UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1237

The GMOC Model

Supporting Development of Systems for Human Control

SIMON TSCHIRNER

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2015

ISSN 1651-6214
ISBN 978-91-554-9192-5
urn:nbn:se:uu:diva-246719

Dissertation presented at Uppsala University to be publicly examined in Room 2446, ITC, Lägerhyddsvägen 2, Uppsala, Monday, 4 May 2015 at 10:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Professor Sebastiaan Meijer (KTH School of Technology and Health).

Abstract

Tschirner, S. 2015. The GMOC Model. Supporting Development of Systems for Human Control. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1237. 158 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9192-5.

Train traffic control is a complex task in a dynamic environment. Different actors have to cooperate to meet strong requirements regarding safety, punctuality, capacity utilization, energy consumption, and more. The GMOC model has been developed and utilized in a number of studies in several different areas. This thesis describes GMOC and uses train traffic control as the application area for evaluating its utility.

The GMOC model has its origin in control theory and relates to concepts of dynamic decision making. Human operators in complex, dynamic control environments must have clear goals, reflecting states to reach or to keep a system in. Mental models contain the operator's knowledge about the task, the process, and the control environment. Systems have to provide observability, means for the operator to observe the system's states and dynamics, and controllability, allowing the operators to influence the system's states. GMOC allows us to constructively describe complex environments, focusing on all relevant parts. It can be utilized in user-centred system design to analyse existing systems, and design and evaluate future control systems.

Our application of GMOC shows that automation providing clear observability and sufficient controllability is seen as transparent and most helpful. GMOC also helps us to argue for visualization that rather displays the whole complexity of a process than tries to hide it.

Our studies in train traffic control show that GMOC is useful to analyse complex work situations. We identified the need to introduce a new control strategy improving the traffic plan by supporting planning ahead. Using GMOC, we designed STEG, an interface implementing this strategy. Improvements that have been done to observability helped the operators to develop more adequate mental models, reducing use of cognitive capacity but increasing precision of the operative traffic plans. In order to improve the traffic controllers' controllability, one needs to introduce and share a real-time traffic plan, and provide the train drivers with up-to-date information on the surrounding traffic. Our studies indicate that driver advisory systems, including such information, reduce the need for traffic re-planning, improve energy consumption, and increase quality and capacity of train traffic.

Keywords: Human-Computer Interaction, User-Centered Design, Human Control, Train Traffic Control, Usability, System Development, Human Work, Railway Traffic, Rail Human Factors, Mental Models, Situation Awareness

Simon Tschirner, Department of Information Technology, Division of Visual Information and Interaction, Box 337, Uppsala University, SE-751 05 Uppsala, Sweden.

ISSN 1651-6214
ISBN 978-91-554-9192-5
urn:nbn:se:uu:diva-246719 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-246719)

To Li and Elli

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **GMOC: A Conceptual Model for Human Control of Dynamic Systems, Applied to Train Traffic Control**
Simon Tschirner, Anders Jansson, Bengt Sandblad, Arne W Andersson. *Submitted for Journal Publication*. 2015.

II **Recognizing Complexity: Visualization for Skilled Professionals in Complex Work Situations**
Arne W Andersson, Anders Jansson, Bengt Sandblad, Simon Tschirner. *Building Bridges: HCI, Visualization, and Non-Formal Modeling*. Springer, 2014. pp. 47–66.

III **Solutions to the problem of inconsistent plans in railway traffic operation**
Simon Tschirner, Bengt Sandblad, Arne W Andersson. *Journal of Rail Transport Planning & Management* 4, 2014. pp. 87–97.

IV **Analysis of Collaboration Applied to Train Drivers and Train Traffic Controllers in Sweden**
Simon Tschirner, Anders Jansson, Bengt Sandblad, Arne W Andersson, Peter Hellström, Gunnika Isaksson-Lutteman. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*. Taylor & Francis, 2013. pp. 389–398.

V **Designing Train Driver Advisory Systems for Situation Awareness**
Simon Tschirner, Arne W Andersson, Bengt Sandblad. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*. Taylor & Francis, 2013. pp. 150–159.

Comments on my Contribution

Paper I

I am the principal author. All authors participated in the development of the GMOC model. The studies described in the latter part of the paper where led by Bengt Sandblad and Arne W. Andersson.

Paper II

I participated in the discussions and development of guidelines for visualization. I was involved in the studies related to train traffic and in the writing.

Paper III

I am the principal author. I was involved in all studies and led the analysis of collaboration between train drivers and traffic controllers, and the development of new concepts for driver advisory systems. Peter Hellström led the analysis of the Swedish train traffic control process.

Paper IV

I am the principal author. I developed the extension to the GMOC model and collected and analysed the data.

Paper V

I am the principal author. I investigated the existing approaches to driver advisory systems, studied the train drivers' work and was one of the process leaders in the vision seminars.

Other Publications

- **Automation in train traffic control**
  B. Sandblad, S. Tschirner, A.W. Andersson. *5th International Conference on Applied Human Factors and Ergonomics,* 2014. (abstract).
- **Authority and level of automation: Lessons to be learned in design of in-vehicle assistance systems**
  A. Jansson, P. Stensson, I. Bodin, A. Axelsson, S. Tschirner. *Human-Computer Interaction. Applications and Services*. Springer International Publishing, 2014. pp. 413–424.
- **Development of Train Driver Advisory Systems**
  S. Tschirner. *First German Workshop on Rail Human Factors*, 2014. (abstract).
- **Improved Railway Service by Shared Traffic Information: Design Concepts for Traffic Control and Driver Advisory Systems**
  S. Tschirner, A.W. Andersson, B. Sandblad. *International Conference on Intelligent Rail Transportation (ICIRT)*, IEEE, 2013.
- **A socio-technical comparison of rail traffic control between GB and Sweden**
  D. Golightly, B. Sandblad, N. Dadashi, A.W. Andersson, S. Tschirner, S. Sharples. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*. Taylor & Francis, 2013. pp. 367–376.
- **Reducing unnecessary cognitive load in train traffic control**
  G. Isaksson-Lutteman, B. Sandblad, A.W. Andersson, S. Tschirner. *FALF*, 2011.
- **Credo methodology: Modeling and analyzing a peer-to-peer system in credo**
  I. Grabe, M.M. Jaghoori, B. Aichernig, C. Baier, T. Blechmann, F. de Boer, . . . W. Yi. *Electronic Notes in Theoretical Computer Science,* 266, 2010. pp. 33–48.
- **Model-based validation of QoS properties of biomedical sensor networks**
  S. Tschirner, L. Xuedong, W. Yi. *Proceedings of the 8th ACM international conference on Embedded software*. ACM, 2008.

Acknowledgements

This thesis, as well as my research and my time in Uppsala, has been influenced and supported by many people. First of all, there are my supervisors, Bengt, who has always been supportive and motivating and Anders, who helped me to look at my research from a different perspective. I would like to thank Arne; it has always been fruitful and a pleasure to discuss with you. Ingela, the head of our division, who in the end played a role that I did not quite expect when our divisions were merged, Mats, who reviewed the final version of my thesis, and Lena, who always helped out with practical matters.

I would like to thank my wife Li and my daughter Elli, who bring so much joy to my life, and my family back in Germany – without their support I would not have been able to start my graduate studies. And I have to thank dfH Felix.

This thesis has directly benefited from my colleagues who were so kind to support me with discussing and reviewing parts of it, or helped out with references and so on. These are "Seminal Lax", Ida, Action Anton, Thomas, and Ted. Martin Joborn was the internal opponent commenting on an early draft of my thesis. I am also thankful for all the interesting discussions, courses, or just nice social events I had together with my other colleagues from the HCI-group, especially Rebecka and Anette, Vi2, and the whole IT department.

I have to mention Pavel, who was extremely helpful and a very good friend when I came to Uppsala, the Bang/Whist group, Pan, Stephan, and Jonas. I probably do not have to mention David.

My research was relying on the cooperation of many train drivers and traffic controllers at SJ, LKAB, Trafikverket, and Green Cargo. I hope that the results of my research will eventually be beneficial to your work environments. I would also like to thank Jörgen Hwargård, Magnus Wahlborg, and Robin Edlund as our main contact persons at Trafikverket. My research was financed by Trafikverket (partly through the KAJT programme) and the On-Time project (FP7-SCP01-GA-2011-285243).

Finally, I would like to thank the two persons who made it possible for me to join Uppsala University. These are my former supervisor Wang Yi and Jorge Marx Gómez, Professor at Oldenburg University, where I pursued my undergraduate education.

Summary in Swedish

Bakgrund

Denna avhandling har fokus på människor som arbetar med styrning av komplexa, dynamiska processer och system. Jag ser detta som ett sociotekniskt system, bestående av en stor mängd komponenter i form av delsystem, delprocesser, organisationer och aktörer. I centrum ligger en dynamisk process som med hjälp av olika stödsystem ska styras så att givna mål uppnås. Att processen är dynamisk betyder att den fortsätter att utvecklas beroende på sin egen dynamik och all påverkan som har skett tidigare. Att kontrollera och styra en process mot ett önskat tillstånd kräver en djup förståelse och hög yrkesskicklighet samt ett bra samarbete mot gemensamma mål mellan de olika aktörerna. När vi utvecklar stödsystem för operatörerna, aktörerna i processen, måste vi ta hänsyn till den komplexitet som finns i deras uppgifter och miljöer. Vi måste ge människorna det bästa möjliga stöd för att förstå och hantera komplexiteten och dynamiken så att de kan utföra ett bra arbete.

En lång forskningstradition med att utveckla system för erfarna operatörer har lett till utformningen av MMSO-modellen (GMOC på engelska) för mänskligt styrning i komplexa miljöer. Modellen är baserad på kontrollteori, där fyra elementer är nödvändig för att utforma tekniska kontrollsystem: mål, modell, styrbarhet och observerbarhet. I forskningen i dynamiskt beslutsfattande har modellen används för att beskriva nödvändiga förutsättningar för människor som ska styra komplexa system. MMSO är resultatet av en omfattande forskning inom operatörsarbete och processtyrning.

Forskningsansats

Vårt forskningsområde ligger inom människa-datorinteraktion (MDI). Ett intresse inom MDI är att stödja användarcentrerad systemutveckling. Målet är att stödja systemdesigners och utvecklare, att i samverkan med användare, skapa system som upplevs vara användbara, vilket betyder att systemen ska vara effektivt, ändamålsenligt och tillfredsställande att använda och lära sig. Vi följer en aktionsforskningsansats vilket innebär att vi dels vill lösa konkreta problem, dels hitta allmängiltiga resultat och därigenom vidareutveckla de metoder som har används. Som komplement till aktionsforskning använder vi etnografiska metoder som hjälper oss att förstå människors arbete i deras miljö.

Mitt forskningsintresse har varit att undersöka, förtydliga och utvärdera MMSO-modellen samt hur den kan stödja systemutvecklingsprocessen. Trots modellens tidigare användning i flera projekt har den inte beskrivits i detalj. Första steget har varit att beskriva MMSO och ta fram en metod för att använda MMSO i systemutveckling. Andra steget har varit användningen och utvärderingen av resultaten, för att svarar på den första forskningsfrågan: Vari ligger nyttan av att använda MMSO i användarcentrerat systemutvecklingen? Mitt tillämpningsområde är främst tågtrafikstyrning. Genom att använda MMSO har jag analyserat tågtrafikledarnas och lokförarnas arbete, samt deras samverkan, för att kunna svara på den andra forskningsfrågan: Vilka principer och system kan förbättra kvalitet och arbetsmiljö inom tågtrafikstyrning?

MMSO

MMSO är en akronym för de fyra komponenter som är nödvändig för att möjliggöra mänsklig styrning av komplexa, dynamiska system. En operatör behöver tydliga mål, tillstånd att uppnå och hålla en process i. Dessa mål kan vara av olika art, t ex personliga eller organisatoriska. Det är vanligt att målen behöver brytas ner i delmål och att man måste prioritera mellan mål som står i konflikt med varandra. Dessutom behöver operatören en mental modell, som representerar kunskap om processen, rutiner, organisationer, styrsystem, omgivningen, och mycket mer. Mentala modeller är oftast begränsade, det vill säga att de inte exakt återspeglar hela verkligheten. Istället är de oftast förenklingar som förklarar de relevanta aspekterna av systemet för operatören. Systemet måste vara nog styrbart, dvs. ge operatören möjligheter att påverkar processen. Genom systemets observerbarhet kan operatörerna identifiera processernas och styrsystemens tillstånd. Observerbarhet kan vara visualiserad information, men också fysisk information, ljud, vibrationer eller liknande.

Dessa fyra komponenter är tydligt kopplade till varandra. Centralt är betydelsen av mentala modeller: de skapas genom interaktion med systemet och är nödvändiga för att besluta om åtgärder för att styra ett system mot det önskade målet. Mentala modeller ger operatören förståelse av det observerade systemets tillstånd och beteende. Genom att de får feedback (via observerbarhet) på åtgärder (användning av systemets styrbarhet) kan operatören få en bättre bild av hur systemet fungerar, vilket skapar möjligheterna för att utveckla den mentala modellen.

Att använda MMSO i systemutveckling

Som systemutvecklare eller designer har man möjlighet att påverka systemets observerbarhet och styrbarhet. Mål fastställs till viss del inom organisationen men tillsammans med mentala modeller kan de påverkas genom utbildning och erfarenhet. Dessutom kan man stödja prioriteringar mellan mål genom att skapa nödvändig observerbarhet. Eftersom observerbarhet och styrbarhet påverkar möjligheterna för utveckling av den mentala modellen, är det viktigt att sätta upp hypoteser om vilken design som är lämplig i det aktuella fallet. Genom utvärderingar kan vi sedan undersöka om hypoteserna var korrekta och om användarna har utvecklat de mentala modeller som man förväntade sig.

Systemutvecklingen bör följa en användarcentrerad designprocess som är byggt runt MMSO-modellen. Därigenom kan vi beskriva operatörernas mål, deras mentala modeller, samt ange krav på styrbarhet och observerbarhet. En sådan analys kan göras av den nu rådande situationen för att skapa kunskap om dagens arbete, system, problem och brister. Tillsammans med användarna försöker vi sedan ta fram nya lösningar för framtidens system. Detta sker iterativt och i samverkan med operatörerna. Under senare utvärderingar kan vi studera den nya arbetssituationen, efter införande av ett nytt system, för att se förbättringar samt för att identifiera ytterligare brister som behöver åtgärdas. MMSO-modellen hjälper oss att identifiera de aspekter på arbete och styrsystem som är nödvändiga för effektiv styrning och en för en god arbetsmiljö.

För att analysera och utvärdera system och arbetssituationer använder vi oss av etnografiska metoder, främst observationer och intervjuer. Syftet är en beskrivning av arbetssituationen som är så objektiv och detaljerad som möjligt. Under designen använder vi metoder som t ex målbilds-seminarier. Här deltar grupper av erfarna användare i utvecklingsprocessen. Designlösningar ska utvecklas baserat på den initiala analysen. Dessutom ska deltagarnas kunskap och erfarenhet utnyttjas för att validera data och utvärdera designprototyper.

Tillämpningar

Jag har tillämpat MMSO i studier av tågtrafiksystem. I Sverige styrs tågtrafiken idag från åtta olika trafikledningscentraler. Här arbetar trafikledare med att planera och styra trafiken baserad på dagliga tidtabeller och information om den aktuella trafiksituationen. En huvuduppgift för trafikledarna är hanteringen av störningar. Deras huvudverktyg för planering är idag en tid-sträcka graf på papper. Grafen innehåller den aktuella tidtabellen där ändringar ritas in för hand. Dessutom finns olika informationssystem som informerar om t.ex. förseningar. Trafikledarnas viktigaste verktyg för styrning är tågledningssystemet med dess trafikbilder, som innehåller en spårplan där aktuella beläggningar av blocksträckor visas. I styrsystemet kan trafikledarna specificera tågvägar för att styra infrastrukturen och på det sättet indirekt också trafiken.

Lokförarna ska följa trafikledarnas planer och köra tåget enligt plan mellan stationer. Deras information omfattar bl a körorder som består av en trafikplan baserad på dagens tidtabell samt kända förändringar och störningar. Lokförarna ska följa planen och köra enligt de signaler som finns ute på banan. Det sker idag en utveckling mot fler informationssystem på loket. Huvudmål är vanligtvis att ge lokförarna information som ska stödja eco-driving, dvs energisnål körning.

Mitt arbete omfattar huvudsakligen fyra olika delstudier. Den första är en analys av STEG systemet. STEG har utformats i samarbete mellan Trafikverket och Uppsala universitet. Det är ett nytt system för att styra tågtrafiken via direkt omplanering i en elektronisk tid-sträcka graf. Den ständigt uppdaterade planen kan sedan exekveras automatiskt. Under många år har systemet designats, utvecklats, använts i skarp drift och utvärderats. Hela processen har genomförts med hjälp av MMSO-modellen. Den andra studien, efter utvecklingen av STEG, omfattar en analys av hela trafikstyrningsprocessen och dess organisation. Kärnan var en analys av samarbetet mellan lokförarna och trafikledarna. Vissa principer har tagits fram för att stödja detta samarbete. Delar av principerna har realiserats i STEG som införts i Boden, en trafikledningscentral i norra Sverige, och CATO, ett lokförarinformationssystem utvecklat av Transrail. Min tredje studie är en undersökning av införandet av systemen med syfte att utvärdera våra koncept. Sista studien är en analys av lokförarnas arbete med syfte att identifiera användarnas krav på nya informationssystem.

Resultat

Användningen av MMSO-modellen inom tågtrafikstyrning visade sig vara framgångsrik. Våra studier hjälpte oss att definiera rollen som MMSO spelar i en användarcentrerad designprocess. Genom att beskriva och analysera ett arbete med hjälp av MMSO-modellen, blir det möjligt att identifiera brister som borde åtgärdas för att göra ett arbete mer effektivt. Jag kunde analysera den komplexa processen tågtrafikstyrning och identifiera hur de olika aktörerna samverkade samt hitta brister i deras samarbete som borde stödjas bättre av nya styr- och informationssystem.

Med hjälp av MMSO kan vi argumentera för viktiga grundprinciper när det gäller att visualisera information och automatisera styrning av komplexa, dynamiska processer. Visualiseringen är starkt relaterad till observerbarhet. Det är viktigt att man visar den nödvändiga komplexiteten för användarna. Försök att gömma komplexiteten kan leda till att operatörerna saknar den informationen som behövs för att kunna styra effektivt, vilket paradoxalt nog kan leda till ökad kognitiv belastning. Att visa mycket information samtidigt innebär inte alls automatiskt en överbelastning för operatörerna, om designen är korrekt gjord. Det gäller att utforma visualiseringen av komplexiteten genom att t ex skapa kända mönster, att framhäva viktig information, medan statisk information av mindre vikt kan tonas ner i bakgrunden. Att visa mycket information stödjer också utvecklingen av mentala modeller.

I mina studier av tågtrafikstyrning visade det sig att de traditionella system bara visade en mindre del av den information som trafikledarna behövde för att effektivt kunna styra trafiken. De fick därför utveckla komplexa mentala modeller som hjälpte dem att återskapa den information som saknades, baserad på den information de hade. Resultatet var att den återskapade informationen ibland kan vara felaktig, speciellt i oväntade situationer och att detta resulterade i en högre kognitiv belastning. STEG är därför utvecklat för att visa all den viktiga information som saknas i traditionella system.

Automatisering är relaterat till styrbarhet. Idag finns det många exempel på system som inte längre är styrbara utan automation. Viktiga grundprinciper för automation är att automatiska system är utformade på ett transparent sätt, vilket betyder att operatören måste kunna påverka automaten så att den agerar enligt operatörernas mål, s.k. icke-autonom automatisering, och att automaten måste tydligt informera om sitt tillstånd och planerade aktioner. Bara om en automat är styrbar och observerbar har användarna möjlighet att förstå automaten och integrera effektivt med den.

Automatisering är ett viktigt tema i tågtrafikstyrning. Automater kan t ex hantera möten mellan tåg på en mötesplats och därigenom avlasta trafikledarna från manuellt styrarbete. I praktiken visar det sig dock att automater ibland agerar mot trafikledarnas planer och därför måste stängas av under störningar, så att trafikledarna kan behålla styrbarheten. I STEG finns en automatisering som enbart strikt exekvera trafikledarnas planer utan att påverkar dem. Det har visat sig att den här typen av automation ger bra stöd. Trafikledarna behöver inte längre komma ihåg att utföra styrkommandon vid en viss tidpunkt, eftersom automaten tar hand om det. Det frigör kognitiv kapacitet för omplanering av trafiken och höjer kvalitén på trafikplanen.

Huvudkoncepten inom STEG är en ny styrprincip: Styrning genom operativ omplanering och automatisk exekvering. Styrprincipen betyder att trafikledarna kan planera med större framförhållning. Det ger dem mer tid att identifiera och hantera störningar på ett bättre sätt. Bristen på information och stöd i traditionella system gör det svår med framförhållning, vilket leder till en reaktiv styrning. För att möjliggör den nya styrprincipen behövs en automatisk exekvering. Studier av införandet av STEG i Boden påvisade vikten denna funktion. Tekniska problem i Boden gjorde att den automatiska exekveringen delvis inte fungerade. Resultatet blev att många trafikledare inte använde funktionen. Följden blev att trafikledarna i Boden ofta fortsatte att styra på ett reaktivt sätt.

