study was to see how performance and perceived difficulty of novices would be affected by the absence or presence of trend lines and introduction of a target. A between subjects study was conducted. 32 participants tried to solve a logical problem in a microworld. They had 40 trials. In summary, we may conclude that the presence of trend lines do not improve performance and slow down learning among the novice users for 40 trials. Experimental evidence also confirms that when trend lines are visible to the participants, introduction of a target may impose a heavy cognitive load and cause the users to perceive the task harder.

7.3 Future work

First of all, in this study the strategies used by the participants were only roughly observed, and couldn’t be reflected much in our discussions. For that reason, the main data we put into account to comment on the strategies used by the participants were the interviews which come along with other issues about the ability of users to verbalise their solutions. If we had more time, an algorithm would have been developed in order to better analyze and investigate the decisions made by the participants during the experiment as they interacted with the software. In our future studies, we are planning to write a script that could recognize different strategies, by for instance counting the number of clicks, holding whether the trains were moved below or above a threshold in terms of speed, and the order in which the trains were initiated etc.

Secondly, based on what has been learned in the project, we may modify our game to better control the cognitive load among users. In this experiment, the pace of the game might have brought some unintended extra complexity to the participants. It would be especially interesting to explore if we would get different results when we decreased the pace of the game.

Moreover, a point that was not discussed in this thesis was the relation of performance and long periods of practice. Therefore, such a study that takes two or three days of practice, with 2 blocks of gameplay per day would be interesting to conduct.

Additionally, before we move on to scenarios with disruptions, looking at the effects of visualising history as a feedback, that is to present the previous moves of the participants in the interface, would be interesting.

During the evaluations in Boden, when deploying STEG into a new traffic center, the importance of training have been experienced (Tschirner, 2015). The results of our studies can help improve STEG, or open possibilities to work with the training departments of Trafikverket, perhaps based on our findings by starting a project on designing a separate STEG interface for training purposes only.

A new project to develop and deploy STEG system for complex stations is already in progress (Tschirner, 2015). When STEG will be introduced nationwide starting from late 2015, the results of such studies would assist the design of the user training and give insight on how STEG could be integrated in different train traffic control centers.

References

Anderson, J. (2010). Cognitive psychology and its implications (7th ed.). New York: Worth.

Andersson, A. W., Sandblad, B., Hellström, P., Frej, I., & Gideon, A. (1997, November). A systems analysis approach to modelling train traffic control. In*Proceedings of WCRR* (Vol. 97, pp. 16-19).

Andersson, A. W., Jansson, A., Sandblad, B., & Tschirner, S. (2014). Recognizing complexity: Visualization for skilled professionals in complex work situations. In Building Bridges: HCI, Visualization, and Non-formal Modeling(pp. 47-66). Springer Berlin Heidelberg.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological review, 86*(2), 124.

Baddeley, A. D. (2001). Is working memory still working?. *American Psychologist*, *56*(11), 851.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta psychologica*, *81*(3), 211-241.

Brehmer*, B. (2005). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science*, *6*(1), 73-93.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2), 171-184.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55-81.

Evans, J. S. B. (1989). *Bias in human reasoning: Causes and consequences*. Lawrence Erlbaum Associates, Inc.

Eysenck, M. W., & Keane, M. T., 1961. (2005). Cognitive psychology: A student's handbook. Hove: Psychology Press.

Gigerenzer, G. (2008). *Gut feelings: Short cuts to better decision making*. Penguin Uk.

Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013). A socio-technical comparison of rail traffic control between GB and sweden.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept.*Cognitive Psychology*, *6*(2), 270-292.

Halverson, C. A. (1994). Distributed Cognition as a theoretical framework for HCI: Don't throw the Baby out with the bathwater-the importance of the cursor in Air Traffic Control. *Cognitive Science Department, UCSD. Report*, *9403*.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174-196.

Hutchins, E. (1990). The technology of team navigation. *Intellectual teamwork: Social and techno-*

logical foundations of cooperative work, *1*, 191-220.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive science*,*19*(3), 265-288.

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science*.