Ett viktigt resultat av min analys av samarbetet mellan lokförarna och trafikledarna är att trafikledarnas styrbarhet kan förbättras rejält om lokförarna vore medvetna om ändringar i trafikledarnas trafikplaner i realtid. Idag är det enda möjlighet att kommunicera ändringar i planen via telefon. Speciellt i situationer med trafikstörningar finns det ett stort behov att kunna kommunicera förändringar, men då har trafikledarna en hög arbetsbelastning som gör talad kommunikation olämplig. Det behövs alltså ett nytt sätt att överföra planer mellan trafikledarna och lokförarna. Genom STEG finns det nu möjligheter att ge förarna tillgång till trafikledarnas realtidsplaner. Förarsystemet CATO ger idag delvis den möjligheten. Preliminära undersökningar visar att kombinationen av STEG och CATO kan minska behovet för trafikledarna att planera om, ett resultat av att lokförarna har bättre möjlighet att följa planerna i realtid, samt att lokförarna har bättre möjlighet att köra energisnålt. Eftersom införandet av STEG i Boden och utformningen av det nya arbetssätten inte är helt avslutat ännu, återstår en mer omfattande utvärdering av visningen av planen i CATO samt av effekterna av automatiskt överföring av realtidsplaner.

Jag har presenterad MMSO-modellen och visat möjligheterna att använda den inom användar-centrerad systemutveckling. Mina studier inom tågtrafikstyrning visar att MMSO stödjer hela utvecklingsprocessen från analys, via design, utveckling och införande till utvärdering. Med hjälp av MMSO blev det möjligt att analysera den komplexa tågtrafikstyrningsprocessen och utveckla principer och system för framtidens tågtrafikstyrning.

Contents

Part I:
Introduction

1. Introduction

1.1 Studying Human Control in Complex, Dynamic Environments

Imagine that you are asked to describe the task of driving a car by someone who is not familiar with doing so. How would you start? Maybe you would begin by explaining the first things you do when you enter a car: you adjust the seat so that you can reach the controls comfortably; you adjust the mirrors so that they reflect the important areas around the car; you buckle the safety belt and adjust the strap over your chest. You might also begin explaining the main driving controls and the instrumentation on the dashboard. The steering wheel, brakes, and speedometer are quite straightforward to explain – you skip technologies such as servos, ABS brakes, and electronic stability controls for now. The rev meter and clutch can be harder to describe; even though a driver knows well how to handle these, he or she might not know exactly how they work.

Let us switch to a situation in traffic. To drive your car safely, you need to be aware of a lot of things at the same time. You need to know the state and position of your car, its speed and brake condition. You need to know the traffic situation; the number and behaviour of cars, bikes, and pedestrians around you; their speed and the probability with which and how they might interfere with your route. Your planned route is another thing you need to know. This is already a long list, even though we have not yet mentioned traffic rules and weather conditions.

The purpose of this short exposition is to show how complex driving a car is and, even though you are probably familiar with it, how hard it can be to describe this task. The main topic of this thesis is the improvement and development of new support systems for human operators in complex, dynamic environments. Such improvements must start with an analysis of the complex environment. *Complexity* refers to the (high) number of interconnected variables in a process. In the example of road traffic, among these variables are the different participants and their parameters, such as speed and direction. These variables are interconnected as each has the potential to influence the others. *Dynamics* here refers to the dynamic properties of the process, meaning that it evolves from its inertial dynamics – in our example, the movement of different traffic participants – and that the process dynamics are influenced by our own actions. When we want to analyse or influence such complex, dynamic environments, we face many challenges.

A comprehensive analysis of a complete environment is often not feasible, as involved organizations, actors, and systems are simply too large, or budget and time constraints for a project would be too tight. So we have to focus on the part of the environment which is most relevant to study. However, if we restrict ourselves too much, we risk ignoring the bigger picture. In the best case, this means missing considerable potential for improvement; in the worst case, it means sub-optimizing, resulting in a negative impact on the process from a holistic perspective. Thus we need methods that help us focus on the important parts but still keep a holistic view.

And there are further challenges. When dealing with larger organizations, it is often not obvious who will be able to provide which part of the required information. The actors who participate in the study are often not involved in all processes. Revisiting the example of driving a car, the planning of road repairs and the construction of new infrastructure are such processes, into which a car driver usually has very limited insight. However, many drivers still have an opinion or idea about how these processes function. As researchers, we have to identify and separate suppositions and opinions from detailed knowledge. In the ideal case, we should be able to consult actors participating in the processes in question, if necessary.

Additionally, knowledge and experience differ, and views are naturally biased, that is, based on personal understanding, meaning that two persons can do exactly the same tasks but have completely different strategies and ideas about their execution. And although experienced operators usually have no major problems coping with the complexity of their environments, they may be in trouble when asked to explain their actions. The explanation for this can be a strong internalization of common tasks and actions. This reduces actors' mental workload for such actions, but it also makes the actions hard to explain consciously. An experienced car driver, for example, intuitively knows when to change gears or when to start to brake. To obtain a sufficiently complete and correct description of the work of human operators, we need methods that help us to reduce biases and to observe and understand automated actions.

1.2 Train Traffic Control

Train traffic control is an example of a very complex, dynamic process. It was the main area of application of my research. For most people, the details of this process are much less familiar than those of driving a car. I will discuss train traffic control in Sweden in detail in Part III; here I will just introduce some aspects of its complexity.

Railway traffic involves many actors in different roles and organizations. To achieve optimal results, these actors have to act collaboratively, using compatible strategies, towards the same goals. Railway systems have historically grown and differ in many aspects from one country to another. Even inside the same country, significant differences can be observed between signal boxes, signalling systems, and so on. The main reason is that infrastructure has been constructed and upgraded at different times with the technology available and reasonable at the time.

Inside railway traffic, many sub-processes and topics for research and development exist: development of safety systems, timetable production, resource planning, maintenance planning and execution, operative train traffic control, train driving, customer information, real-time optimization and automation, supply of electricity, freight distribution, simulation of many different kinds, and so on. This thesis concentrates on operative traffic control and train driving.

Train traffic control is a real-time process that transfers the timetable into actual traffic. The goal is to keep the traffic as close as possible to the timetable and to efficiently deal with disruptions. These range from small perturbations, for example, delayed departure from stations, to large disruptions, for example, infrastructure failure at busy junctions, with consequences ranging from smaller delays to re-routing or cancellation of trains. Within traffic control, infrastructure is controlled, meaning that points and signals are set to generate train routes, setting the stage for the train drivers.

1.3 Research Objectives

Behind this thesis lie two main efforts: firstly, the theoretical contribution to the research field of human–computer interaction, and secondly, the development of concepts and systems for future train traffic control. These interests are mutually beneficial. As we aim for systems supporting human operators, human–computer interaction supplies us with methods with which to analyse the operators' environment and develop new or improved support systems. In turn, an evaluation of the results and experiences from these studies can provide feedback on the applied methods.

This is particularly important, as we combine methods and models in a novel way. Therefore the evaluation of methods and produced results is a necessary step to improve and legitimate our approach.

1.3.1 A Model to Support Development of Systems for Human Operators in Complex Control Environments

As pointed out, when analysing human operators in complex, dynamic environments, we need guidance to focus on the relevant aspects without losing the big picture. In our research group, a model for human work in process control called GMOC has evolved – an acronym for *goal*, *model*, *observability*, and *controllability*. By 'model', we mean the mental model (see Section 4.2) of the human operator. These four elements are seen as necessary prerequisites for human operators to achieve control.

The roots of GMOC lie in control theory. For us, the most influential work has been its adaptation to dynamic decision making [11, 13]. Since then, it has been used in several studies [5, 68, 54, 83, 48]. However, despite its regular use, GMOC has never been clearly defined. One main goal of this thesis is to define and anchor GMOC so as to make it available to the research community. To do so, I take the following steps:

1. I provide a basic description of the model by definition of the four elements goal, model, observability, and controllability as well as the importance of their interplay (see Chapter 5).
2. I relate GMOC to different theories and concepts from human–computer interaction (see Chapter 2).
3. I provide a generic method for applying GMOC in a user-centred system design process (see Chapter 9).
4. I present case studies in which GMOC has been applied to analyse work environments and develop systems for human control of complex, dynamic processes (see Part III).
5. I evaluate the model and the used methods, based on the results of the case studies (see Part IV).

Research Question 1 What is the utility of GMOC for analysis of human control of complex, dynamic environments and design of new support systems?

1.3.2 Development of Concepts for Future Train Traffic Control

Our research group has a long, ongoing collaboration with the Swedish transport administration Trafikverket.[1] Trafikverket fills the role of infrastructure manager and is responsible for planning, construction, maintenance, and operation of the railway infrastructure in Sweden. Our research is financed by Trafikverket via its programme KAJT[2] ('Kapacitet i järnvägstrafiken'; in English, Capacity in the Railway Traffic System) and by the European research project On-Time.[3]

KAJT coordinates research in Sweden concerning the capacity of railway traffic systems. Projects under this programme cover, for example, infrastructure analysis and planning, timetable processes, algorithmic planning support, and traffic management. The On-Time project aims for increased capac-

[1] http://www.trafikverket.se/.
[2] http://www.kajt.org/.
[3] http://www.ontime-project.eu/.

ity in railway traffic as well. It focuses on reduction of delays by improved timetabling processes, automatic perturbation management, and train driver support systems. A main result has been a technical framework for future train traffic control and management that facilitates integration of automatic systems and controls across European borders.

These projects provide the application area for the user-centred design process developed around GMOC. We use them in an exploratory way to find new concepts to improve human control of railway traffic. Our goal is to develop support systems that improve the work situation in terms of GMOC. This means especially increasing observability and controllability of the process and supporting the human operators to develop better mental models, that is, their understanding of how to control the process (see Chapter 5). Additionally these new systems should improve the quality of railway traffic, for example, in terms of reduced delays and increased capacity (see Section 10.3 for a further definition of optimization criteria in railway traffic).

In Chapter 10, I introduce the traditional situation of operative train traffic control in Sweden, including an analysis of its organization and the roles of traffic controllers and train drivers. This chapter shows the complexity of train traffic. I describe three consecutive studies in this setting in Chapters 11–14 and evaluate our general findings in Chapter 15.

Research Question 2 Which concepts and future support systems for train traffic controllers and train drivers can improve their work environment and the quality of train traffic operation?

1.4 Delimitations

The domain of railway traffic offers a broad field for research. There exist plenty of different roles and many different support systems and challenges for human factor research. Our analysis of the organization of railway traffic revealed that train traffic control is dependent on collaboration between train drivers and traffic controllers. Despite this, their support for collaboration is very limited. Therefore I chose to focus on analysis of the collaboration between train drivers and traffic controllers and the development of concepts and systems to support their collaboration.

Starting in late 2013, Trafikverket introduced new roles to the traffic control process. These organizational changes are still ongoing, and their effects could not be sufficiently analysed at the time of research. This thesis describes the situation until 2013. The presented results are still relevant, as they mainly concern the collaboration of train drivers and traffic controllers – these roles still exist unchanged.

Our studies are part of long-term projects with a horizon of several years. It was not possible to repeat them or conduct similar projects applying a different method. Therefore I chose to evaluate GMOC by the results obtained from our studies and not by a deeper comparison to other approaches and development methods.

1.5 The Relation between the Research Objectives and Papers

Figure 1.1 shows how the research papers included in this thesis are related to the thesis's theoretical contributions (research question 1) and application (research question 2). Paper I describes the conceptual model GMOC for analysis and design of human work environments. It is the main paper showing the theoretical contribution to human–computer interaction. The model itself had already inspired the work of my research group for many years; my main contribution is to clearly describe this model and further explore its theoretical foundation and the importance of the interplay of its elements. Closer to the theoretical side is Paper II, dealing with the visualization of complex, dynamic processes. It is based on the thoughts behind GMOC and includes aspects of the application, as well as being informed by the results of our studies in the development of systems for train traffic control.

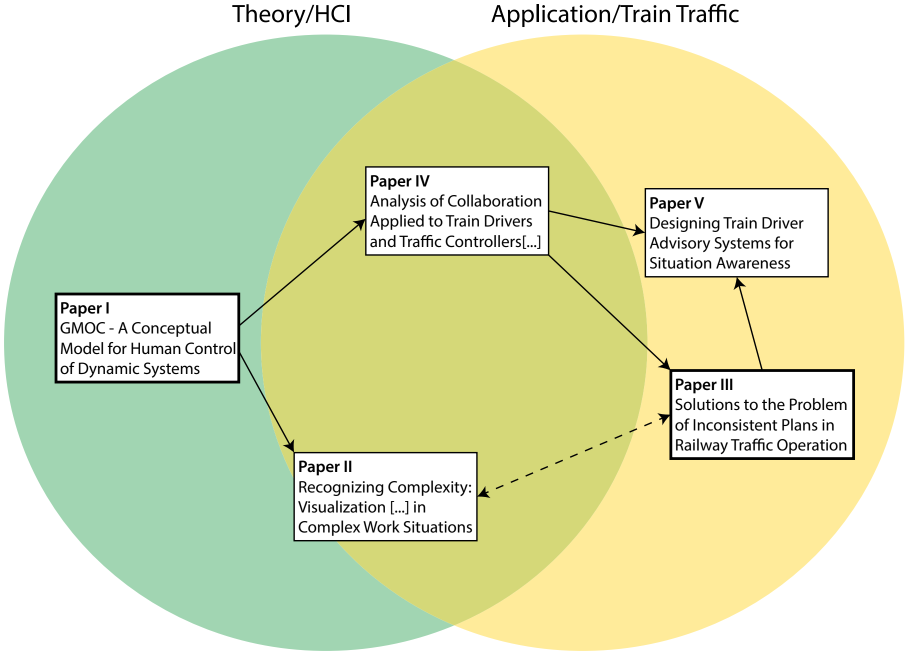


Figure 1.1. Relation between the papers included in this thesis and the research objectives. Paper I is central for the contribution to human–computer interaction, Paper III summarizes the contribution to future train traffic control.

While Papers I and II concentrate on contributions to human–computer interaction, Papers III–V are directly related to our studies in train traffic control. Paper III is most comprehensive, summarizing our progress towards the development of future concepts and systems in Swedish train traffic control. An important part of the paper describes the organizational context of operative traffic control. Owing to the fragmented organization of railway traffic, often a result of deregulation, awareness of this bigger picture is important. Especially research on systems and algorithms for (partial) automation of railway traffic tends to forget or even avoid this bigger picture. Our experience shows that missing this perspective easily leads to solutions with questionable value for practical application. Paper IV is a result of the study described in Chapter 12. It is related to the theoretical side as it extends GMOC to analysing collaboration. This is done through the example of train traffic controllers and drivers, which in turn was the basis for Papers III and V. Paper V summarizes the design of an advisory system that can give train drivers a better understanding of the surrounding traffic. It is a result of the study described in Chapter 14.

Part II:
A Model for Analysis and Design of Systems for Human Process Control

This is the theory-oriented part of this thesis. At its core, this means the GMOC model. It is a conceptual model to describe human work in complex control situations. The attempt of human operators to achieve control manifests in their actions influencing the system. This part starts with a description of the research field and the applied research method. It continues with a closer look at the operators' decision making, the process pre-pending the action. After introduction of the GMOC model, we discuss design in accordance with this model. GMOC has a number of implications for automation and visualization. The part concludes with a proposal of methods that are useful to utilize GMOC in the user-centred design process.

2. Research Field

This chapter frames the research presented in this thesis. A discussion of different paradigms in human–computer interaction (HCI) marks the start of this chapter. The contents of this thesis are then related to important goals of HCI. The chapter concludes with a definition of the main properties of the studied environments.

2.1 Human–Computer Interaction

HCI is a multi-disciplinary field, combining themes, theories, and methods from a multitude of other disciplines. Among these are, for instance, social sciences and computer science in general, and more particular, psychology, cognitive science, system development, ethnography, design disciplines, and many more. All these are combined to aim for the development of more usable systems, including usefulness, user experience, and efficiency. Additionally, this work is rooted in human factors. A simplified distinction is to see HCI as more oriented to the user in general, focusing on user experience and interaction design, while the field of human factors focuses more on functional aspects, such as ergonomics and safety. The relation of this thesis to human factors therefore lies in its application domain, process and train traffic control. Here traditionally such functional aspects are of higher priority.

Harrison, Tatar, and Sengers identify three paradigms in HCI [41]: (1) *human factors*, which, in the tradition of ergonomics, identifies problems in the coupling between human beings and systems, aiming for an optimal fit; (2) *classical cognitivism/information processing*, with interest in abstract models explaining interactions between human beings and computer systems in the relation to their information processing capabilities; and (3) *phenomenologically situated* interaction. So, firstly, this view sees human factors as an integral part of HCI, a view that fits well into the constructs of this thesis.

The GMOC model we present has as a goal to support the analysis of environments for process control. We are aiming for identifying and solving problems in the interaction between operator and system. Thus follow the goals of the human factors paradigm as described by Harrison et al. [41]. This is in contrast to the cognitivistic view, where models aim for a detailed description and explanation of (human) information processing. The third paradigm recommends viewing interaction in the social and cultural context. This emerging view builds to a large degree on Suchman's 'Plans and Situated Actions' [90].

We definitively agree with the importance of context when analysing and designing work and systems. Methodologically, we have strong overlap with the third paradigm, placing the stress on ethnographic methods.

One way to interpret the goals of the three paradigms is through their common theme of supporting the development of 'better' systems. This might not cover each and every researcher but seems to be a valid generalization. To be more concrete, *better* has to be defined. In the context of HCI, the notion of usability is central to the quality of a system, so the next step is to define *usability*. This, then, leads to the question of why achieving usability is a problem. Here we hit the central problem of HCI, namely, that computers and human beings 'speak' a different language.

2.1.1 Usability

The first question when trying to assess usability is the question of *purpose*. We need to know what users want to achieve to be able to judge if a tool provides usability for that purpose. Let us say, for the sake of illustration, that the users want to have a coffee and we are evaluating the usability of a coffee maker.

We can expect that the coffee maker will offer the functionality of producing coffee. But can we say that it meets users' purpose if the users do not know how to produce coffee with the machine? The important point is that a person has to know or be able to learn the use of a system. Thus *learnability* (fast and easy to learn) and *memorability* (easy to remember how to use) are relevant criteria for usability [85].

Further aspects of usability, regardless of the object being a coffee maker or a software system, are *effectiveness* and *efficiency*. If the coffee maker produces great-tasting coffee and keeps produced coffee warm and fresh, it is certainly very effective. In efficiency, a machine that brews one cup at a time might beat a machine that brews a whole can at once, but if we need coffee for a whole group, it might be the other way around. A last point is *satisfaction*. Using the metaphor of coffee making, this means that it is probably more pleasurable to have the coffee served in a nice mug by a nice-looking coffee maker that works silently without spill than in a thin and cheap disposable plastic cup by a huge and noisy machine.

In the ISO-standard, usability is defined as

> the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. (ISO 9241-11, 1998)

'Can be used' resembles the criterion that the usage of an object has to be known or learnable. 'Specified goals' and 'context' are equivalent with the purpose. On this basis, the general definition for usability of an object in the context of this thesis is as follows:

Purpose The goals to achieve by usage of the object in a specific context, for example, by a certain user group in a certain task.

Learnability and memorability The users' knowledge of (memorability) or ability to learn (learnability) ways to utilize the object towards the purpose.

Effectiveness Quality and degree of goal achievement possible by utilizing the object, given the users' ability.

Efficiency Effort needed to achieve a goal by utilizing the object, given the users' ability.

Satisfaction Degree to which the users are satisfied by the object and enjoy its use.

2.1.2 The Gap between User and System

My personal view on HCI is that the central problem in the interaction between human beings and computers or users and systems is communication. Computers are built using the logical language of zeros and ones, whereas human beings are conscious, autonomously thinking, influenced by emotions, and able to communicate on many different levels. Everyone who once got angry at a system that was not easy to understand or did not behave as expected (e.g., a complicated ticket machine close to departure of a train or a computer crashing during editing an important manuscript) – so basically everyone who has ever had contact with a computer system – knows that computers can create emotions. However, they cannot do so consciously.

Many people are involved in the development process until a user communicates with a computer. Hardware designers make the zeros and ones accessible via memory, CPUs, registers, and so on. Developers of operating system cores and programming languages make these technical components available to programmers, who can communicate with them on the level of programming languages. Their work results in software that is available to the end user. This software presents the highest level of language a computer is able to speak. Our goal is to make this as understandable to human beings, or at least the designated users among them, as possible. Seen from the opposite perspective, if we want to make computers 'usable', we have to make them understand and speak the language of their users. Researchers in HCI often describe this problem as a *gap*. Gaps differ in certain aspects, but they always describe a scenario where seamless interaction between human beings and computers is hindered by their different natures.

One such gap is the *socio-technical gap*, as described by Ackerman: "There is a fundamental mismatch between what is required socially and what we can do technically" [1, p. 198]. This means, for instance, that tasks that we can easily perform in social situations in everyday life are quite hard and cumbersome to support with computer systems. Ackerman even argues for making the gap 'an explicit intellectual focus' in the disciplines of HCI and computer-supported collaborative work (CSCW) [1, p. 192].[1] Further research on the phenomenon and understanding the implications of the socio-technical gap are necessary to improve usability.

Instead of a gap, Norman sees the issue in the form of 'two gulfs that must be bridged: the *Gulf of Execution* and the *Gulf of Evaluation*' [66, p. 38]. In this picture, we have the users' goals on one side and the physical state of a system on the other. These gulfs are bridged by users and system designers. Users perceive and interpret the system state and translate their goals into action sequences supported by the system. System designers develop the interface which is the users' point of interaction. This means that a part of the bridging is done on-line during the use of the system, while the other part is done in advance, during the design process. Being system designers aiming for usability, we should make users' 'on-line bridging' as efficient and satisfactory as possible.

2.1.3 The Gap between User and Designer

The described gap between users and systems is well known and an issue often broached. When we design or develop a system, we approach another gap, the one between users and designers. Norman [66] demonstrates this gap very well: system designers create a design model which is the basis for a system to build (this is further discussed in Section 8). Users then form their own model of a system, based on their interactions with it. In the ideal case, these two models are equivalent. However, examples where this is not the case are not difficult to imagine. The simplest examples are systems where the users' model is only a subset of the design model; that is, the system provides functionality of which users are not aware. In the best case, users are able to identify and understand this functionality when they need it. Discrepancies between the users' model and the design model become critical when users interpret a function differently from what is implemented in the system. Take the common paper basket metaphor on desktop systems as an example. A user can drag files there to delete them but recover them if he or she deletes the wrong files. Now imagine a system that permanently deletes any file dragged to the paper basket and the surprised user who made a mistake.

User-centred methods usually have as a goal to close the gap between user and system. When there is a gap between user and developer, this is impossible. Therefore these methods implicitly mean to close the gap between user and designer as well. Common solutions involve users in the design process.

[1] CSCW is closely related to HCI. It deals explicitly with computer systems in collaborative scenarios and related social and organizational aspects.

One way to explicitly address this gap is to ensure that designers have a correct understanding of the user's model on which they base their design. Laaksoharju thus recommends that designers formulate their understanding in terms of hypotheses that can be falsified [59]. This is a utilization of the concept of mental models to improve system design. A further approach is to formulate hypotheses on future mental models. This means to hypothesize how the users' mental models will be affected by a new functionality or system. It is this approach that we aim to support with the GMOC model.

2.1.4 This Thesis in Relation to HCI

A central contribution of this thesis is GMOC. Figure 2.1 depicts its relation to concepts and aspects of HCI. GMOC is intended for systems supporting human control of complex, dynamic processes. We therefore focus on concepts and aspects that are typically most relevant in this type of application.

We identified the development of systems with the users' needs in focus as a central aim of HCI. Therefore *user-centred system design* (UCSD), a process with the goal to create such user-centred systems, is set at the top of the figure. GMOC is supporting UCSD, as is discussed in Chapters 8 and 9.

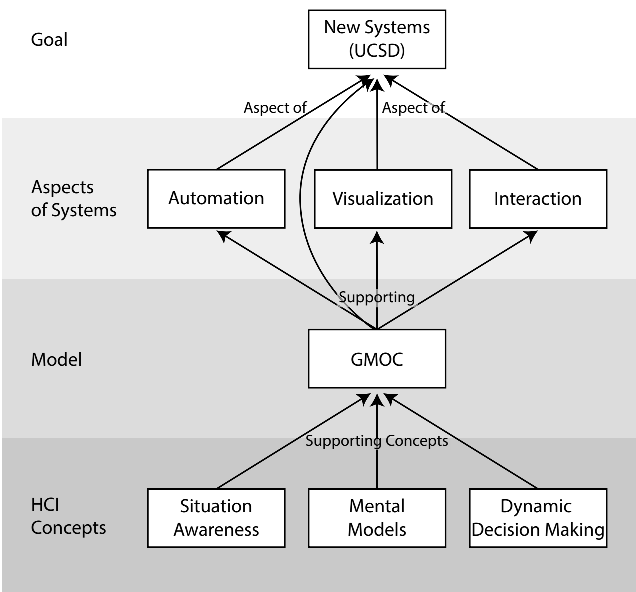


Figure 2.1. Relation of the concepts presented in this part, reflecting the relation of GMOC to HCI.

Below the goal lie the important aspects of the systems we aim for. *Interaction* with the system and *visualization* of the available information and system state are relevant aspects of systems in general. Especially important in the domain of process control is *automation*. It is a part of most systems for process control, often as a sub-system between the user interface and the process to control. Development of these aspects is supported by GMOC. Automation is covered in Chapter 6 and visualization in Chapter 7. Interaction is not covered separately; it was slightly less stressed during the work on this thesis. A discussion of interaction is included in the descriptions of automation and GMOC (see Chapter 5).

During the development of GMOC, three concepts seemed particularly relevant. These are popular in or have emerged from areas belonging to HCI. *Dynamic decision making* is a specialization focusing on decision making in complex, dynamic processes. It thus follows the same theme as our research and even provides the basis for our work. Dynamic decision making applies the framework of control theory to human work. We followed this approach, which led to the concrete formulation of GMOC. *Mental models* are another basic concept. These can almost be seen as a direct part of GMOC, as is discussed in Section 5.2. *Situation awareness* is the third influential concept. It emerged in the domain of aviation and became popular in human factors and process control in general. Situation awareness is widely seen as a precondition for good performance in process control. A system designed following GMOC will most likely facilitate situation awareness as well. So, to some extent, the facilitation of situation awareness gives an explanation for the utility of GMOC. Situation awareness therefore supports our reasoning, but it is much less influential on GMOC itself than mental models and dynamic decision making. The three concepts are presented in Chapter 4.

The second main part of this thesis describes the development of new concepts and systems for train traffic control. This work lies in the field of rail human factors. It is a specialization of human factors in the domain of railway traffic. The focus of our work is on operative train traffic control. This domain is introduced in Chapter 10.

2.2 Dynamic and Complex Systems

The concepts discussed in this thesis are seen in the context of human control of complex, dynamic systems. This section defines what we mean by the terms *complex* and *dynamic*, and by *system* and *environment*.

2.2.1 System and Environment

The term *system* covers a large range of possible definitions. At the core of the systems we discuss lies a *process* to control. A process can be small with a simple structure or large with several sub-processes. In the example of train traffic control, a train is part of the process train traffic. The train traffic controllers at the infrastructure manager's control centre monitor train movement and control the infrastructure according to traffic plans. This is a basic necessity to allow the trains to move. Even if this is a different level of control compared to a train driver's task, it still is a part of the train traffic process. Other examples of processes to control are chemical processes, such as in the pulp and paper industry, production of electricity, as in nuclear power plants, or firefighting.

The most basic system would comprise a computer or machine sitting between the operator and the process. However, in the areas on which we focus, usually many actors and systems are involved in control of the same or integrated processes. Many combinations are possible: it can be several operators in the same role controlling the process via their own or a shared system, close to each other or far away, or actors in completely different roles and organizations. Thus we are talking about *socio-technical systems*.

Vicente [93] describes 'complex' socio-technical systems generalized as consisting of the four layers: technical system, workers, organization, and the environment (see also [65]). We basically agree with this view. The process would, though, cover technical system, workers, and organization.

Another aspect is the *environment* in which the process is embedded. In the layered view this would be external effects on the process, caused, for example, by weather conditions such as temperature changes. In train driving, such influences could be rail conditions (to a high degree influenced by weather conditions) or slopes of the track. However, in this context, the term *environment* can also cover the surrounding traffic situation. Thus environment is interpreted more as a work environment which is different for each actor in the process. As we analyse systems from a user perspective, we use the term 'environment' or 'control environment' following this interpretation. These terms include not only the environment in which the process is located and the conditions prevailing around process and system but also the process itself. So, in some sense, the terms cover the socio-technical system as a whole.

2.2.2 Complexity

Complexity refers mainly to the often large number of involved components, systems, actors, and large problem spaces, consisting of information sets, subprocesses, and state variables and their properties, interconnections, and so on. Different aspects of complexity, similar to our view, are discussed, for example, by Perrow [75], Woods [97], and Vicente [93]. Besides the main definition of complexity – many interacting components – other aspects can also contribute to complexity. Examples are non-linearities, control actions that can only be taken at specific points in time, external stochastic disturbances, or interactions with other operators or automatic systems.

If a system is complex, this means that many components, their relations, and their properties are involved and must be considered by the operator. The operator interface must visualize appropriate information, which often means a large amount of information. We have found (see Chapter 7) that it is important to accept the complexity and visualize all information the operators need, and not to try to hide information or simplify the presentation. The design of interfaces must support the operators to cope with the complexity. Complexity is one of the main reasons why a system is *complicated*, that is, difficult for human operators to understand, learn, and control. Ironically, systems that try to hide complexity often make a task more complicated.

2.2.3 Dynamics

Dynamic systems most often manifest in processes that evolve over time, not only as consequences of influences, for example, control actions, from outside, but also as changes caused by internal properties. The opposite of a dynamic system is a static system, where the behaviour at any point in time is directly related to the interaction with it at exactly that point in time. A dynamic system can be stationary, that is, not 'moving' or developing over time. Dynamics is a property of the system and is not related to whether it is 'moving'.

Dynamic responses to interactions may not have effect immediately, but later on, and the behaviour is at any moment a combined effect of system properties and all earlier interactions. To control a dynamic system towards a specific goal, appropriate control actions must be taken in advance. The movement of a heavy ship is, for example, not only the result of the rudder movements just made but also results from movements over a long time period. To manoeuvre the ship to a certain position, a series of well-planned control actions must be taken in advance. In control of a traffic system, for example, train traffic, which is highly dynamic, control actions of different types must be taken at specific points in time, sometimes long in advance and in the right sequence.

A dynamic system 'has a memory', and in mathematical terms it can only be modelled by differential equations. The development over time is determined by the dynamic model for each state variable, initial values for all variables, and the collective effects of all previous control actions. Similar to our view, aspects of dynamic systems relevant to human decision making and control have earlier been discussed by, for example, Edwards [24], Rapoport [79], and Brehmer [11].

Dynamics are often seen as reasons for complexity (cf. [97, 93]). We chose to see these two as separate properties of processes. Instead we prefer to see dynamic properties, in addition to complexity, as reasons why the control tasks are complicated, difficult, and require highly skilled operators, long experience, and well-designed support systems.

3. Research Method

The research presented in this thesis basically follows an action research model, but our research methods have also been influenced by ethnography. This chapter describes how the ethnographic methods have been combined with action research.

3.1 Ethnography

Originally, ethnographers were researching and studying societies that developed in a cultural setting that was distinct from the Western culture with which the researchers were familiar. A goal of ethnographic research was to produce a detailed picture of those societies without influencing them. Methods ranged from observations that were as unobtrusive as possible to direct participation in activities, but still with the goal not to influence the culture under observation. Later, the ethnographic approach was adopted by researchers who were interested in studying human work [15]. Here the study object was no longer part of a distinct culture and society but a group in the same cultural environment as the researcher. From the studies of human work, the application of ethnography in HCI has evolved. Researchers realized that it was not always sufficient to study users interacting with computer systems in laboratory setups but that studies in the context of the users' work environment, including social and collaborative aspects of their work, gave further insight [10]. Ethnography was providing fitting methods.

Ethnography is a process, covering information collection, analysis, and documentation during field studies. It is important that the analysed situation be described with as much detail and knowledge as possible. Therefore the process of data collection is quite important. A basic principle is to use 'natural' environments for data collection. In human work this usually means data collection at the workplace or, if this is not possible, at least in simulated scenarios that feel realistic. It is further important to be open-minded and observant so that the collected data are relatively 'objective' and as few details as possible are missed. Typical methods used in the context of HCI and adopted from ethnography are interviews, observations, and participatory design. These methods have been important for our work as well and are described in Chapter 8.

Ethnography belongs to the third paradigm of HCI, the one concerned with being phenomenologically situated [41], as described earlier. But as its methods are spread further in the field of HCI, one has to consider the following quote from Blomberg: “For some, ethnography is simply a fashionable term for any form of qualitative research. For others, it is less about method and more about the lens through which human activities are viewed” [10, p. 964]. When we apply methods such as observations and interviews, we certainly have the intention of reproducing someone’s work as knowledgeably and objectively as possible. Respect for the human operators, their work, and their experience is a ground rule. However, our main goal is to create new work environments, which often lead to new routines and changes in the users’ tasks and work. Thus the informed analysis and description of someone’s work is just the first step that necessarily precedes the design and introduction of the changes. Respect for the operators is necessary to generate meaningful designs and evaluate them. But it is the central part of our research to eventually change the work environment. This striving for change is a basis of our research method.

3.2 Action Research

Action research is a process that combines theoretical research with practical problem solving. This means that there are always two sides to consider: the side of the practitioners and the side of the researchers. In the context of HCI, practitioners are the users of a system. The basic principle in action research is the presence of researchers aiming for changes in the users’ work environment. The purpose of these changes is to improve the users’ work environment towards certain objectives while generating new knowledge with relevance for the research objective. Additionally, the users should be empowered by the knowledge generated during the process, allowing them to improve and gain control in their work environment. Action research can be seen as an iterative process where phases of planning, that is, analysis of the current situation and formulation of changes; intervention, that is, the introduction of changes; and reflection, that is, analysis of the results of the intervention and knowledge generation, alternate.

A strong advantage of action research is its ability to deal with the complex and ill-defined problems that are usually found in the real world [6]. It allows researchers to test theories in real-life situations, to adjust them, and to test them again. This gives the researchers a genuine impression of the users, their tasks, the systems – basically the complete work situation – leading to the more holistic view that is also promoted in this thesis. As action research produces both knowledge and skills for researchers as well as for users, it can also be seen as a win-win approach [62]. However, action research has a couple disadvantages. Most important is probably that action research is prone to result in consultancy rather than research [6]. Similarly, it is sometimes not seen as ‘scientific’ enough, as results might be biased, lack rigour and validity, or be hard to generalize [62].

Table 3.1. *Elements of the action research used in this thesis*

Framework	As described in the previous chapter (HCI); GMOC seems to be useful in developing systems for human control in complex, dynamic environments
Research methods	Action research and ethnography
Problem-solving methods	User-centred system design utilizing GMOC (see Chapters 8 and 9)
Problem situation	Human control of complex, dynamic environments – analysis and development
Problem-solving interest	Improving the quality of train traffic control

Reason [81] and McKay and Marshall [62] provide different approaches to avoiding these pitfalls. Reason sees action research as a process based on choices with four characteristic dimensions: (1) worthwhile practical purposes, that is, that the problem to solve has a significant relevance in practice; (2) democracy and participation, that is, that the users have explicit influence in and benefit from the process; (3) many ways of knowing, that is, the way knowledge is created from the combination of theory and practice; and (4) emergent developmental form, that is, that the research design is not unalterably fixed during the process but has to adapt to emerging circumstances. It is the researchers' task to continuously make choices regarding the aspects of these dimensions on which the study should focus. By making these decisions transparent, the researchers can guarantee the scientific value of their research [81].

McKay and Marshall [62] discuss action research from an information systems perspective. They suggest dividing the process into two concurrent but distinct cycles, the first concerned with the problem-solving interest, the second with the research interest. Whereas the first cycle can be similar to consultancy, the second cycle clearly is not [62]. From this, implications are drawn on a model for action research. This model consists of five elements: (1) a real-world problem situation, (2) a theoretical framework, (3) a research method, (4) a problem-solving method, and (5) a problem-solving interest. Table 3.1 shows a mapping of this thesis to the five elements. The framework of this thesis is given by the research field and our research objectives. The first objective is to support the analysis of human control of complex, dynamic systems and the development of new support systems. This objective fits best as a description of a real-world problem. A certain focus lies on the development of the GMOC model as a tool supporting the user-centred system design process. This makes GMOC part of both the research interest and the problem-solving method. The development of concepts for future train traffic control, the second research objective, gives the problem situation.

It might be surprising that, despite the elementary conflict between ethnography and action research – non-intrusive observation versus intrusive action –

both have been chosen as the research method for this thesis. First, obviously, action research is a part of the research method only, while the methods used for problem solving represent an ethnographic view. But second, ethnography has even influenced the research method, as we see it as important to gain an adequate understanding of the environment to study before proposing any changes. Furthermore, there are some substantial similarities between the two approaches. Examples are the importance of performing any studies in the natural environment and the deep respect for the study subjects:

> A key value shared by action researchers, then, is this abiding respect for people's knowledge and for their ability to understand and address the issues confronting them and their communities. [14, p. 14]

4. Decisions in Dynamic Environments

Making a decision requires certain information and knowledge. Different approaches exist to explain human *decision making*. Two popular concepts can be utilized to try to describe what is needed to make an informed decision: *situation awareness* as a notion of the operators' understanding of the current situation, that is, the status of the system, and *mental models* as a notion for the operators' understanding of the way the system functions. These concepts are quite common in research on human process control. This section gives a short overview of these three topics and discusses a way to relate them to each other.

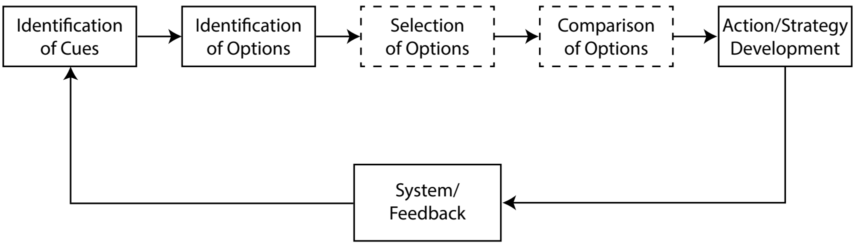
4.1 Decision Making

A fundamental article in cognitive science is Simon's article on bounded rationality [86]. It questions the picture of the 'economic man', who makes thoroughly rational decisions based on all relevant information. Instead it paints a picture of choices made by simplified optimization based on assumptions and the information at hand, using, for example, simple payoff functions to estimate acceptable outcomes of a choice. On the basis of this idea, decision making explores the ways human beings make decisions, that is, which assumptions and simplifications are made, and how they lead to a decision.

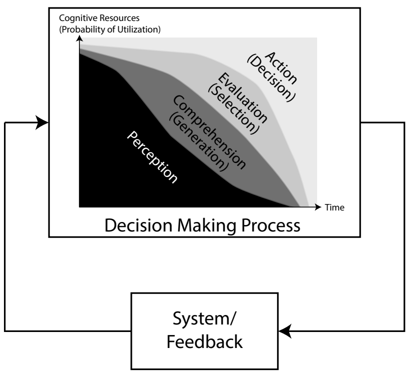
4.1.1 An Overview of Different Approaches

Figure 4.1a shows a simplified decision-making process. It is based on the assumptions on which Klein et al. started when analysing decision making in firefighting [57]. The cycle starts with identifying cues in the environment that give a picture of the situation with which one is confronted. This understanding leads to the identification of options for action. Depending on the underlying theory, a list of options is generated and ordered. Finally, the most promising option is selected and, probably with some adaptation, implemented.

This view is quite simplistic. First, it is unlikely that the human mind goes through these steps separately to come to a decision, so at least the visualization with blocks is sub-optimal. Second, it has been shown that the selection and comparison of options, the *decision event*, is unlikely to happen in real-life scenarios [70]. Instead, experienced decision makers seem to generate, in a process of pattern matching, just one intuitive option that is based directly on the available cues. In this *recognition-primed* decision model, the option is evaluated by a mental simulation and, possibly with small a adaptation to the situation, applied. Only when a decision maker is unfamiliar with the situation, that is, there is a significant mismatch between the received cues and the generated option, is a conscious process of evaluation and mental simulation of different options triggered [57].



(a) A generic but insufficient cycle for decision making [57].



(b) A universal model for a decision-making process, based on naturalistic decision making. Close to the dynamic decision loop presented by Brehmer [12].

Figure 4.1. Two alternative visualizations of a decision-making cycle.

Figure 4.1b shows a different visualization of the decision-making cycle. This view avoids the mentioned problems. Still, it is not intended to be an exact model. It shows human decision making as a process with four stages: perception, comprehension (these two stages correspond to identification of cues and generation of options in Figure 4.1a), evaluation (integrating selection and comparison of options), and action. Instead of showing a process with clearly separated stages, the stages are merging, as it seems that they are not conclusive. The generation of an option can, for example, start while new information is still perceived.

The core of Figure 4.1b depicts the decision-making process as a whole. On the horizontal axis runs the time scale, which ranges from the start of a decision-making process to the point when a decision has been made. The vertical axis describes the probability with which cognitive resources are spent. The diagram shows tendencies of how likely it is at a certain time during the decision-making process that cognitive resources are spent on one of the four stages. In essence, it expresses that at the beginning of the decision-making process, perception of information as cuing for a decision is much more important than, for example, evaluating (different) options for a decision, that is, that more cognitive resources are likely to be spent on perception. On the other hand, at the end of the process, formalizing the decision, and possibly preparing for an action, is much more important than perceiving new cues. If an additional important cue is perceived 'late' in the process, that is, when an option has already been evaluated, the amount of resources spent on perception can be increased again. But for simplicity, Figure 4.1b only depicts the idealized process.