Jeffries, R., Polson, P. G., Razran, L., & Atwood, M. E. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology, 9*(4), 412-440.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S., & Passingham, R. E. (1994). Motor sequence learning: a study with positron emission tomography. *The Journal of Neuroscience*, *14*(6), 3775-3790.

Kahneman, D., 1934. (2011). Thinking, fast and slow. New York; London: Farrar, Straus and Giroux.

Karat, J. (1997). Evolving the scope of user-centered design. *Communications of the ACM*, *40*(7), 33-38.

Kauppi, A., Wikström, J., Sandblad, B., & Andersson, A. W. (2006). Future train traffic control: control by re-planning. Cognition, Technology & Work, 8(1), 50-56.

Klein, G. A. (1999). *Sources of power: How people make decisions*. MIT press.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of Competence in Solving Physics Problems*. *Cognitive science*, *4*(4), 317-345.

Lo, J., Van den Hoogen, J., & Meijer, S. (2013, December). Using gaming simulation experiments to test railway innovations: implications for validity. InProceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World (pp. 1766-1777). IEEE Press.

Lovett, M. C., & Anderson, J. R. (1996). History of success and current context in problem solving: Combined influences on operator selection. *Cognitive psychology, 31(2)*, 168-217.

Newell, A., & Simon, H. A., 1916-2001. (1972). Human problem solving. Englewood Cliffs, N.J: Prentice-Hall.

Norman, D. A., & Draper, S. W. (1986). User centered system design.*Hillsdale, NJ*.

Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. Basic Books.

Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, *27*(3), 303-316.

Rogers, Y. (1997). A brief introduction to distributed cognition. *Retrieved July*,*24*, 1997.

Rosenbaum, D. A., Carlson, R. A., & Gilmore, R. O. (2001). Acquisition of intellectual and perceptual-motor skills. *Annual review of psychology*, *52*(1), 453-470.

Sandblad, B., Andersson, A. W., Frej, I., & Gideon, A. (1997). *The role of human-computer interaction in design of new train traffic control systems*. Universitetet.

Sandblad, B., Andersson, A. W., Kauppi, A., & Wikström, J. (2007). Implementation of a test system for evaluation of new concepts in rail traffic planning and control. JR Wilson, B. Norris, T. Clarke, & Mills, A.(Eds.), People and Rail Systems: Human Factors at the Heart of the Railway, 301-307.

Sandblad, B., Andersson, A. W., Kauppi, A., & Isaksson-Lutteman, G. (2010). Development and implementation of new principles and systems for train traffic control in Sweden. Computers in Railways, 12, 441-450.

Stuhler, E. A., & DeTombe, D. J. (Eds.). (1999). *Complex problem solving: cognitive psychological issues and environment policy applications* (Vol. 5). Rainer Hampp Verlag.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning.*Cognitive science*, *12*(2), 257-285.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means–ends analysis and learning. *Journal of experimental psychology: Learning, memory, and cognition*, *8*(5), 463.

Thomas, J. C. (1974). An analysis of behavior in the hobbits-orcs problem.*Cognitive Psychology*, *6*(2), 257-269.

Tschirner, S., Sandblad, B., & Andersson, A. W. (2014). Solutions to the problem of inconsistent plans in railway traffic operation. *Journal of Rail Transport Planning & Management*, *4*(4), 87-97.

Tschirner, S. (2015). The GMOC Model: Supporting Development of Systems for Human Control.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453-458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), 293.

Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press.

Von Geijer, Erik. (2014) Trafikverkets trafikledning historie och nulägesrapport. Borlänge: Trafikverket.

Appendix A

Background Questionnaire

Please fill out the details below:

What year are you born? _______

What is your gender?

- Female
- Male

What is the highest degree or level of school you have completed?

- Primary School
- Secondary School
- Higher Education (Bachelor or Master)
- Post graduate education (Ph.D. or equivalent)

How many hours a week do you spend playing games or solving puzzles (including computer games, mobile phone games, board games, and newspaper puzzles like Sudoku etc.)?

________ hrs

How much programming experience do you have?

- None
- Less than six months
- Six months to one year
- More than one year

In which programme are you currently enrolled?

____

At what year are you in your current programme?