Kahneman and Tversky describe human beings acting based on the interaction of two systems. System 1 can rapidly and effortlessly identify situations and trigger adequate actions. System 2 instead allocates cognitive resources to more deeply analyse the current cues and actions suggested by System 1 [50]. Their studies describe many situations where System 1, prone to biases and fallacies, leads to sub-optimal or even incorrect judgements and decisions.

Todd and Gigerenzer follow close to the idea of Simon. Decisions have to be made 'using realistic amounts of time, information, and computational resources' [91, p. 741]. For them, decision making comprises two phases: searching alternatives and selecting one alternative. Heuristics are tools to deal with bounded resources; they guide the search, decide when to stop it, and control the actual decision making. Additionally, they favour formal modelling of decision making, as this will contribute to the theoretical foundation of the field by increasing the clarity and falsifiability of suggested models [92].

4.1.2 Dynamic Decision Making

Brehmer and Dörner developed a different way to analyse decision making [11, 13, 12]. They suggest microworlds which are computer simulations of complex, dynamic scenarios. The subjects involved in the studies have as a task to control these settings with or without certain given goals. Microworlds allow studies halfway between laboratory and real-world experiments. They compromise between the advantages of complex, opaque, and dynamic scenarios in field studies and the advantages of controllable conditions and faster generation of results in laboratory settings [11, 36].

The interest lies in dynamic scenarios. Here, especially, feedback on actions influencing the state of the microworld and delays in this feedback are relevant for the decision makers' performance [12]. This marks a main difference from the more traditional research done by, for example, Kahneman and

Tversky or Todd and Gigerenzer but places dynamic decision making closer to naturalistic decision making. Thus, for the domain of our research – the control of complex, dynamic processes – the work of Brehmer and Dörner is most relevant. Indeed, the authors also suggest control theory as a metaphor for human control and describe human control in terms of goals, models, observability, and action [11, 13], which is the basis for the GMOC model described later in this thesis (Chapter 5).

Despite the interesting possibilities and the enthusiasm in the work by Brehmer [11], the use of microworlds to analyse human decision making has been limited (cf. [36, 12]). For our purposes – the analysis of human work – the utility of microworlds is limited as well. A critical factor is creating faithful models. Someone who is familiar with a scenario simulated in the microworld might perform badly if the model used by the microworld itself is flawed. However, this was not the purpose for which microworlds were invented. Instead they are an interesting tool to further investigate dynamic decision making and possibly also mental models and their construction – at least, if we are able to relate models in mircoworlds to real-world scenarios, we can draw conclusions on dynamic decision making and how to support it.

4.1.3 Relation between the Different Approaches

A review of decision-making literature makes clear that many different approaches to research on decision making exist. The views differ widely between the context of the decision making and the aims of the research, but there are possibilities and attempts to identify similarities between the different approaches. Examples are Brehmer, who identifies similar tendencies between dynamic decision making and representativeness heuristics [11, p. 89] and the 'failure to disagree' in Kahneman and Klein's attempt to explore differences between heuristics and biases and naturalistic decision making [51] (see also [56]).

As the main differences between approaches to decision-making research are the settings and methods on which the studies are based, the implication I see is simply that different conditions have a reasonable impact on the way decisions are made and on their outcomes. This view conforms with Orasanu and Connolly:

> Decision performance in everyday situations is a joint function of two factors: (1) *features of the task*, and [(2)] the subject's *knowledge and experience* relevant to that task. [70, p. 7]

For instance, when a decision is made under strong time pressure, it is more likely that a fast heuristic will be applied. If the decision is made by experienced persons who have been able to learn from feedback on their decisions, biases will be less likely. If stakes are high, it is more likely that more sophisticated approaches with more careful evaluation of options will be chosen – especially if time pressure is low and the decision maker lacks experience.

At this point, I want to make a critical remark regarding too restricted lab-based studies of human decision making. In the literature (among which is Kahneman's seminal book *Thinking Fast and Slow* [50]), one can find many examples of experiments showing that test persons perform surprisingly poorly. Some of these experiments seem too constructed without adequate representations in our natural environment – it is no surprise that people fail at tasks in which they are not trained if they are constructed to aim for failure. However, I do not say that these results are not interesting. In fact, they illustrate very well what happens if computer systems are designed badly, that is, in an arbitrary way that does not take users' characteristics and behaviour into account.

So, to utilize the results of the different themes in decision-making research in system design, I suggest identifying the conditions under which decisions will be made. One can consider, for example, the dynamic properties of a scenario, the expected expertise of the users, and the nature of the tasks (e.g., whether they are likely to lead to fallacies). These properties can then be related to the appropriate research results to find and avoid probable pitfalls – scenarios where users have 'no other choice' than to fail. Another implication for system design is the necessity of 'valid cues and good feedback' for developing 'skill and expert intuition' [51, p. 524]. This comes close to recommendations in research on dynamic decision making [11] and is one key aspect of observability and the interplay of the elements in GMOC (see Section 5.2). The complex and dynamic properties of our research application explain why dynamic decision making has been most influential for our research.

4.2 Mental Models

Mental models can be seen as the component that explains how we can meaningfully interact with systems and environments, how we can understand them, and how we can learn about them. The ability of a system to support construction of mental models is a main factor of its usability. Thus mental models are quite a central theoretical concept in HCI. With a close connection to cognitive psychology, the concept appeals to both researchers and practitioners, but when it is discussed from a theoretical perspective, there is a consensus that the term is used widely but vaguely (cf. [8, 74, 100]).

4.2.1 Defining Mental Models

Bainbridge defines mental models in general:

> Cognitive psychology is concerned with understanding tasks in which a stimulus is processed in some way before a response is chosen; the brain of the

> person doing the task contributes something which is not the original stimulus. [...] In many tasks, knowledge about structures or cause-effect relations which underlay what can be observed play a central part in cognitive processes. When doing the task uses knowledge about the state(s) of a potentially changeable world, these structures of knowledge may be called mental models. [8, p. 119]

Thus, whenever we act, and this action is based not only on the perceived information but also on knowledge about the environment, we use mental models. This is indeed a very broad definition but erects a frame for discussion of the term. Consequently, Bainbridge proposes further themes to specify one's view on the mental model: (1) its role in cognitive processing, (2) the kind of knowledge it represents, (3) implementation of the knowledge (more specifically, its accuracy), (4) affecting factors, and (5) our model, in the role of the researcher, of the user [8].

The first theme falls within cognitive psychology and so is beyond the scope of this thesis. Without being too specific, I can say that I see mental models as persistent knowledge that will be accessed if cues from the environment trigger this. The knowledge can then be used for certain tasks – for example, to predict a development or to decide on an action – and even altered. This leads to properties of mental models that are discussed in Section 4.2.2; this section covers themes 2 and 3 as well.

Our purpose is to utilize the concept of mental models in system development. Ways to develop (i.e., construct and improve) mental models belong to the fourth theme. This is covered in Section 4.2.3. I will not take the perspective of cognitive science but rather will discuss factors that influence development, which is most relevant for the utility of mental models in design, covered in Section 4.2.4. Themes 4 and 5 are very relevant for us if we want to analyse mental models. These are covered in Section 4.2.5.

4.2.2 Properties of Mental Models

Norman summarizes his observations on mental models as follows:

> 1. Mental models are incomplete.
> 2. People's abilities to "run" their models are severely limited.
> 3. Mental models are unstable: People forget the details of the systems they are using [...]
> 4. Mental models do not have firm boundaries: similar devices and operations get confused with one another.
> 5. Mental models are "unscientific": People maintain "superstitious" behavior patterns [...] because they [...] save mental effort.
> 6. Mental models are parsimonious: Often people do extra physical operations rather than the mental planning that would allow them to avoid those actions; they are willing to trade-off extra physical action for reduced mental complexity. [33, p. 8]

Though these findings are interesting and I mostly agree, we have to consider that Norman usually deals with the 'typical' user. The relevant users in our domain are trained experts controlling complex, dynamic processes. By combining Norman's observations with those of Bainbridge [8], experiences from dynamic and naturalistic decision making [11, 57], and observations from our own research, I see the following as the most important properties:

Imperfect... Mental models are imperfect in that they can be (and usually are) incomplete and even wrong. Reasons can be that there is no obvious need or possibility to learn and maintain knowledge of certain behaviour and details about an environment. It can also be that operators understand a system completely incorrectly, if this understanding still explains the observed correlations between input and output.

...but efficient Mental models are abstractions of the system and environment, which makes them imperfect but efficient. They typically give the right response with reduced mental effort. (This correlates very well with Norman's points 5 and 6.)

Developable It is possible to construct new and develop existing mental models. If operators, for example, cannot explain observations or the systems' behaviour, there is the opportunity to develop mental models. (See also Section 4.2.3.)

Adaptable If operators approach a new system, situation, or mechanism (i.e., they do not have a developed mental model), they can use existing mental models as analogies.

Completely different Very much related to imperfection is the fact that two operators with the same control task and environment can have completely different mental models about (parts of) their environment.

Runnable Despite Norman's second observation, I see that operators are able to run their mental models. The reason for this difference is likely to lie in the dynamic nature of process control. By 'running', I mean that they have the ability to produce predictions about future states of the system, predict outcomes based on their actions, and infer earlier actions or events from current observations.

I do not see any specific boundaries on what a mental model can represent. In process control, I see mental models covering the complete environment, including organization, tasks, systems, processes, rules, natural laws, behaviour of co-workers, and so on. It is arguable if these objects are represented in different ways, but I do not aim to make any specific claims about how mental models represent anything.

4.2.3 Construction of Mental Models

Mental models are not fixed. Every action and observation of a user can influence her mental models. There are three ways in which users can construct or develop their mental models about the system: (1) applying and developing already existing mental models, (2) interaction with and observation of the system, and (3) training, manuals, and explanation. The first point is the utilization of already existing mental models, but their application in a different way or environment. This kind of construction is thus heavily dependent on the individual operators' previous knowledge. It can be supported, for example, by using metaphors in design.

The second way is heavily dependent on the system and the environment and is a central point in Norman's discussion of mental models [33], but also in the field of dynamic decision making [11]. A system's reaction to operators' input gives certain feedback. This can either be consistent with the operators' mental model (i.e., an expected response) or not. Consistencies will make the operators more conscious. Inconsistencies might be alerting, and the mental model might suggest that there is a certain problem in the process, or it might point out a deficiency in the mental model. Operators who ascertain the reason for inconsistencies have the chance to develop their mental model.

The third way is an external influence on the operator. This can take the form of written (or illustrated) instructions and manuals or can come from another person, for example, in training or through suggestion or explanation by a colleague. The form of this kind of instruction seems to have a considerable impact on the development of the mental model. An example is the work by Halasz and Moran. They trained a test group using a calculator with additional material intended to support construction of an adequate mental model. As a result, this group performed significantly better (i.e., made fewer mistakes) in solving more complex tasks [38].

4.2.4 Utility of the Concept of Mental Models

According to the preceding definition of mental models, they are a part of most meaningful interactions with an environment. This shows their relevance to our goal: supporting analysis and development of human control of complex, dynamic scenarios. For this purpose, a detailed explanation of mental models and their exact function in cognitive processes is less important. Instead, I provide some examples of utilization of the concept.

Norman discusses mental models in the context of system design:

> The problem is to design the system so that, first, it follows a consistent, coherent conceptualization – a design model – and, second, so that the user can develop a mental model of that system – a user model – consistent with the design model. [66, p. 46]

So if we, as designers, have a better understanding of the users' mental models, we should be able to design systems that are easier to learn and more efficient to use than systems designed without a good understanding of the users' mental models. And based on the three ways of developing mental models, we can add that we can design the system in a way that is either consistent with existing users' mental models or so that the design itself supports development of mental models (cf. also [59, p. 58]). Or, by the third way of learning, we can support the users with adequate instructions. Chapters 5 and 11 give more concrete examples of how to design work environments that support the development of mental models. Bainbridge gives similar recommendations for considering mental models in system development, but she adds that it might be more beneficial 'to include other aspects, such as pattern handling, cued recall, the time to translate from one representation to another, the non-mathematical optimum use of evidence, and physical activity' [8, p. 140].

The other important aspect of utilizing mental models in system design is identifying problems in current work environments. When we analyse mental models, we can draw conclusions about their complexity, their accuracy, the information that is needed to 'run' the model, and the results (e.g., in the form of actions). Problems can, for instance, be that a mental model seems overly complex or inaccurate or that the information is unavailable or an action impossible to implement. This gives us clues for further investigation.

4.2.5 Models of Mental Models

Both Norman and Bainbridge point out the importance of stepping back and considering the process of analysis. When we investigate users' mental models, we develop our own models (which have the same properties as mental models). Bainbridge suggests relating models of mental models clearly to the purpose of the investigation, as this purpose angles the model and thus might make it useless for persons with other purposes [8]. Norman summarizes three factors that help to verify correspondence of the researcher's model to the analysed mental model [33]: (1) it should contain a model of the relevant parts of the person's beliefs; (2) just as aspects and states of the mental model should correspond to the physical system, the observed model should correspond to the system as well; and (3) the researcher's model has to include details about how the users are able to 'run' their systems to generate predictions.

When we analyse work environments, operators' mental models are an important part of this analysis (see Chapter 5). Our goal is to understand the mental model embedded in the work context, aiming for identification of problems. Certainly important for us is the dynamic information that operators' need to run their mental models. We analyse the work environment, including the operators' mental models, using methods based on ethnographic principles (see Chapter 8).

4.3 Situation Awareness

During the past twenty years, situation awareness (SA) has been an important topic for research in human factors. There have been efforts to establish SA as a theory [26], but it has also been called a construct or a folk model [21]. This section describes SA, compares and discusses some of the opposing views, and concludes with a statement about how SA is understood in this thesis.

A generally accepted definition says that having SA means ‘knowing what is going on’ [26]. Well known and commonly used is Endsley’s definition:

> Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future. [25, p. 792]

She further introduces a three-level model [26]. In this thesis, mentions of SA refer to this model, which can be summarized as follows:

Level 1 Perception Operators perceive the important information, meaning that the needed information is available and that attention is directed to its important pieces. Without attention to the right information, operators have no chance to understand and control a system.

Level 2 Comprehension This level describes the ability to understand the perceived information. It is the process of interpreting information towards one’s goals and to remember relevant parts of information.

Level 3 Projection Perception and comprehension of the relevant information allows to understand the dynamics of a system. This will allow projecting future states and support decision making to influence and control the state of a system.

Approaches to SA differ between disciplines; whereas psychologists such as Endsley see SA solely as an attribute of the individual, engineers see SA as an attribute of the environment, for example, in the used systems [87]. I agree with the view of system ergonomics to the extent that one has to consider the whole socio-technical system as a source of SA. SA can be distributed over several agents in such a system [88]. However, I would like to point out that I see systems and the environment as mere sources for an individual’s SA, that is, that a system cannot ‘have’ SA, it can only provide information that can contribute to someone’s SA. Therefore I disagree with Stanton’s view on this point.

Despite its widespread use and application, SA has continuously been a focus of criticism. Quite early, researchers had pointed out that there were problems in using the term, leading to circular reasoning:

> How does one know that SA was lost? Because the human responded inappropriately. Why did the human respond inappropriately? Because SA was lost. [32, p. 151]

Loss of SA seemed to become a common explanation for human error, so Flach warned about seeing SA as a cause for anything. Instead, he saw more value in exploring whether (loss of) SA could be used as a category for a certain type of incident, which in turn might be helpful for understanding these types of errors. Another positive aspect of SA is that it points out the need for a holistic view of human performance and 'the inseparability of situations and awareness' [32, p. 152]. This implies that SA has to be seen in the light of the operators' current goals and demands on them. Thus, because SA is highly dependent on the situation, this also leads to the conclusion that (experimental) assessment of SA has to be done in environments that are as realistic as possible.

Years later, this criticism has been renewed [20, 21], showing that – despite the early warnings – there still is a need to clarify what SA is and what it is not. Dekker and Hollnagel [20] see SA as a 'folk model', meaning that it explains by substitution. An example is an accident that is explained by 'lost SA'; SA, as a seeming commonsense construct, is used as a substitute for a proper explanation. They also see immunity to falsification and over-generalization as typical problems of folk models and problematic in the use of SA.

This criticism does not stand without opposition. A main argument supporting SA is the existence of a large number of sound empirical data on SA. According to Parasuraman, Sheridan, and Wickens [73], these data are obtained by researchers who have a clear picture of the theory behind SA, while improper use of the term and seemingly circular reasoning are characteristics of 'consumers' of that research. These are valid points, but still the research community has a responsibility for the over-generalized usage of SA (cf. [21]).

At last, there is the argument of availability of measurement methods. On one side, they are seen as adequate because there always exists a 'ground truth' (i.e., the current situation) against which an individual's SA can be tested [73]. But, as Dekker [21] emphasizes, applying this ground truth in measuring SA is not trivial – it requires an 'all-knowing', neutral observer – and not at all a legitimate basis for judging people for loss of SA.

I definitively agree with Dekker that accountability cannot be based on a concept such as SA. However, similar to Flach, I see some value in it. In the design of systems and user interfaces, especially for control of dynamic processes, paying attention to SA is important. A system should always be designed in a way that it displays all critical information, reflecting the current situation, and supports the operator in gaining awareness of 'what is going on'.

4.4 Discussion

This chapter has summarized major themes in HCI and human factors with relation to manual control. A main task of operators in control of complex, dynamic processes is to get and keep an overview of the process and to interact with it according to certain control goals. In terms of the discussed concepts, operators have to gain situation awareness and make control decisions. Mental models can (with reservation) be seen as a central component of both SA and decision making.

4.4.1 Relation between Situation Awareness and Mental Models

The first part of the claim is easy to explain. Endsley herself provides an explanation, suggesting SA as a tool for analysing the mental model [28, 26]. In relation to Endsley's three levels of SA, a mental model has access to the perceived information and determines the information to attend to (perception), it is involved in interpretation of the perceived information (comprehension), and it is necessary to derive future states of the system (projection). This allows us to interpret mental models as the 'static' structure[1] lying behind the construction of SA. SA can thus be interpreted as the current configuration of a mental model (cf. [100, p. 62]). For this work, this implies that an adequate mental model is necessary for development of SA. Thus developing systems with the goal to support the development of mental models will also facilitate SA.

4.4.2 Relation between Decision Making, Mental Models, and Situation Awareness

Endsley mentions the significance of SA as a precondition for dynamic decision making [26]. Indeed, it is the dynamic scenario that builds the common ground between SA, dynamic decision making, and naturalistic decision making. It is the model of recognition-primed decision making (RPD) [56] in the context of naturalistic decision making that directly acknowledges this connection: "In the RPD model of time-pressured decision making, situational awareness becomes very important" [57, p. 201].[2] SA allows the operator to generate options (or the one most reasonable option) for decisions. The basis for the option generation is the operators' experience of similar situations they have encountered earlier.

Continuing this process leads us to an acknowledgement of utilization of mental models in naturalistic decision making: when options are selected and applied, they often go through a process of mental simulation and have to be slightly adapted to fit the current situation [57] – a process that is quite

[1] Static in this context is to be seen as relative, as we just have discussed that mental models can be and are continuously developed. However, SA develops basically every moment, whereas development of mental models is a slower process (cf. [95, p. 398]).

[2] I have to mention that it is not completely clear which concept of 'situational awareness' Klein [57] uses. However, in the postscript, Klein does not object that it resembles the concept of situation awareness that also has been the basis for the discussion in this thesis.

characteristic of using mental models (cf. [74]). The importance of mental models in dynamic decision making is of no doubt. They satisfy the 'model' precondition in human control:

> Both in the case of human and automatic controllers, the model can, of course, take ma[n]y forms, but all of them will have the effect that the controller will behave as if he/she/it has a model with some specified relation to the system it controls, and the level of control will depend on the extent to which the model includes the important aspects of the system to be controlled. [11, p. 217]

Furthermore, there is a suggestion that this type of analysis is an action of System 2 [56], which might allow us to relate mental models to the systems thinking Kahneman [50] proposed. The connection seems intuitively logical but is not explored further in this thesis.

5. GMOC

The GMOC model describes human work in complex dynamic environments. It plays a central role in this thesis and can be seen as a continuation of Brehmer's work on dynamic decision making [11]. The model has been applied during many studies done by the HCI group at Uppsala University. It has been used for analysis of human work and design and evaluation of new systems for human control. Examples are operation of high-speed ferries [68], train traffic control [5, 54], and train driving and truck driving [48]. This chapter is an attempt at a more formal definition of GMOC, a goal shared with Paper I.

5.1 Utilizing Control Theory for Human Control

GMOC is an acronym for goals, models, observability, and controllability. It follows the basic idea that these four elements are the necessary prerequisites for human beings to achieve control over a task and a system. This is closely related to basic concepts in control theory [52]. A key application for control theory is automatic control (i.e., by a technical system) of dynamic systems. Probably because of its technical, mathematical background, control theory has drawn very little interest on application of its frameworks to the work of human beings. This is understandable, as technical control often follows well-defined rules and goals, whereas human beings often follow changing, fuzzy goals and apply rules differently based on their own interpretation, experience, and mood. Still, human control of complex, dynamic environments puts human operators in quite similar environments as technical regulators in control systems. They often have to collaborate with such technical control systems as well.

In human control, as in control tasks in general, observability and controllability are dependent on the environment and the system through which control is done. But in contrast to a regulator following defined mathematical equations, human operators use their capabilities to dynamically define and prioritize goals and following mental models that can develop over time to improve and adapt to new and unpredicted situations. We therefore follow and continue the work of Brehmer and Dörner [11, 13], who define goals, models, observability, and action as prerequisites for dynamic decision making.

5.2 The Four Elements

Figure 5.1 shows the four prerequisites for human operators to efficiently achieve control over a process: operators need a clear goal in the form of a state to reach or keep a system in. They need a model of the process, the system, and the environment to assess the current state of the system and to predict future states resulting from the process dynamics and executed control actions. The system has to provide observability, allowing the operator to efficiently assess the current state of system, environment, and process, as well as controllability, allowing the operator to manipulate the development of the process.

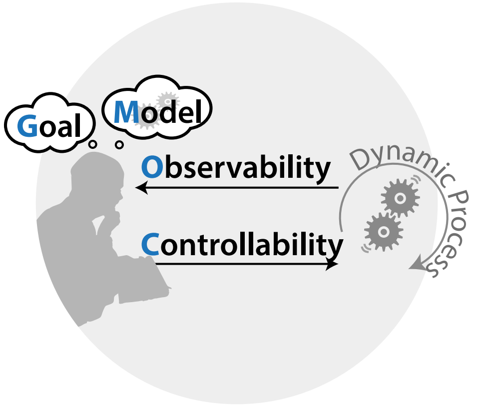


Figure 5.1. In a complex, dynamic control environment, the human operator needs clear goals and adequate models, while the system has to provide sufficient observability and controllability to allow efficient control.

Following the described theories from control theory, we claim that these four prerequisites are sufficient to exhaustively describe a human operator's work environment in process control. Our main interest is to use this complete description to support the development of new, or improvement of existing work environments for human operators. Central for our influence on the control environment is the system used by the operators to perform their tasks. This allows us to change observability and controllability. But our influence is not limited to this. Goals and models can be influenced by the organisation in which the control task is embedded, and they are also related to the control system. This means that we even influence the development of goals and models by changing observability and controllability in the control system and environment. However, to do so, we have to assess the degree to which four prerequisites occur in the control environment (including the operators). Assessment of the four prerequisites in human control is a much more complex task compared to the well-defined technical domain. Before going further, the four elements are described in detail – in the following, the prerequisites are called elements, as they are the fundamental, elementary components of GMOC.

5.2.1 Goals

The goal is the basic specification of what the operator will, must, should, or wants to achieve – a specification of the objectives of the control process. Without a goal indicating what should be achieved, no purposeful control is possible. As simple as this description sounds, as complex does the term *goal* become on closer examination.

In decision-making research, the relevancy of goals and their influence on the decision is recognized. However, the definition of goals is often something intuitive and given. Dörner, as an exception, does comprehensively discuss goals in the context of dynamic decision making (see [23]). In the concept of situation awareness, goals are an important element. They define which information actually contributes to situation awareness. Measurement methods, such as SAGAT, therefore can be based on a *goal-directed task analysis* [29]. Goals are an important element in GMOC. Our research has shown that goals are important to understanding the operators' selection of models and actions. At the same time, goals and sub-goals develop complex structures. It is important to see through these structures during analysis and to design systems that support users in organizing goals.

Our findings on goals correlate very well with Dörner's [23] description. According to Dörner, goals can be about reaching or avoiding a state, defined by few and fuzzy constraints, or very specific, including many clear criteria. Often goals are split into several sub-goals to make them easier to handle. Goals can be conflicting and of different importance. With an increasing number of goals, the possibility of conflicts rises. Additionally, goals are dynamic. The development of the process will change the urgency and importance of certain goals and may lead to a resolution and eventually to formation of new goals.

During experiments with micro-worlds, Dörner observed that it is common for human beings to focus too much on one sub-goal, leading to situations where other sub-goals are neglected or realized too late, with severe consequences for reaching the overall goal. A good understanding of the different goals and their relations and conflicts is necessary for an operator to take the right decisions. An operator in complex, dynamic environments does usually not have the time to elaborate on goals to find the optimal solution, which brings us back to Simon [86] and the topic of bounded rationality.

The motivation of goals is an aspect that is important in real-life scenarios. They are mainly of an organizational or personal nature. Organizational goals are typically those that are related to the outcome of the process and primarily beneficial to the organization, whereas personal goals are primarily beneficial to the operator. Often the resulting goals are not conflicting, for example, the personal goal of performing well (to earn a bonus) is usually oriented at the organizational goals. Further goals can be imposed on a social layer, by co-workers or family. In systems with several actors, it is important to ensure that goals are not in conflict.

In conclusion, systems should be designed to support the whole process of creation, prioritization, and de-composition of goals and understanding of their relations. This can be clear visualization of information that is related to goal achievement or setting of parameters that inform the system about users' goals. If we take the GPS navigation of a car as an example, we can set where we want to go and if we prefer a fast or an economical route. However, if our goal changes during the journey – we might get a call that we should arrive earlier – we have to reconfigure the GPS navigation. Depending on the situation and design of the system, reconfiguration might conflict with the goal of arriving safely.

5.2.2 Models

The model element refers to the operators' understanding of their work, the systems, the process, their co-workers, the organization, or in short the entire work environment. To a large extent, models are mental models as described in Section 4.2. We often use these terms synonymously, but models as part of the GMOC model actually go beyond the definition of mental models. In this case, they refer to models of the work environment in the widest meaning as models that operators (logically) need to make sense of the environment and to interact with it in a meaningful way.

When we analyse a work situation, we will face operators who have developed their own kinds of models and expertise. Often they have some kind of declarative knowledge available, a description which they are able to verbalize. However, we cannot assume that such verbalizations are a perfect description of the operators' models. Taking Rasmussen's model of skill-, rule-, and knowledge-based behaviour as an example, knowledge-based behaviour can be observed when operators are confronted with unfamiliar situations in which they have to plan their actions based on their knowledge; rule-based behaviour means that they have developed and can apply rules or procedures from previous experience; and skill-based behaviour represents unconscious and automated actions. Rasmussen states that a person is 'unable to describe how he controls and on what information he bases the [skill-based] performance' [80, p. 259].

We aim neither for a perfect definition of the term *model* nor for a thorough explanation how models function and how they are represented in the human mind. Instead we aim for a working definition in the GMOC model and its integral role in a work environment. When we analyse and design work environments, it is important to recognize how the environment influences the construction of models.

As mental models, models are often neither a complete nor an entirely correct description of the operators' environment, but they allow for efficient performance of common tasks. Models help operators to interpret the information available in the environment. With this interpretation, and guided by goals, operators can decide on actions to achieve their goals. When approaching unfamiliar situations which are not entirely covered by existing models, operators can develop new models. They may start from a model that already exists in their mind and try to construct hypotheses about the situation. By empirically testing their models with such hypotheses, operators can further develop their models, which is elementary for learning and gaining expertise. This flexibility of models is important for operators to handle unexpected situations and identify failures.

During analysis, models are important to understanding operators' actions. As they reflect their understanding of the system, they are also key in the development of future systems. Firstly, if we can identify flaws in models, we can probably find deficiencies in the work environment. This can be that a system is not transparent enough and therefore leads its operators to draw incorrect conclusions, or that they have knowledge that goes beyond the system which can be used for improvements (e.g., if a system is inefficient and operators know how to make a process more efficient). Secondly, when we design new systems, we should be aware of existing models to design the new system in a way that is easy to learn and not design components in misleading ways (i.e., appealing to inappropriate models). Thirdly, if we have an understanding about how models are constructed, we can support development of models and expertise by system design. Fourthly, when we design new systems, we can motivate components of new systems and interfaces with hypotheses about how these components will be used and understood. This means that we can base our design on hypotheses about models, which we can evaluate later, when the system has been in use.

5.2.3 Observability

Observability refers to the ability of the system to produce information that can be sensed by the operator. Most common is the visual observation of information, for example, by a direct view on the process, computer displays, signal lights, analog and digital instruments, or lever and switch positions. Sound is often used for alarms or notifications that require some kind of action or acknowledgement. If the workplace of an operator is close to the process itself, it is often also possible to observe its state via aural and haptic feedback.

It is a common opinion that too much information is negative, as it leads to information overflow. Endsley, for example, describes the ‘information gap’ [29, p. 4], where she discriminates between the number of data a system can produce and the information an operator actually needs. She claims that large numbers of data require an extensive search of relevant information and thus leave the user less informed.

These points are quite valid. However, it is important to keep the balance and the purpose of the system in mind. We have encountered several situations where operators did not have the needed information available; this was sometimes caused by technical limitations but also because developers of the original system deliberately hid information to achieve what they believed was a user-friendly design. The result is a lack in observability that requires the operators to develop additional models that help them derive the needed information from the actually available information. Instead of hiding the complexity of a process by hiding (too much) information, we suggest appropriate and careful designed visualization of this information (see Paper II).

In GMOC, observability follows two main intentions. The first is to raise awareness of the fact that operators might need more than the visual information that is available via screens, which is of particular importance when new systems become more isolated from the original process. The second is to raise awareness of the fact that observability is needed not only to help the operators decide on their actions but also to develop and evaluate (sub-)goals and models.

5.2.4 Controllability

Controllability refers to the ability of the system to provide means for the operator to gain control over the process. It means that the system offers mechanisms for interaction, for example, physically via keyboards, buttons, levers, or steering wheels, or any other way, such as voice control. The mechanisms for interaction need to be perceivable, known, and accessible to the user. In addition, the system has to react in deterministic ways to control actions – without predictable results, control is not possible.

Another dimension is the extent to which a process can be influenced by the offered means for interaction, and the accuracy in doing so. If not all (input) parameters of a process can be controlled or if the accuracy or range to control certain parameters is too low, this will result in low controllability. Analogously, who would be happy with a toaster if it could not be manually stopped when the bread starts to smell burned, if its controls would only allow settings where the bread either remains cold or gets burned, or if its timer can only be set to a value so small that a slice of bread has to be toasted several times to achieve the desired browning?

Related to controllability, but strictly speaking a part of the observability, is the feedback that the system produces as a response to a control action. The first step would be an acknowledgement that an action has been registered; the second step would be feedback on the result of an action. As a rule of thumb, one can say that the closer (in time) feedback follows an action, the easier it is for the user to associate response and action. If a process is of a nature that response delays are unavoidable, the system might have to deliver additional reminders or cues that a response to an action is still expected.

It is especially interesting to look at systems that include automation. Automation often sits between operator and process and takes over a part of its control. Therefore it can have a huge impact on controllability. Some processes might not be controllable at all without automation, for example, if some (re-) actions have to be executed very quickly or accurately. In other systems, situations can occur where automation acts against the intention of the human operator and thus reduces controllability [9]. Automation might also encapsulate the original process controls and so alter the nature of the operators' task to control of automation rather than a process. 6 discusses further aspects of automation.

5.2.5 The Interplay of the GMOC Elements

As important as each element separately is the interplay between them. Each element has to be geared towards the others to construct an efficient work environment. Similarly, it is almost impossible to analyse the four elements separately, because there are strong dynamic effects between them. The model is the central element in GMOC and can be seen as mapping goals and observability to controllability.

If operators encounter a work situation, they have certain goals and their experience in the form of (mental) models, and they are confronted with a system offering observability and controllability. Observability provides the information that can be used to interpret the current state of the system. Models interpret observability and decide on appropriate actions, using a subset of controllability that promises the most success towards attaining their goals. Goals and models influence observability and controllability as they can direct attention towards certain sources for observation and means for control – if our goal while driving a car is to hold a certain speed, we would rather check the speedometer than the fuel gauge, and we would rather use the throttle and brake pedal than the hand brake.

Models are also needed to break goals down into sub-goals. This might be necessary if a goal cannot be directly related to observability or controllability. As long as we do not have an 'autonomous' car, we cannot control its destination directly; instead, we can use the steering wheel to control if the car goes straight or turns. Thus our model of the environment (e.g., in the form of a map) will set sub-goals such as turning left or right.

As already mentioned, models hold knowledge and experience. But we need the other elements to support the development of adequate models. Appropriate feedback to control actions is an important aspect for the ability of the operator to develop an adequate model. Thus observability has to reflect controllability. If a means to control the system (e.g., a switch) is neither part of an existing model nor providing feedback allowing the development of such, the control will simply be disregarded. However, if the switch produces clear feedback, for instance, in the form of a dashboard symbol, and possibly direct feedback from the process, operators can form a model according to the switch and, by understanding its function, even integrate it into existing models.

5.3 Application of GMOC

GMOC describes the four basic elements that are necessary for human control of complex, dynamic processes. It helps us find deficiencies in a system. These can lie in any of the four elements or in their interplay. As described in the previous sections, such shortcomings have a negative impact on the work situation of the human operators; this can be inefficiencies in the process control, difficulties in keeping the process in certain states to reach desired results, or difficulties in gaining expertise. The main goal of applying GMOC is to find solutions that eliminate the existing deficiencies to create a better, namely, a more efficient and satisfying, work environment.

In practise, the first step in applying GMOC is to analyse the current work situation. The resulting description should help in sufficiently understanding tasks and conditions of the human operators' work. The next step of analysis will focus on finding deficiencies in this situation. Theoretically, the best way would be to describe an optimal description as well and to identify discrepancies between this and the current situation. However, as it is possibly impossible to define such an 'optimal' situation, a reasonable compromise in most situations is to classify the deficiencies according to GMOC and suggest areas for improvement.

During the design, the need and possibility for improvements in certain elements will be determined closer. It is good to start with an open-minded, unlimited view in the design process. This means not to start considering today's environment, systems, and way to work, including the operators' understanding of the environment and technical limitations. Instead one should consider the overall goal of the task and key elements of the work process. This should lead to an envisioning of the future work situation. Realization of this vision related to the current conditions will be part of the later design process.

Chapters 8 and 9 present a process and methods to support system development with GMOC. Some key points of such a process are as follows:

User involvement Users or operators should be involved in each phase of the process. The focus lies on understanding the users' work, and the role of the users is to ensure that this understanding is correct. In the beginning of the design, however, the users' role should be more passive, as the users' view usually is too oriented to the current work situation and thus limits the envisioning process. It takes time for users to let their current views go.

Involve different roles If possible, persons in different roles should participate in the process. This includes persons higher up in the organization's hierarchy to ensure that the organization understands and supports the system's development, as well as representatives from the system development team to ensure that ideas about and requirements for the system generated during the design process will be implemented correctly in the final system. All participants should have a basic understanding of GMOC.

Iterative and continuous evaluation and deployment The analysis and design of a system with GMOC has as a goal to change the work environment. It is not possible to completely change an environment and the operators' approach to their work in a short period of time, nor is it possible to predict each effect a new environment will have (e.g., a change in controllability might imply different requirements on observability). Therefore the introduction of new systems and work routines has to be divided into smaller, feasible steps that should be evaluated continuously.

Large changes take time At its core, GMOC is a quite scalable process. If time and resources for a project are limited, one can focus on certain aspects of the work environment instead of envisioning a completely new work environment. However, if done thoroughly, the analysis of an operator's work can take months.

5.4 GMOC at Different Organizational Layers

The main goal of GMOC is to describe human work, where, naturally, the individual is the focus. In this view, goals and models always belong to the individual. And even observability and controllability – as characteristics of the system – are connected to the ability of the individual to attend and understand any information or means for control offered by the system.

Beyond the view of the individual, there exist two additional layers (see Figure 5.2). The first layer at the top is the individual. The second layer represents collaboration, and the third layer represents the organization. This is a socio-technical system, as discussed in Section 2.2.1. The collaborative layer consists of any number of individuals working in the same context. These individuals can have the same or different roles in a given context – they can even be part of different organizations. An example is the collaboration between train drivers and traffic controllers, who work on the same process – train traffic – but from completely different perspectives. In such scenarios, conflicts between goals are not uncommon but are problematic. Sharing and consideration of others' goals are therefore quite important. Additionally, operators develop mental models of collaborators and interdependencies between roles and organizations. Observability and controllability can be extended via communication between the collaborators, for instance, in asking colleagues about their observations of a process. Paper IV further discusses the team layer.

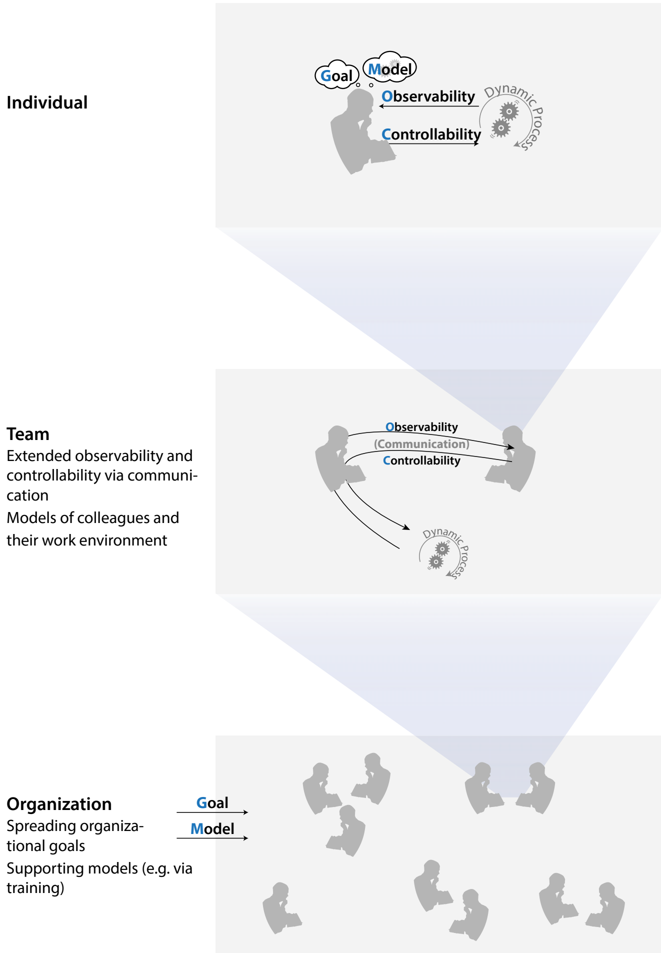


Figure 5.2. Meaning of the GMOC elements in different layers of a socio-technical system: individual work, work in a team, influence of the organization.

The influence of the third layer, the organization, is quite different. An organization has the possibility to influence individuals' goals by communicating the organizational goals which they want their employees to adopt. Additionally, an organization can offer training supporting the operators' learning, that is, construction of (mental) models. These two influences can have a direct impact on operators' efficiency and should be included in an analysis.

6. Automation

Modern process control without automation is unthinkable. Early automation was of a mechanical nature, with the goal to make humans' work more efficient by reducing its physical demands. Later, automation was designed to take over mental tasks as well, even more so with the rise of electronics and computer systems. Nowadays, large portions of an operator's work can be automated. Some reasons for automation are to ease the human operator's tasks that machines are better suited to handle, for example, tasks where monotonous repetition, fast reactions, or very exact controls are required. Such classification has led to the development of concepts for task division between humans and automation, such as the famous Fitts list [19], a tool for task distribution between human beings and computers according to their abilities (e.g., flexibility or accuracy). Common goals are still to increase efficiency, safety, or quality. While physical demands and mundane tasks are reduced, automation often increases cognitive demands, as control tasks are shifting to higher levels [98], operators have to cover a larger part of a process [9], or, ironically, the operators' tasks become more mundane [39].

6.1 Effects of Automation

Besides the desired support, many unexpected, negative side effects of automation have been documented. Basic problems are incorrect implications during the design of automation and either leaving the operators with a suboptimal set of tasks or not estimating the impact automation has on human work. On one hand, automation is introduced for tasks where it supposedly performs better than humans. However, human operators still have to monitor the system to detect failure of the automation. They might even have to take over when a situation occurs that the automation cannot handle. On the other hand, the more passive tasks can cause skill regression. This leads to the irony that the decreasingly proficient human operator has to supervise the supposedly more proficient automation [7]. A possible solution is to design automation so that it is capable of supporting the operator in identifying failure and recovery procedures. In general, automation can be seen as a second process above the original control process. This implies that the operator now has to deal with two control tasks. When controlling a process including automation, appropriate feedback is needed to allow the operators to construct adequate mental models of the system and to stay 'in the loop', meaning to be aware of the system state [67, 27].

With more sophisticated automation, the system is increasingly seen as an additional operator in the process. This implies that deficient automation is 'a failure to design for a coordinated team effort across human and machine agents as one cooperative system' [84, p. 1941]. And consequently, 'ten challenges for making automation a "team player"' have been summarized [58], leading to four basic requirements. These are very close to the four elements of GMOC.

Some conclusions that can be drawn for the impact of automation on the work of the human operators are as follows:

- While reducing the physical strain, the operators are also moved further away from the original process, meaning that means for observation and control change.
- Collaboration with automation has several impacts on mental models and situation awareness. The operators have to be able to understand the automation and to interpret their current state of operation.
- Increased area of responsibility means that operators need to be aware of a larger set of variables and their correlations, requiring additional cognitive resources or appropriate support systems.