____

Appendix B

Consent Form - Research Project on Visual Information and Interaction

Fair Processing Statement

Information collected as part of this research will be used to inform the development of decision support systems, undertaken by the Uppsala University. The information collected as part of this study will be retained for scientific research purposes and will only be made accessible to authorized personnel who are involved in the project. Personal and identifiable information will be processed and securely stored by the Uppsala University in accordance with the provisions of the PuL 1998:204.

Statements of Understanding and Consent

- I understand that my participation is voluntary and that I am free to withdraw at any time during the study without giving any reason. If I withdraw my data will be removed and destroyed.
- I understand that if I wish to withdraw from the study after taking part, I must explicitly request this by emailing Sercan Caglarca (sercancaglarca@gmail.com) within seven days of the study being completed. If I withdraw my data will be removed and destroyed.
- I agree to be voice-recorded during the interview.
- I agree that my gameplay on the computer screen is recorded during the study.
- I understand that the data and recordings collected in this study will be treated as confidential. Any recording will be stored securely. The identity of the participants will remain anonymous in the study outcomes, and only relevant research team members and collaborators working on the project will be granted access to these recordings for legitimate research purposes.
- I understand that my personal data will be processed for the purposes detailed above, in accordance with the PuL 1998:204.
- Based upon the above, I agree to take part in this study.

Name of Participant .... Date...

Signature ....

Additionally, by signing below I agree that data recorded during my participation in the study can be used in scientific papers, conferences and events

Signature .... Date...

...
...

Researcher/ Witness

A copy of the signed and dated consent form should be given to the participant and the original should be retained by the researcher to be kept securely on file.

Name of Researcher .... Date...

Signature ....

Appendix C

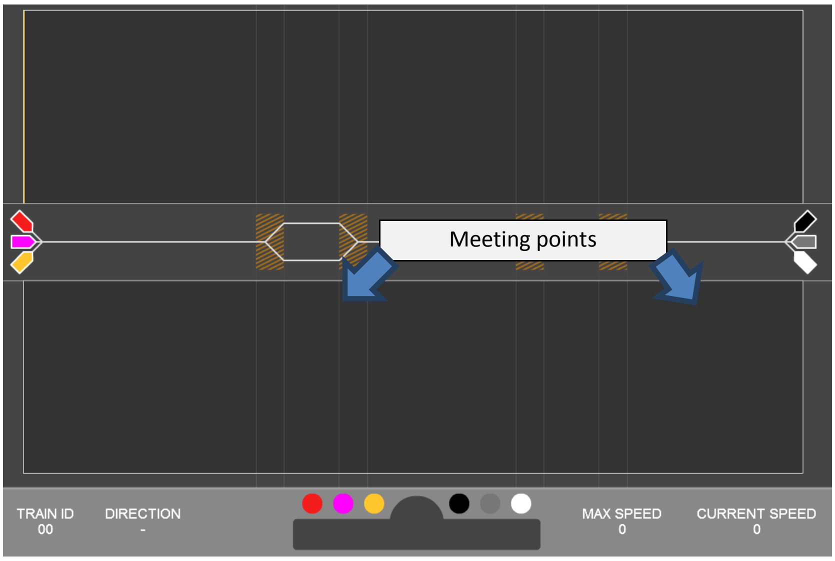
Instructions Sheet

Welcome!

You will play a game with the aim of solving a logistic problem as fast as possible and you will in total be given 60 minutes to improve your performance. There are no 'lives' and you will not 'die' in the game. ***(Participants under with target condition saw this extra sentence here: "It is possible to solve the problem in less than 32 seconds. Please aim at a score that is at least as slow as 32 seconds.")***

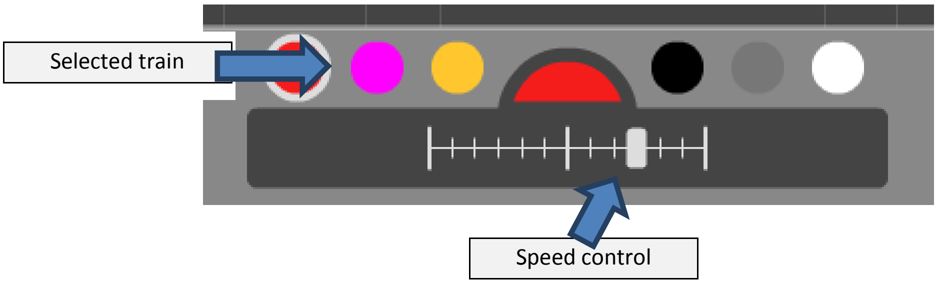
The game consists of a simulated railway track and six trains, three at each end of the track. Your task is to drive each train to the opposite end of the track compared to where the train started.