6.2 Levels and Dimensions of Automation

When analysing automation, three different dimensions are commonly used. The first two dimensions are *level* and *type*. Popular is the definition by Parasuraman, Sheridan, and Wickens, who define ten levels of automation, increasing with the capability of the system, and four types of automation, adapted from human information processing [72]. The major *technological stages* – manual control, supervisory control, and control of multi-layered networked systems [98] – are the third dimension. In train traffic control, we see the technology getting ready for the shift from the second to the third technological stage (even though there also are enough examples where control is still in the first stage). Design principles as presented in Balfe et al. [9] support this shift.

A general problem at this stage is the question of accountability [39]. This mainly arises from the system designers' view of autonomous automation. In practice, a higher level of automation is often described as a more autonomous system. In terms of GMOC, this kind of autonomy is problematic, as it conflicts in particular with the controllability condition.

6.3 Autonomous versus Non-autonomous Automation

The term *autonomy* has to be used with care, because truly autonomous behaviour is a feature reserved for human beings, who are able inherently to make their own decisions [89]. In contrast, automation can only act heteronomously, according to built-in, programmed rules. Despite this, literature often characterizes automation with a certain degree of autonomy. We try to be careful with our definition. When we refer to autonomous automation, we mean that it restricts users' autonomy. In essence, this results in a reduction in controllability of the human operators – having to collaborate with such an automation, they will not know entirely what the system will do. This can be because of specific requirements or caused by bad system design.

Autonomous systems that are collaborating with human beings need to be especially carefully designed and developed to give the human operator the most possible support. If a task requires, for example, monitoring too many variables or extremely fast reactions to certain events, so that automated support is needed, the system will have to take actions too quickly to give the human operator the possibility to influence the automatic action. However, if the system is well designed, the operator will still be able to predict or understand, and even to influence, the action by parameters set earlier. Conversely, bad design will lead to actions by the system that the user cannot understand or predict, leading eventually to the situation that the automation is conflicting with the user's plans. It is exactly this degree of 'autonomy' that we have to avoid to develop supportive automation.

6.4 Automation from a GMOC Perspective

Automation is a substantial part of many systems for process control. It has a severe impact on human operators' work environments and can therefore be quite visible in an analysis with GMOC. It is important to be aware of the effects that automation has and helpful to consider each separately. Problems in controllability, for instance, can be the result of complicated process properties, insufficient interfaces, or badly designed and implemented automation. This can also be motivated by the technological stage of automation, as in higher stages, the operators' main task moves from controlling the process to controlling the core process and controlling the systems that are controlling the process.

A desired property of automation is transparency. Transparent automation allows users to develop adequate mental models, which in turn helps to make it predictable and to be aware of its current status. As we can conclude from the interplay of GMOC elements, observability and its relation to controllability are key factors for the development of models. Another component of transparency is the communication of goals (cf. [58]). In the best case, automation never acts according to goals conflicting with users' goals. However, if it does, this has to be clear to the users so that they can take the necessary countermeasures. Systems that lack transparency might be interpreted as autonomous, even if this was not the intention. In system design, we should aim for transparent systems that do not appear to be autonomous.

7. Visualization

Visualization is a wide field (see, e.g., [94]). Despite this, there is relatively little work regarding observability for controllability in process control. This chapter gives recommendations for visualization in control of complex processes. Most recommendations can be closely related to GMOC (Chapter 5) and the interplay of its elements. Many of them are derived from our work in train traffic control (Chapter 11).

Visualization is the main channel a system has to generate observability. It is critical to create situation awareness, as it indicates the state of a system and its automation. It is critical to support the development of models, as it is an important source for feedback, showing the results of control actions.

Paper II discusses important implications for visualization in complex work environments. Complexity often arises from a large number of dynamically developing variables with extensive interconnections. A common idea in user interface development is simplicity, which can easily lead to hidden complexity. This can certainly be a reasonable strategy for systems that are (also) directed to inexperienced and untrained users, who would otherwise be overwhelmed by the number of available options, but in the domain we are aiming for, we deal with trained experts. In such domains, hidden complexity will impose cognitive strain on the users as they have to actively search the interface for variables on which they base their decisions or to derive needed information from the actually visible information.

A basic point in our discussion is the confusion between complexity and complication. Problems in organizations and work environments can let complex scenarios appear to be complicated. Such complicated environments make a task indeed hard to cope with, especially when the information needed for decision making is complicated to collect. However, complexity often comes with an intrinsic possibility to cope with it. If visualized correctly, it is, for example, possible to recognize patterns in complex information. As a result, operators in such environments have the possibility to develop skilled intuition based on recognition-based decision making [51]. We argue that a main task in designing systems in complex, dynamic environments is to distinguish between complexity and complication. The latter has to be transformed into complexity and visualized in ways that support learning and development of models, allowing the operators to cope with the complexity.

The following general recommendations for visualization have been given in Paper II:

Show the whole and the details simultaneously
If the user has to focus on a specific detail, it is important still to show an overview of the available information. This will allow relation of detail to the whole and reduce cognitive effort when the user has to switch focus to other information. It also reduces the risk that a user will focus too much on one detail while missing other important developments in the process.

Show all information needed simultaneously; support continuous overview of the whole process
Experienced drivers of a car can overview a great amount of information at once. They suffer more from missing information, for example, when navigating to a certain street, than from an additional source of information, such as a GPS provides. This shows that users can handle a large amount information if it is displayed in the right way, especially when they are experienced. Having a large amount of dynamic information available even supports development of mental models. If the information is secondary and displayed in a less accentuated way, it will not offer a disturbance during tense situations but will remain available if it is needed. Such a scenario allows users to attend to the information during less tense situations and can help to challenge aspects of users' mental models.

Show dynamic information
Dynamic information, for example, trends in and speed of change of a value, is usually most important in process control. If dynamics are not shown, users have to remember values and compare them to current values. If the dynamics of a value are displayed, the user just has to scan the display to get to know the desired information.

Emphasize what is important
If large amounts of information are displayed, as recommended in the preceding points, users need guidance to find the important information in critical situations. In general, static information (as static structures or entities) is often less important than dynamic information. In certain contexts, high values can be more important than low, and vice versa. Contrast, colour, and font size are suitable visual elements for coding importance.

Time-related information
In many domains, it is important not only to know the current state of a variable or if an event has happened but also when this state was reached or, if this information is available, when this state will change again. It is often also interesting to know if other events happened synchronously. To properly visualize this, time-related information should be displayed at identical scale and arranged according to other information.

Show effects of alternative decisions
In process control, users often have different options – different actions that they can perform at different times. The system should support the users in their decision. A possibility is to separate decision from execution. This would allow the users to 'test' an action first, and see possible results, before implementing the action. This requires a certain ability of a system to simulate the process.

Support development of mental models
Mental models are, as discussed before, often incomplete and not entirely perfect. By displaying more than the purely necessary information, we can support users in developing their mental models. A very efficient method to support the users is to give them freedom to perform smaller control actions and allow them to observe the dynamic response of the process to such actions.

Efficient coding of information
Similar to emphasis of the important, we have to efficiently code large amounts of information. This reduces the cognitive effort required of users in identifying important changes, events, or situations. It is hard to give concrete advice, as efficient coding depends strongly on context. Therefore the visualization has to be evaluated in the specific context of use.

Visualization of automatic systems
As discussed in the previous chapter, automation adds another layer of complexity to users' work environment. Key points to reduce this complexity are to design automation in a way that clearly communicates its state of operation, to allow for control of this state, and to design the automation in a strictly non-autonomous way.

8. Methods Supporting System Design

So far, concepts and aspects relevant for the design of control systems have been presented. This chapter introduces analysis and design methods that have been applied to conduct the studies in the course of this thesis. We follow a user-centred design process. This process and the role that GMOC played in it are described first. This is followed by some specific methods that can be used during the different phases of this process.

8.1 User-Centred System Design

The main goal of user-centred system design (UCSD) is to support the integration of the users' needs into software development projects. According to Norman, two different conceptualizations of a system exist: one is the view of the designer or developer (design model), which is the basis for the system to develop, and the other is the view of the user (user's model), which develops during the user's interaction with the system. Norman describes an important principle of UCSD: "The Design Model is the conceptual model of the system to be built. Ideally, this conceptualization is based on the user's task, requirements, and capabilities" [66, p. 47]. We see GMOC as a link between the users' and designers' models, as shown in Figure 8.1, which depicts users' work in terms that should be used by system designers and developers to develop a system that gives needed support for the users' work.

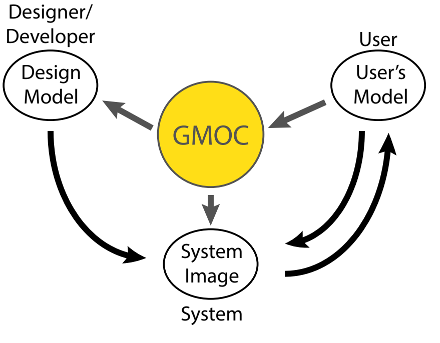


Figure 8.1. First role of GMOC in system design: helping the designer develop a design model based on users' requirements. Adapted from Norman [66].

A problem with UCSD is that the term is widely used but vaguely defined. It also seems to come with obstacles so that even committed projects risk failure in meeting users' requirements [37]. To avoid this, UCSD should 'be defined in terms of a process where usability work and user involvement are tightly integrated with the development process' [37, p. 406]. Furthermore, a UCSD cycle is shown, and twelve key principles are formulated to support such a definition; see Figure 8.2, which shows our adaptation of the UCSD cycle with the following four phases:

Analysis User requirements and needs are analysed and translated into a project plan and a vision, formulating the goal of the project.

Design The new work environment is designed, ideally involving users, developers, designers, and representatives from all stakeholder groups in the development process. The design phase usually leads to (requirements for) a prototype.

Evaluation In this phase, a prototype is implemented, deployed, and evaluated with end users in the original work context.

Feedback Originally, feedback is a phase at the end of a system's life cycle [37]. We interpret the cycle slightly differently so that, in projects with close iterations, this phase can fade directly over into the analysis.

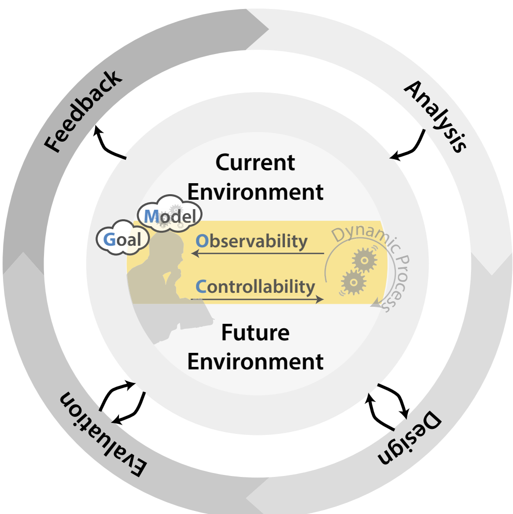


Figure 8.2. Our interpretation of the user-centred system design process presented by Gulliksen et al. [37]. The centre of the figure symbolizes the descriptions of the current and a future work environment following the GMOC model.

In our interpretation, the cycle is seen as primarily guiding the iterative development of prototypes rather than as a complete system life cycle. In such a view, the feedback phase is optional and will often be skipped. Instead, we will have several iterations going through analysis, design, and evaluation before a system is ready for operative use. Even these three phases are neither strictly defined nor distinct from each other. A project might begin with a long analysis phase, while these phases become much shorter in future iterations. The first prototypes will be very low fidelity and can be iterated directly during one design phase, without going through an evaluation in context. The latter instead becomes important during the later development of a project, when a number of prototypes have already been designed and the development of requirements for a final system has begun.

During our system development process, we couple each phase to the same description of the human operators' work. The description is made following GMOC, that is, by describing the operators' goals and models and the observability and controllability the system provides. We describe two work situations, the current and the improved future situation. This description informs every phase of the process, and further information gathered during the process will be fed back into the description. Every iteration of the development process does also mean another iteration of the description, which should minimize the risk of missing any important user requirement during the complete development process.

During the analysis, a description of the current work environment is produced. During design, this description might be completed, but the main goal is to find problems and produce the description of a future, more optimal environment (i.e., optimal as it at least avoids the identified problems, but the process will possibly also reveal further potential for improvement). This description will be used to implement a prototype during the evaluation phase. The resulting prototype will be deployed and tested with the users and evaluated towards their goals and the GMOC description of the improved work environment. During the feedback phase, at the end of a system's life cycle or after a substantial iteration of a prototype system, our description of the work environment switches; the 'current' description will be updated to the current version (at the end of one iteration in the UCSD process), and the future description will be updated based on information on further problems and improvements, obtained from the evaluation.

8.2 Analysis and Design Methods

This section describes methods that can be used within different phases of the design process. It is possible and even necessary to apply different methods, depending on the circumstances of a project and the possibilities the designer has in these projects. The methods presented here were most important during our projects.

8.2.1 Observations

Observations are an analysis method aiming for a description of human behaviour. They are commonly used in ethnography to understand human behaviour in its natural environment [10]. Different options are available when applying observations in HCI. As researchers, we have to plan observations with respect to our research objectives, the application domain, and opportunities or limitations arising from the domain.

Our goal is to analyse the work of operators in complex, dynamic environments. As we are interested in a person's work, person-focused observations are most adequate. This means that we follow and observe an operator during a working shift, if constraints, such as safety regulations, will allow. Being able to observe operators at their original workplaces allows us more natural and holistic insight into a person's work. With such, we are able to observe additional external factors, such as social behaviour and communication with colleagues or technical and environmental aspects deriving from processes and systems. Observations in controlled environments such as laboratory studies can be useful in researching certain aspects and to testing hypotheses and new systems, but then we are leaving the ethnographic perspective.

A main reason for observations is the more objective view of the researcher as a neutral observer than the subjective view of the operators themselves. If they report about their work, for example, in interviews, they might leave out facts that they regard as unimportant or 'bad work practise', or they may simply not be able to report every detail. They might not be aware of certain behaviour, actions might be so internalized (tacit knowledge) that they are hard to recall, or other cognitive constraints hinder them from delivering a 'perfect' report. Observations are thus a good method even to identify internalized behaviour.

The role of the observer in complex control situations is more likely to be passive and unobtrusive than participatory. Active observation might violate safety regulations and can influence the work situation in a way that the observed details might no longer represent the typical situation. Nonetheless, if research objectives and external constraints allow, observations can be combined with think-aloud, where operators verbalize their thoughts and actions, or interview techniques, where operators are asked about certain observations. We use a combination of these techniques. For example, when we observed a behaviour we could not understand, but that seemed to be relevant (e.g., because we observed it repeatedly or in situations that seemed critical), we tried to ask about this behaviour in an adequate situation, that is, close in time to the observation, so that the operator was likely to recall the situation, but after a normal, uncritical state had returned. The motivation was to obtain an explanation for an observation in the natural setting and close to the original event, giving a larger probability that the report would be 'correct'. Similar results can be obtained with verbalization methods (see, e.g., [30]).

Of interest is the additional possibility of, rather than observing the operators directly, but screen captures of their interaction with the control system. This gives access to a large number of data without the usual problems that come with direct observations at the workplace, such as safety regulations, working hours, and limited space. However, this radically reduces the observable amount of information in the interaction with the system as a whole. Other factors, such as communication, conditions at the workplace, or the operators' facial expressions, are not observable. This can make it hard to understand the operators' behaviour, and thus screen captures are only recommended when the observer is already familiar with the operators' work. During our work, screen captures proved to be a useful method for analysing and evaluating the work situation when a new system had been introduced.

In relation to GMOC, observation clearly focuses on controllability, as actions are the easiest to observe. Even observability can be revealed, especially if additional tools, such as eye-tracking devices, are used. Goals and mental models are, in contrast, difficult to observe. We can only guess about the relation between observability and actions, between goals and the direction of attention, or between facial expressions and models (i.e., a sceptical expression might indicate that an action (controllability) did not lead to the expected reaction (observability)). However, these guesses can be used to formulate hypotheses about the relation of collected observations to goals and mental models. These hypotheses can be validated or falsified later (e.g., by other observations or in interviews).

8.2.2 Interviews

Using interviews, we can get a deeper understanding of the operators' work, the work domain, users' attitudes, and their opinions about certain aspects. Interviews can be useful in analysing the work environment and in the qualitative evaluation of systems. Interviews are a good frame for observations: preceding observations, interviews can give us a general understanding of the work domain, which makes it easier to interpret observations and also helps to focus on certain behaviour during observations (note that this will influence the observer's objectivity). Subsequent to observation, interviews can provide deeper understanding of the collected observations (cf. [10]).

We use mainly unstructured or semi-structured interviews. Depending on the progress of our studies, we start with open questions. We want to achieve a narrative flow, where the operator reveals topics that did not come to mind as we prepared the interview questions. If we are new to the domain, such questions also help us understand the domain; if we are facing an operator for the first time, such questions help to create a good atmosphere for the rest of the interview. After several open questions, we can ask more focused questions. These can consist of detailed questions about topics that came up during the open phase, questions on certain topics that we prepared beforehand, or follow-up questions on observations. We usually prepare a number of topics or questions we would like to talk about with the operator. There is a good chance that such topics will be covered during the open-ended questions, which offers a slight advantage, as the answer will be less biased than if we had elicited it with concrete questions.

Documentation should consist of notes taken during an interview, for example, in the form of new questions or key phrases, and an audio recording of the complete interview. It is advantageous for the interviewer to write down his or her own reflections on an interview immediately after its conclusion to document impressions and key information the interview provided. Recordings are necessary to confirm and complement the interviewer's impressions and for additional analysis of parts of the interview where certain notes have been taken. Furthermore, transcribing the interview can lead to new insights.

Instead of a complete transcription of the interview with formal analysis of the text, we prefer a 'light-weight' method for our purposes. With this, we follow the main topic of GMOC – 'trading the rigour for the relevant': we listen through our recordings again and transcribe interesting quotations, statements, and topics (with an interpretation or conclusions) from the interview and note time stamps for later validation. The next step is to go through these transcriptions and assign the parts to the GMOC elements describing the work environment (e.g., if the operator has described a goal). Our motivation behind this is to shift the focus from the more mechanical work of transcription towards analysis of the content. Disadvantages of this form of transcription are that it is possible to miss topics that did not seem relevant during the transcription process and that wrong conclusions during this process are hard to identify later (as we do not always work with the exact wording, but with our interpretation).

8.2.3 Vision Seminars

Vision seminars are a design process where system operators are directly involved in developing solutions for new work environments [40]. First of all, the name 'vision seminars' is a bit misleading. The process does not consist of typical seminars where a seminar leader presents certain topics. Instead, the seminars are a kind of workshop, where the participants decide on the outcome, contributing their own knowledge, experience, ideas, and thoughts. A group of researchers or system designers guides and documents their contribution towards a design of new systems and organizations forming the participants' future work environment.

The process covers the design phase of the UCSD cycle presented in Figure 8.2. The researchers and designers participating in the process are supposed to familiarize themselves with the operators' work domain and environment before and at the beginning of the seminar process (i.e., during the analysis phase), and the process will result in a description of a future work situation and additional design documents such as prototypes [69]. As the process ends before system design and technical development, that is, before system implementation, it is necessary to involve system developers in the process, who can carry the knowledge into development, or to include operators and, especially, the designers involved in the vision seminar process in system development.

Vision seminars are inspired by participatory design regarding operators as central actors during the design, and contextual design/inquiry emphasising the importance of the designers' knowledge of the work domain and enhancing the operators' work environment. Besides this, the vision seminar process does focus on specialized systems, implying that the projects are often in-house developments for a well-defined group of experienced operators [69].

The workshops (see Figure 8.3) are the central element in the process. Here a group of participants, typically operators and some other actors in the organization, meets with the process leaders, typically researchers or system designers. Workshops are held with several weeks in between. Workshops can have certain themes (e.g., organization, communication, technology), and the time in between gives the participants opportunity to reflect on the workshop. Discussing results with other colleagues and asking for further opinions and viewpoints are endorsed as well. The participants can be given assignments to support preparation of the next theme and to keep them continuously interested in the process. The process leaders will use the time to document and analyse the process.

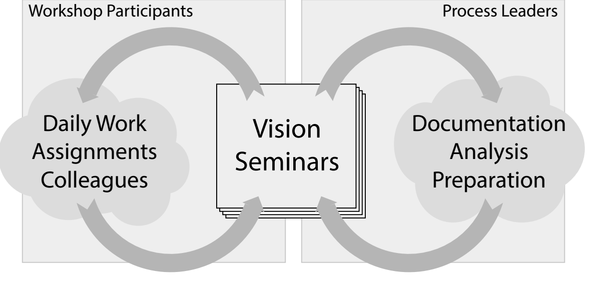


Figure 8.3. Centrality of seminars to the process. Between seminars, there is time for reflection. This is facilitated by assignments given to the participants and by documentation together with analysis performed by the process leaders [40].

The goal of the vision seminar process is to document the current work environment, including 'present work, tasks, cooperation, and workers' skills' [69, p. 9], and to envision future work. Searching for 'barriers' to efficiency, productivity, and a good work environment is a typical approach to identifying improvement for the future work situation.

In relation to GMOC, the vision seminar process works very well. Commitment to context awareness and user participation are common goals. The description of a work situation in terms of GMOC is a good way to document the seminar results in a form that is understandable to the participants. The comparison of the current situation and envisioning a more optimal, future work situation are also key ideas behind GMOC.

A disadvantage of the process is the amount of time necessary to develop its full potential. It takes time to activate the participants to envision truly creative solutions. Often the designers have to give thought-provoking prompts. It takes time to get the participants to think out of the box and take different perspectives than just one reflecting their personal daily work. Additionally, because the process does often identify the need for organizational changes, interest and commitment at all layers of the organization are needed if the project is to lead to a successful development endeavour.

9. A Generic Method to Apply GMOC in User-Centred System Design

This chapter introduces a method that allows the application of GMOC in user-centred design. This method evolved from our own studies and has been used in similar ways during the studies presented in Part III. To successfully use GMOC, it is important to understand its basic elements and their interplay. As every application and domain has different requirements that have to be addressed with different methods, a thorough understanding of GMOC allows its adaptation to a different set of methods. This allows application in a wide range of projects with different contexts. Therefore the method described in this chapter should only be seen as an example to illustrate how GMOC can be used – a starting point for further studies using GMOC.

9.1 Overview

The basic idea follows the UCSD cycle as described earlier, with the goal of creating a description of the current work situation and a description of the improved future work situation. We want to be able to describe concrete actions to take when implementing a new or improved system to support the work situation. The description will be done in terms of GMOC; this means that we constitute for each element goals, models, observability, and controllability, how they exist in the current system, how they are understood by users, and which problems are experienced. Respectively, this means for the future system how the elements should be supported (to improve the work environment) and how better support of the elements can be implemented.

The UCSD cycle has four phases. In general, the methods used in certain phases might differ less than the objects to which they are applied. The basic objectives in analysis, evaluation (i.e., to evaluate the developed prototype, not to construct the prototype), and feedback are quite similar: we always want to analyse a work environment. It is the work environment and the objectives that differ.

The methods used in our studies can be divided into analysis and design methods. The process starts with an analysis of the existing work situation with the goal of defining user requirements and planning the further process. Afterwards, we design solutions in the form of prototypes of increasing fidelity. For evaluation of our solutions, we generally analyse the situation based on an iterated prototype or system in its context of use. We aim for an evaluation of the changes and improvements in comparison to the original work situation. The feedback phase (interpreted according to [37]) would then be an analysis of the work environment at the end of a product's life cycle. However, this phase was not necessary in our studies.

9.2 Analysis

Analysis aims for a description of the current situation. We usually use a combination of observations and interviews as described in the previous chapter. Verbalization and questionnaires can be useful as well. Observations are a good method to use in identifying observability and controllability based on the users' actions. We can see which controls are used and which information is attended to. It might also be possible to identify connections between observability and controllability, allowing conclusions towards goals and models.

In interviews, we can then ask about the users' subjective views on observability and controllability and try to verify our observations as well as our hypotheses about goals and models. We can get a better impression of the work situation, the interplay of the elements, and possible shortcomings. We usually perform semi-structured interviews and start with open questions such as "How would you explain your work to a novice?" The goal is to motivate the users to start a narration of their work, giving us a comprehensive and sometimes lively view of their work. We can then ask more specifically about common or problematic situations. Such narrations usually allow numerous conclusions according to the GMOC elements. Later during an interview, we can ask concrete questions in relation to the elements, such as "What are the goals of your work?", "How do the systems support you in achieving … ?", "Can you explain, how … works?", or "What happens if you do … ?" Here we also can ask concrete questions based on our observations. Interviews were recorded and partially transcribed for a consecutive analysis of the information we received about the GMOC elements and related problems.

Using verbalization techniques, we ask the users to comment on their or their colleagues' actions, usually using video recordings. Verbalization by different users but based on the same footage has proven to be useful in identifying (differences in) models [46]. Questionnaires have as an advantage their ability to obtain more data with statistical relevance. We can reach a much wider group with questionnaires than we are able using observation, interview, or verbalization methods. Repeated questionnaires during different phases of the project can be a good tool for evaluating if certain aspects have improved or declined.

9.3 Design

Design aims for system requirements. The goal is to develop a system that allows changes in the work environment towards the improved, future situation defined in terms of GMOC. This is initially given by the analysis results and refined during vision seminars (see the previous chapter). In the beginning, vision seminars are useful to test and improve the description of the current version. Together with the users, the researchers can evaluate their hypothesis about problems identified during the analysis. Reasons for and the importance of problems, and their impacts and possible solutions, can be discussed. This will result, on one hand, in a description of a future work environment, following the GMOC model, which can be seen as a goal for the system to develop. On the other hand, concrete solutions will be developed. It should be possible to explain these solutions using GMOC.

We then implement these solutions in prototypes. In the beginning, these can be sketches on a whiteboard during a vision seminar to brainstorm different possibilities together with the participants. Later these can be more advanced prototypes produced between the vision seminars. At a certain stage, a prototype will be ready to be implemented, either as a high-fidelity prototype, for example, a simulation, or in an operative system. At this stage, we will go over to the phase of evaluation. Thus the design process mainly consists of a combination of vision seminars, design sessions by the researchers between the seminars, and prototype evaluations. Progress will be documented in our GMOC description of the existing and future work environments, as well as in a list of problems, hypotheses, solutions, and design prototypes.

9.4 Example

Here, to illustrate the process, we give a hypothetical, simplified example of progress during the development of an information system for passenger train drivers. GMOC elements in the following description are marked with their initials: goal (G), mental model (M), observability (O), and controllability (C).

During observation, we notice that a driver brakes relatively strongly as the train approaches a stop signal just before a station. The stop signal is part of observability (O), the braking part of controllability (C).

During an interview, we ask for typical 'disturbing' moments. The train driver mentions, "You know, sometimes when you are late, you drive as fast as possible to make up some minutes (G), just to realize that the train to Uppsala (M) is late as well (O/M). Then you have to stop right before the station. You will be late again (C/M) . . . and your passengers will become nervous (M)."

From this description, we can derive a typical problem. Train drivers have as a goal to be on time (G). When they realize that they are late (the possibilities are not explained here) (O), they try to control the situation by increasing their speed (C). However, if the preceding train is late as well, the train has stop just in front of the station, as the track has not been cleared. During an interview or a vision seminar, we can further identify negative effects: increasing speed and then having to brake wastes energy, increases wear on the material, and reduces passenger comfort (G/M). Stopping and starting again might in the end cost more time than driving slightly more slowly without having to stop (M). Stopping just before a station makes passengers more nervous than a continuous movement between stations (M). So we can conclude that this problem is worthy of study.

As a main reason for this stop is that the driver is not aware ('lack of situation awareness') of the preceding train's delay (O). We can hypothesize the following: "If this delay were observable, the driver's mental model would suggest that he not increase his speed too much." We can continue the process by trying to falsify our hypothesis, refining it together with the users or by creating a prototype of a system that would provide the additional observability of the delay. Our example concludes here, but the designs that will be presented in Chapter 14 actually show different attempts to solve this problem.

Summary

The first objective of this thesis is to provide a model that supports the development of systems for human operators in complex, dynamic environments. This part defined such a model and proposed a method for system development based on this model. In the next part, this method is used to analyse train traffic control and train driving and to develop new support systems. Assessing the results and the used method serves as a basis for evaluating the defined model and concluding if the research objective has been met.

Key points of this part are as follows:

- Situation awareness can be described as the operators' knowledge of 'what is going on'; this is important to make correct control decisions.
- Mental models are a concept that describes the operators' understanding of the system, the environment, their work, and the world around them.
- Situation awareness can be seen as a configuration of mental models; in other words, mental models are a functional description of the world and situation awareness the result of filling the function with concrete information.
- Control theory has introduced the notions of goals, models, observability, and controllability; researchers in dynamic decision making have applied these notions to humans in complex, dynamic environments. Our research continues this work by making this model, GMOC, applicable in system development.
- Goals, models, observability, and controllability are the elementary prerequisites for human control in complex, dynamic environments. All elements have to be geared together to give the operators control above a process and to allow the development of adequate mental models.
- Automation can have a substantial impact on human operators; a process description in terms of GMOC can reveal much of this impact.
- (Seemingly) Autonomous automation should be avoided.
- Adequate information visualization is important for an operator's performance; GMOC helps us to describe important factors for visualization.
- GMOC can be applied during user-centred system design. Analysis methods such as observations and interviews can be used to analyse and evaluate work environments towards GMOC.
- Using GMOC, we can hypothesize how changes in systems can improve work environments. Prototypes can be created and used during vision seminars to evaluate hypotheses.

Part III:
Developing Systems for Future Train Traffic Control

This part presents our work towards the second research objective: development of concepts and systems for future train traffic control. It presents some background on train traffic (control) and four studies in this field. The studies have been based on methods embracing the GMOC model from the previous part of this thesis. The results are used in the last part of this thesis to evaluate the usefulness of the model in analysis and development of systems in complex, dynamic environments.

10. Introduction to Train Traffic Control

Train traffic, road traffic, aviation, and maritime traffic are complex, dynamic processes. All of these have complex underlying structures, organizations, and roles with different tasks, rights, and obligations. Apart from this, each type of traffic is inherently different. This is already obvious from their degrees of freedom of movement: whereas a plane can (within certain boundaries) freely move in three dimensions, boats and cars can move in two, and trains are essentially bound to one dimension.

In human factors research, aviation and road traffic are established fields, whereas rail human factors is a relatively young area of inquiry. Today the study of rail human factors is firmly established in the United Kingdom, and its importance has grown [96]. It offers many interesting challenges from a couple of different perspectives. Off-line processes, such as timetabling and resource planning, can benefit from improved support systems and routines. Operational control includes train traffic control, train driving, planning and conduction of maintenance work, customer information, and operational planning at the railway undertakings, for example, rolling stock management and resource allocation. Among the many research opportunities are information, communication and support systems, organization of the process, safety and resilience, and decision support systems.

Our research covers the organization of train traffic control, aiming for more efficient collaboration, information, and routines. The focus of this thesis lies on train traffic control and train driving. Traffic controllers and train drivers have to collaborate to gain control over a process in which most decisions influence other participants. They sometimes have to work under extreme conditions, for example, they have to deal with very high or very low cognitive workload, fatigue during night shifts, or periods with low cognitive strain, while the process is very sensitive to failure of material and weather conditions. Despite these challenges, operators have to make safety-critical decisions that guarantee smooth traffic flow with infrequent delays and optimized energy consumption.

This chapter gives an introduction to train traffic control, including the organization of train traffic control, tasks and the workplace of train traffic controllers, and tasks of and support systems for train drivers. Finally, a short overview about general goals or optimization criteria in train traffic, related work, and our studies is given. The studies themselves are presented in the remainder of this part.

10.1 Train Traffic Control

Wherever one studies train traffic, one can see that railways have grown historically. Organizations have developed quite differently in different countries [34], and even within country lines, safety systems, stations, signal boxes, and so on can differ tremendously. At some places, one can still find old mechanical lever frames. Even in large European capitals, one can find signalling centres, where traffic is operated by equipment that is forty years and older. At the other extreme, one can find the recent Chinese high-speed lines, with thousands of kilometres of dedicated tracks, or fully automated systems, such as the Lötschberg Tunnel in Switzerland [64].

During the On-Time project, we analysed the differences in train traffic control throughout Europe and described its organization in several European countries. Here we focus on the Swedish organization. Distinct features of the Swedish organization are its centralization and the role of the Swedish train traffic controller. At the end of 2013, the Swedish transport administration, Trafikverket, reorganized train traffic control by introducing additional regional and national control layers. Its goal is a better coordination of traffic control across the borders of different control areas and better communication with the railway undertakings. The effects of this reorganization have not yet been sufficiently explored. The results presented in this thesis are mainly based on the Swedish train traffic control organization until 2013.

10.1.1 Organization of Train Traffic Control in Sweden

Figure 10.1 shows the complete train traffic planning process, beginning with timetabling more than one year before actual operation. Timetabling has only been covered marginally during the work presented in this thesis. Details on the Swedish timetabling process can be found in other reports [43]. Our focus lies on traffic operation and the traffic planning process up to a few hours before operation. This process includes an infrastructure manager (IM) and railway undertakings (RU). Whereas there is only one IM, the Swedish traffic administration Trafikverket, more than thirty private RUs organize all kinds of traffic, from local commuter trains to long-distance freight transportation.

The main role in this part of the process is the traffic controller at the IM. In Sweden, control takes place at eight centralized traffic control centres. The traffic controller fills the role of both dispatcher, rescheduling the current traffic plan with respect to perturbations and disruptions, and signaller, executing the plan and controlling train paths, points, and signals. In most other countries, the roles of dispatchers and signallers are separate and performed at different workplaces. Paper III discusses differences between the Swedish organization and train traffic control with collaborating dispatchers and signallers.

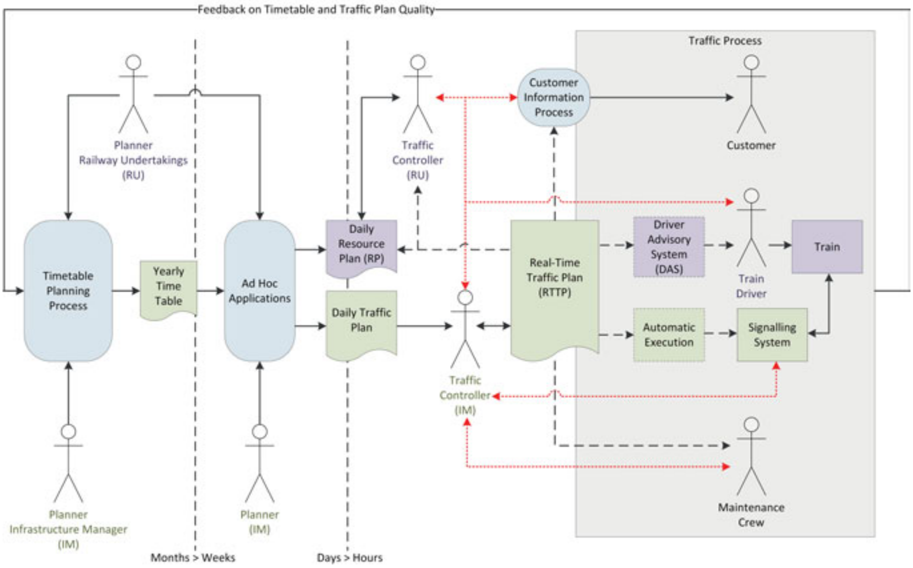


Figure 10.1. A simplified overview of the train traffic planning and operational control process in Sweden. The planning horizon decreases from left to right, starting with the yearly timetable and ending with actual operation, including real-time replanning and train driving.

The RUs also have traffic controllers. Their task is quite different: they organize the resources of the RU, mainly rolling stock and train drivers, to fulfil customers' requests. A basis of their work is often a resource plan based on the original timetable (note that, because there are so many different RUs, details can change from RU to RU). Customer information is a shared responsibility between IM and RU: the IM usually provides more general information (e.g. via information systems and announcements at stations), whereas the RU is responsible for direct communication with the customer.

In this thesis, 'traffic controller' refers to the role played at the IM, with the main responsibility for controlling the traffic in a safe way and as close as possible to the daily traffic plan. This plan is an extract of the timetable, including short-term planned exceptions, such as additional trains or maintenance work. It is not uncommon for the daily traffic plan to contain conflicts and for freight train traffic to deviate substantially from the timetable; therefore a main task of the traffic controller is to constantly replan traffic. This implies that the original daily traffic plan is usually not the final, valid plan. This is instead a plan only existing in traffic controllers' minds and can only be communicated via telephone. This issue is covered in Papers III and IV. Our solution, a real-time traffic plan (RTTP), is explained in Section 12.3.1.

The last role to be discussed here is that of the train driver. A train driver's task is to operate the train safely while holding deadlines or timetabled stops.

They usually have a schedule based on the daily timetable that they get printed on paper at the start of a shift. When operating a train, a train driver's trip is restricted by signals based on the traffic controllers' action, and supervised by an automatic train protection (ATP) system, which ensures that trains stop at red signals and do not violate speed limits.

10.1.2 Traffic Types

The Swedish railway network is widespread, with significant differences between different parts. There are single-track lines with light utilization through the mountains in western Sweden; highly utilized double-track lines in the southern region with a high percentage of passenger traffic; large and complex stations, such as Stockholm Central Station; and the iron ore line in northern Sweden, a single-track line operating at its capacity limits and dominated by long and heavy iron ore–transporting trains with a speed limit of 70 km/h (60 km/h for loaded trains). The traffic on the Swedish rail network is mixed, with no tracks reserved for high-speed trains; the highest permitted speed is 200 km/h. A few tracks are reserved for local commuter trains.

10.1.3 The Workspace of the Train Traffic Controller

Figure 10.2 shows the typical workspace of a train traffic controller in Sweden. The main tools are the traffic control system, a printed graph of the current day's traffic plan, and the telephone. Additional systems and paper forms are used to keep track of delays, to manage phone calls, and to report anomalies.

The traffic control system consists of a track diagram that displays the infrastructure status: free, occupied by a train, or set for a specific train to enter. Depending on the system, the traffic controllers can set train routes either via text commands or directly via mouse on the track diagram. Automation supports traffic control in different forms. This can include locally programmed switch boxes with simple algorithms such as 'first come, first serve' or more complex, centrally programmed systems, similar to automatic route setting (ARS). These systems often create problems as described in Chapter 6; that is, the automation is perceived as autonomous, hard to understand, and counterproductive during traffic disturbances [54, 9].

A main problem of the track diagram is the quality of information. It is, of course, exact in the sense that it displays exactly which segments are blocked by a train. However, because these blocks can have lengths of several kilometres, the display does not provide information about the exact position of a train within a block. Only when a train moves between two blocks does the traffic controller get sufficient information about a train's position and speed. To still be able to control the traffic, controllers must remember the time of transition of a train between two blocks and develop mental models that allow

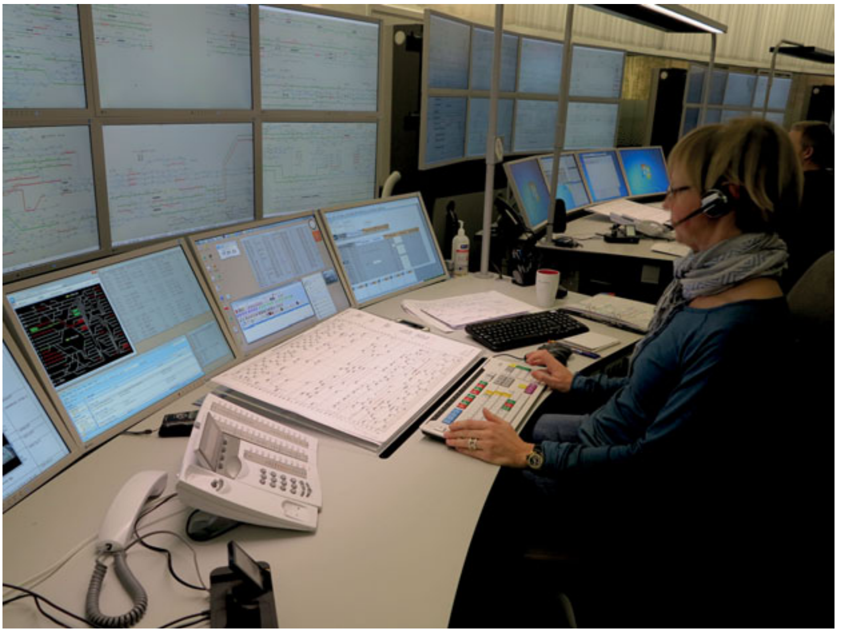


Figure 10.2. The typical workplace of a train traffic controller. The array of screens, slightly in the background, shows the track diagram, displaying the blocks that are blocked by or reserved for certain trains. The screens slightly in the front give access to the different control and information systems. Right in front of the traffic controller is the telephone, connected to a blue-tooth headset, and the traffic graph on paper. This picture was taken in 2013 at the control centre in Stockholm after a complete redesign (but still using the same, traditional information systems).

them to derive a train's speed and position from the sparse information they have. The situation gets even more complicated by the fact that the lengths of track blocks on the display do not relate to their actual lengths. This absence of proper observability is a good example underlining the necessity to develop mental models to support situation awareness.

The printed time–distance graph reflects the daily traffic plan, on which changes can be drawn using a pen. Disadvantages of this are obvious: as replanning and accurate noting of a train's trajectory would require numerous redrawings, this plan can only provide rough support. Additionally, a paper-based plan is impossible to share automatically; all changes have to be communicated via telephone. Busy shifts can require the traffic controllers to answer the phone continuously, for example, to register trains entering the authority area, to approve shunting or maintenance work, and to record infrastructure failure. There is hardly time to communicate noncritical information, that is, most changes in the plan.

To summarize, the main problems at the traffic controllers' workspace are as follows:

- Insufficient observability of the train traffic process can lead to suboptimal planning, increased workload, and the need to develop complicated mental models.
- Communication creates a very high workload during busy shifts, prohibiting the communication of supportive information that, for example, could improve the train drivers' understanding of the traffic situation.
- Today's tools do not adequately support precise planning.
- Information is distributed over several different systems, inducing further workload for attending to the necessary information.
- Automation is perceived as autonomous and does not give needed support, especially during disturbances.