This is what the railway track and the trains look like as you start the game:



If two trains meet along the track they will simply stop. They cannot collide and break. There are two designated meeting points along the track as you can see in the image above. Trains can only meet where the two tracks of the meeting points are parallel. Trains cannot meet in the striped areas at the two ends of the meeting points.

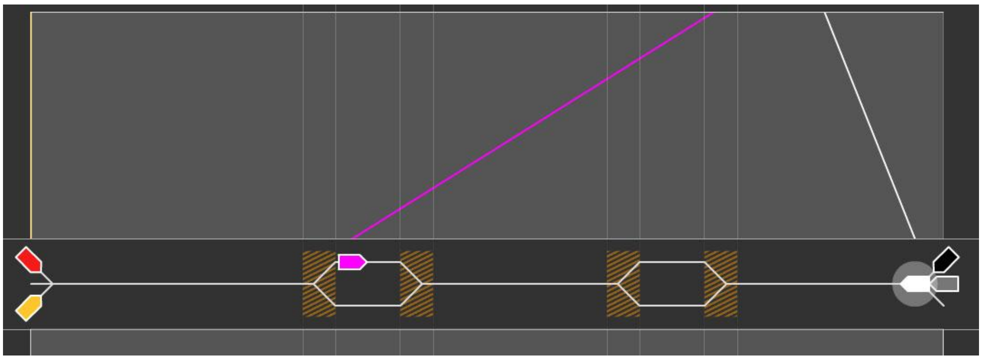
You control the trains in the lower panel of the window. Please note: You CANNOT click directly on a train to select it. You select a train by clicking on the correspondingly coloured circle in the lower panel. The selected colour will be highlighted and also displayed in the speed control. You set the speed of the selected train by clicking in the speed control, see image below.



If you want the selected train to move forward (in the direction of its pointed end) you click somewhere to the right of the centre in the speed control; if you want it to move backwards, you click somewhere to the left of the centre. A click in the middle of the speed control stops the train.

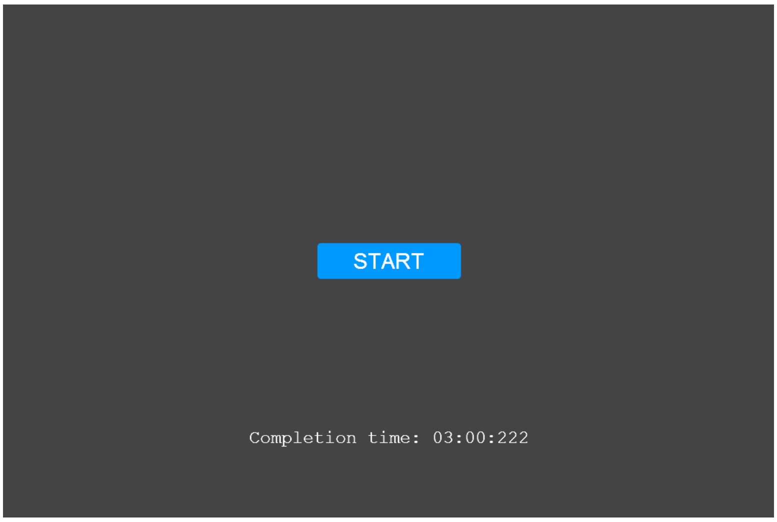
The game is over when you have managed to move the three grey-scale trains to the left end of the track and the coloured trains to the right end of the track.

(Participants under the condition with trend lines saw this extra part here: "In the top of the game screen you will see lines representing a prognosis of the trains' future horizontal positions.



")

When the game ends, you will see the time you spent and will be given a possibility to start a new game by clicking on the start button. The timer starts counting once you select the first train.



If you have any questions, please ask the test leader now.

Good luck!