10.1.4 Local Knowledge

During our studies, train traffic controllers frequently used the Swedish term *lokalkännedom*, which best translates to 'local knowledge'. It seems to be a quite important term, but despite this, to my knowledge, it has so far only been studied by researchers in the United Kingdom [76]. They conclude that local knowledge is 'beneficial to facilitate decision making and/or situation awareness' [76, pp. 365–66].

Based on this, conclusions from our own studies (which were not focused, however, on local knowledge), and an internal document from Trafikverket describing their view, I suggest dividing local knowledge into two types.

The first type is closely related to situation awareness. This includes, for example, knowledge of landmarks that influence operators' speed to locate a train's position (we pulled this example from the U.K. report and from our own interviews; this facilitates situation awareness) and knowledge of states of assets, such as ongoing reconstruction at a station (example taken from Trafikverket's internal report; direct situation awareness).

The second type is closely related to (mental) models. This means, for example, information that can be used as input to mental models, such as the gradient of a track, which is important for calculating a train's speed and ability to accelerate or brake (this example appears in all sources).

I admit that this distinction is debatable. I see as a main difference that the information provided by the first type reflects current conditions, whereas the second type supports predictions of future developments. However, both types of local knowledge facilitate operators' ability to plan. If a clear distinction can be made, it would be interesting to investigate implications for system design, that is, if there are different ways to support these types of local knowledge.

So far, we only know basics about local knowledge; much more exploration is needed. It would be interesting to know the impact of local knowledge on traffic controllers' performance and the factors that facilitate development of local knowledge. From our studies described in Chapters 11 and 13, we can conclude that some types of local knowledge, for example, the gradient or height profile of the track, can be integrated into support systems. However, in the first instance, the height profile helps the traffic controllers to identify places and thus to recall their local knowledge, which may inform them, for example, if it would be appropriate to stop a heavy train at that place. It will hardly be possible to replace the entire local knowledge of experienced operators with information in support systems.

10.2 Train Driving and Driver Advisory Systems

This section introduces the role of the train driver. A focus lies on driver advisory systems meant to support train drivers in optimizing their train control.

10.2.1 Train Driving

In the train traffic process, the main task of train drivers is to operate their trains following signals and rules and the current plan set by the traffic controllers. Additionally, they have to follow schedules given by the RU. These can be passenger timetables, schedules and deadlines for freight delivery, and additional plans for rolling stock management (maintenance, cleaning, shunting, etc.).

Depending on the type of train and the kind of traffic, trains' controls, properties, and equipment can vary greatly. Heavy freight trains can have slow acceleration, low top speed, and long braking distances, while regional passenger trains with several power cars are light and can accelerate and decelerate relatively quickly. Often the drivers of passenger trains inform their passengers about upcoming stations or delays. Train drivers will also commonly function as 'machinists', monitoring the health of their trains and performing simple check-ups and troubleshooting, such as engine or brake malfunctions. They have to monitor and report infrastructure malfunction, too.

Train drivers' actions have large impact on the traffic process. If they cannot or do not control their trains following the traffic plan, their deviations might influence other trains. If they drive in a considerate way, they prevent early wear of infrastructure and reduce maintenance time. And these are just a few examples. Difficulties in train drivers' work are long waiting times for freight trains with lower priority, fatigue due to shift work, and very uneven distributions of workload. Their social life and health can be affected by odd and uncertain working hours and the lone characteristic of their work.

Figure 10.3 shows the cab of an iron ore train. There are different sources of information. It has indicators, meters, and more advanced systems to monitor a train's health. Most important is the safety system, the ATP (Automatic

Train Protection). It can initiate emergency braking if a train is in danger of passing red signals or violating speed limits above a certain safety margin. The ATP is certainly the most important and valued system for the train drivers, but nevertheless, it does not guarantee 100 per cent safety. It is not uncommon to overwrite warnings coming from the ATP, for example, when two trains operate on a track or platform simultaneously, during shunting, or in cases of certain infrastructure failure. In very, very rare events, the system even fails, for instance, in consequence of a mistake in wiring during track maintenance. Thus train drivers are still required to observe the track carefully to avoid accidents. Unfortunately, train drivers often only have limited possibilities to react, as braking distances often exceed the distance to objects, people, or animals on the track.

Figure 10.3. The workplace of a train driver of an iron ore train.

Another source of information are *driver advisory systems* (DAS). These are becoming more and more popular and have as a main goal to support and optimize the drivers' driving styles (e.g., [4, 2, 61]; see also [71] for a review). We see these as a main component in supporting collaboration between train drivers and traffic controllers. Therefore DAS are a major topic in our research.

10.2.2 Centrally Guided Train Operation

DAS can be divided into two classes, *connected* (C-DAS) or *unconnected*. The second class requires an on-board device operating on more or less static information. It can have access to the current position of a train, its speed, a map, the timetable, and the traffic plan. It can even receive messages from the RU's traffic control, but it cannot receive real-time updates from the IM's traffic control. This is in contrast to C-DAS, which in addition requires supplementary systems at the IM's traffic control centre and a reliable communication channel. This makes it more complicated and expensive to develop. As a consequence, most DAS existing today are unconnected [71].

C-DAS have the advantage that they can provide the train driver with much more detailed and interesting information about the current traffic situation. Built on C-DAS is the concept of *Centrally Guided Train Operation* (CGTO). The term CGTO was coined during the On-Time project and is inspired by ATO (Automated Train Operation), which means train operation without an active train driver controlling the train. However, CGTO requires train drivers, but it supplies them with detailed information on optimal driving strategies. This information is based on a centrally optimized traffic plan. The DAS CATO (Computer Aided Train Operation) [49, 99] and the DAS concepts described in this thesis (Chapters 12 and 14) implement CGTO.[1]

10.2.3 Route Knowledge

Similarly to the traffic controllers' local knowledge, as discussed in Section 10.1.4, train drivers develop route knowledge:

> The knowledge and experience of a route that drivers develop over time (what is known as 'Route knowledge') supports anticipation and future-orientated behaviour. Route knowledge allows the driver to think ahead, and helps control the allocation of cognitive and perceptual resources based on expectations about the future. [63, p. 673]

This skill allows train drivers to meet their goals [47], for example, to achieve a smooth and energy-efficient driving style. CATO supports route knowledge by showing the track gradient. Novice train drivers using CATO appreciate this, as it supports their understanding of forces affecting the train and in turn facilitates their development of mental models.

10.2.4 Driver Advisory Systems and Their Motivations

A review of existing DAS, either implemented or conceptual, reveals that most of them are built around one or several main goals, motivating their develop-

[1] Note that CATO is the name of a product, whereas CGTO is an acronym for the presented concept.

Table 10.1. *Different main interests of roles and organizations in train traffic operation*

Interest	Main Interest of	Examples
Safety	Everyone	ATP, ETCS (not strictly DAS)
Energy consumption	RU	CATO [99], Trainguard [77]
Tear and wear	RU	Leader
Punctuality	Everyone	FARE [61]
Infrastructure capacity	IM	CATO, FARE
Replacing paper	RU	EBuLa [42], TrAppen
Awareness of surrounding traffic	TD	RouteLint [4] (CATO)
Ability to plan ahead	TC	
Maintenance of traffic flow	TC	
Controllability of traffic	TC	(CATO)
Gain expertise	TC, TD	(CATO)
Efficient communication	TC	RouteLint
Comfort	TD, RU	Often implicitly included
Passenger information	RU	TrAppen

Note. This table lists main interests, their corresponding owners, and driver advisory systems built to support these interests. TC = traffic controller. TD = train driver.

ment. One of the main reasons for implementation of DAS is 'eco-driving', a term that primarily refers to energy saving but can also refer to preservation of infrastructure or rolling stock. Table 10.1 lists most interests in operative train traffic. These can be personal interests, or, in terms of GMOC, goals, of train drivers and traffic controllers, or organizational interests of the infrastructure manager or railway undertakings. Examples are given for DAS motivated by some of these interests. Note that this list does not claim to be exhaustive, but it is an interesting by-product of our analysis.[2]

Leader is one of the DAS with a longer history; it began as a system in the United States with a goal to inform the driver of long and heavy freight trains about forces affecting the train. This puts the train itself in focus.

There are two main possible ways to save energy with a train. One is to apply a more defensive driving style, with lower high speed, reduced or regenerative braking that can recover energy and (under certain circumstances) feed it back into the overhead contact wire, longer coasting periods, and in general a better utilization of the track gradient. Most existing DAS only implement this option.

The second option to save energy is to avoid unnecessary stops. Unnecessary stops occur when a train approaches a track segment that is still blocked by another train. This can happen when a faster train catches up to a slower train or when two trains are scheduled to meet or overtake. At a meeting, the

[2]This work has not been published in a research paper but was presented at German Rail Human Factors, May 2014, in Braunschweig, Germany.

train that is supposed to pass through must not arrive at the meeting point before the other train manages to stop and the signalling system is able to switch the signal. To give the correct advice, a DAS needs to know a couple of details, such as position and plan of the involved trains and the state of the control system. Basically, CGTO is needed, coordinating the recommendations with the traffic control system. To ensure avoidance of unnecessary stops, all involved train drivers should follow the same advice, which requires a C-DAS to be installed on all involved trains.

DAS aiming for energy saving usually focus on speed advice, either in the form of 'regimes' ([3] – used, for example, in Trainguard MT [77] – that is, recommendations for the driving style to apply in the current situation, or in the form of a speed curve or 'motion profile' (e.g., CATO). Punctuality is another common reason for DAS. The interface of FARE makes its focus on punctuality very clear, as it displays to the driver the amount of delay that could be reduced during the next time slots [61]. Implicitly, punctuality is included in most other DAS as well, as the speed recommendation is usually calculated in accordance with the original timetable or (in C-DAS) a real-time plan.

Often train drivers have to handle quite an amount of paper, including daily schedules, forms to fill out on demand (e.g., to document safety-relevant incidents), documentation, and most importantly the timetable book, which is usually a quite thick book containing all routes in a network, where every segment is listed with its accompanying speed limits and additional information. Railway undertakings, but also infrastructure managers, are interested in reducing the use of printed documents and in establishing electronic versions. Examples are EBuLa (Deutsche Bahn) [42, pp. 281 f.] and TrAppen (a DAS developed by SJ, the main operator for passenger traffic in Sweden).

DAS usually address more than one interest, but often a main interest can be identified, while other interests are handled implicitly. For instance, reducing energy efficiency and wear on material usually results in a smoother driving style. So the interests are beneficial to each other and even lead to improved passenger comfort. Safety is usually not directly addressed by DAS, as this is the task of safety systems such as ATP and ETCS (European Train Control System).

From Table 10.1, one can see that the majority of DAS address interests that are directly beneficial to railway undertakings. This is natural, as they are the ones doing the investing. Rarely (explicitly) addressed are aspects that give the train drivers additional information about the surrounding traffic and primarily have a positive impact on the collaboration between train drivers and traffic controllers. An exception is RouteLint (developed by ProRail). Here a main goal was to improve the communication between train drivers and traffic controllers by improving the drivers' situation awareness [4].

Consideration of the collaboration between train drivers and traffic controllers is a unique feature of our perspective on the development of concepts for DAS. Additionally, to support the main interests (i.e., safety, energy consumption, punctuality, etc.), we try to fill in the blanks in Table 10.1. We want to improve train drivers' work environment by providing information that supports their awareness of the traffic situation. We also aim for a DAS that supports the train drivers in gaining expertise instead of just giving a plain command to follow. But most unique is our effort to even include traffic controllers' perspectives. By providing more efficient communication channels and the ability to automatically communicate the current traffic plan, we give traffic controllers the ability to plan further ahead and to gain better control of the situation. Our goal is to develop CGTO-DAS that supports

- traffic controllers' observability (and subsequent development of mental models) through feedback of information from the train/DAS to the traffic control system
- train drivers' observability and situation awareness by presenting more information about the current traffic situation around them
- traffic controllers' controllability, as we give them the possibility to automatically communicate updates of their traffic plans to drivers, resulting in significantly improved means to influence train drivers' driving styles (see Section 12.3.2)

These goals have partly been met by the CATO system and are further discussed in Chapter 14.

10.3 Goals – Optimization Criteria

Optimization is a common theme in this thesis. To be clear, this section defines what is meant by 'improvements', 'optimal', or 'better' situations.

On one hand, we have the perspective of the user. One common goal of users is to deliver good quality. When the cognitive workload of a task exceeds the capability of the operator, quality will suffer. A too low workload can have similar effects, but this aspect is not a focus of this thesis. Here improvements for the user mean the ability to perform better, that is, to deliver high quality; the ability to maintain a workload that does not exceed one's limits; and in general an improved usability provided by the systems, for example, through more efficient tools and improved controllability.

On the other hand, we have the perspective of our domain, train traffic. In this context, improvements and more optimal solutions mean safety, capacity, punctuality, energy consumption, wear on the material, or passenger comfort. These goals can come in conflict, and persons from different organizations and in different roles will probably not have the same prioritization. Safety is usually named as the most important, but because the Swedish automatic train protection is very reliable, great confidence in this system has somewhat lowered the awareness for priority of safety.

Our work deals strictly with systems operating on top of the existing safety systems. We give a higher priority to punctuality and capacity than to wear on the material and energy consumption and passenger support. Another of our goals is to sustainably improve the situation; this means that we strive for solutions that give users a better understanding of how to achieve such optimization of capacity and energy consumption. With this, we also want to leave the users the autonomy to apply their own prioritization.

10.4 Our Studies in the Area of Train Traffic Control

The HCI group at Uppsala University started their research on future train traffic control about twenty years ago. The basis for this research has been the ongoing collaboration with Trafikverket. During this period, several new concepts for train traffic control have been developed and tested. A result of this collaboration is the STEG (Swedish for ‘Control via an Electronic Graph’) system for train traffic control, which is based on a digital time–distance graph. It incorporates a new control strategy, *control by awareness*; a *real-time traffic plan*, allowing for sharing information about current and planned deviations form the original timetable; an *automatic execution* of the current traffic plan; and ways to communicate with the train drivers. Figure 10.4 shows the parts of operational train traffic control covered by STEG.

From my point of view, the development can be roughly divided into two phases: the time until 2010, during which a prototype was developed, deployed, and evaluated in Norrköping, and the time since 2010, when STEG has been deployed at the train traffic control centre Boden. Until 2010, traffic control and train driving were mainly considered separately. STEG was developed as a single system to replace one traditional workstation, without communication channels to train drivers [83], and train drivers’ work environment was studied from a safety perspective, with a focus on the ATP system and physical conditions such as workload, stress, and fatigue [55]. I joined the project in 2010, which means that I was mainly involved in the work in connection with the deployment in Boden and further development during this period, aiming for better collaboration between train drivers and traffic controllers.

In the following, four studies are presented. In relation to the train traffic process shown in Figure 10.1, the studies cover the parts shown in Figure 10.4. The first study is a summary of the development of STEG until 2010, with a focus on its user interface and the new control strategy it incorporates. The second study evolves around the collaboration between traffic controllers and train drivers. The third study covers the deployment of STEG in Boden, including a few changes to STEG to support collaboration, and CATO, a DAS that is connected to STEG and is key to the collaboration with the traffic controllers. CATO has been developed by a private company, Transrail, whereas

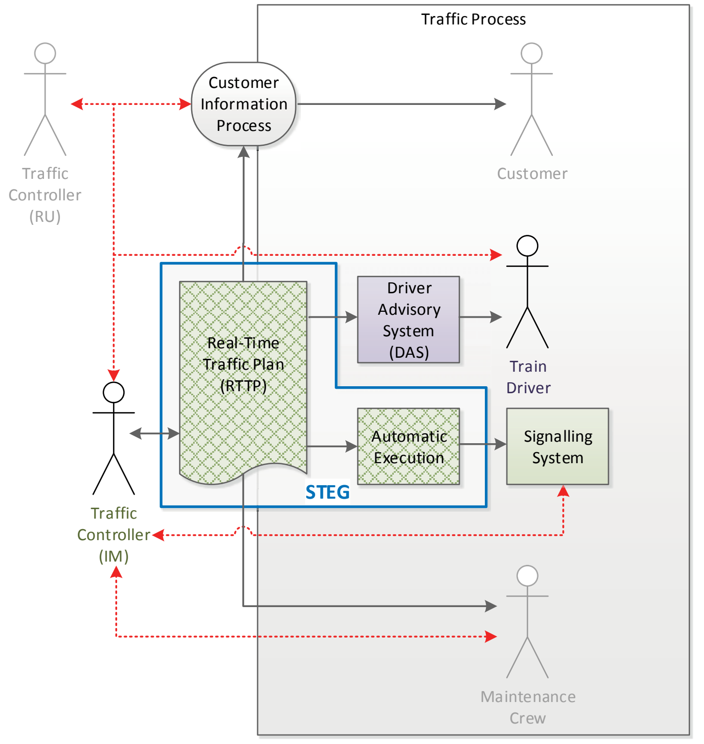


Figure 10.4. An isolated part of Figure 10.1, showing the main topics of this thesis and the part of the process covered by STEG.

its interface design has been supported by our group. The last study concerns train drivers. Its goal is to develop a novel DAS, putting the needs of the human operators in focus (in comparison to the companies' motivations; see Section 10.2.4).

10.5 Related Work

Research often focuses on aspects of train traffic control or train driving. It rarely deals with the development of new concepts and systems integrating traffic control and train driving. Exceptions are studies from Switzerland [61, 78]. However, their focus is automation, a human factors perspective has only been taken by Fenix et al. [31].

When it comes to studies of traffic control systems, there exists research on analysis of aspects of the traffic controllers' and signallers' work, for example, concerning situation awareness [35], analysis of the operators' disruption handling, design of decision support systems [18], and analysis of graphical support tools in traffic planning at stations [16]. Our own research concerns analysis of the traffic controllers' work [5], development of a new control strategy [54], and development and evaluation of systems implementing this strategy [45].

Compared to systems for train traffic control, DAS are small and relatively simple. Several systems have been designed, developed, and evaluated (e.g., [4, 2, 61]; see [71] for a review). In Sweden, the CATO system has been developed, a DAS that shows recommended speed based on the current traffic situation and plan [99].

Collaboration between the roles is a critical prerequisite for effective and safe operation. Sharing of information is an important aspect, as it creates 'shared situation awareness' [82]. We have analysed collaboration between train drivers and traffic controllers (Paper IV) and present concepts for advanced systems for traffic control and driver advice, based on these findings, in Paper III and Chapters 12 and 14.

Another area of interest is the development of algorithms for intelligent railway systems. A recent review of the field summarizes the main challenges [17]. Among these is the limited view on the traffic situation as a whole as well as the deficient integration of these algorithms in the operators' workflow. In the past, automation has often been designed with a very limited consideration of its integration into the work process and the traffic situation as a whole. Therefore it is often not supportive in critical situations, and traffic controllers tend to turn the automatic systems off to be in full control. Exemplary guidelines to avoid the pitfalls of automation in this domain have been developed [9].

In general, not only related to automation, we observe the problem that it is common in train traffic control to sub-optimize. One goal of our work is therefore to take a more holistic perspective. The concepts that are presented are developed on this premise.

11. A New Control Strategy for Train Traffic Control

This chapter summarizes the development of a new control strategy for train traffic control, control by awareness, and a novel system for train traffic control, STEG, which implements this strategy. The development has been documented in earlier reports [54, 83]. Paper I describes how GMOC has been applied to analyse train traffic control and to design and evaluate new systems. Interface elements in relation to GMOC will be highlighted to illustrate how the analysis resulted in a concrete design.

11.1 Development of the Operative STEG Prototype

The development of STEG, the new system for train traffic control, followed the iterative user-centred system design cycle as presented in Chapter 8, proceeding through phases of analysis, design, and evaluation. GMOC was used as a way to structure the results during each phase and to pass them on to the next phase.

Extensive observations and interviews were performed to *analyse* the current work situation of train traffic controllers. The controllers were experienced and came from different control centres, but all of them worked in an environment comparable to the one described in Section 10.1.1. Interviews were un- or semi-structured, performed both during the actual work situation and after a work shift. Observations were made during longer time periods, where the work of different traffic controllers was followed for several workdays. Interview transcriptions and observation protocols were coded with respect to aspects related to goals, models, observability, and controllability. This formed a first version of the description of the current work environment in terms of GMOC.

As a transition from analysis to *design*, a long series of workshops following the vision seminar process (see Section 8.2.3) was performed. A group of eight experienced traffic controllers participated over a two-year period in monthly full-day workshops. The first part of the process was used to describe and analyse today's situation and systems and to identify problems and important areas for improvement. The starting point was the description received from interviews and observations, which were complemented during the vision seminars.

The remaining period was used to iteratively specify ideas for future systems, new control principles, and design requirements. Several alternative prototypes were evaluated, starting with sketches and ending with functional prototype systems. The participants were involved in summarizing and analysing the results during each workshop. Prototypes have also been evaluated by traffic controllers who did not participate in the workshops, that is, by those who were not directly involved in the design.

The discussions during the workshop were guided by the GMOC model; problems and limitations that the participants experienced in the current support systems and organization were analysed using the elements goals, models, observability, and controllability. This allowed for defining an improved work situation according to GMOC and thus motivated each iteration of the design and suggestions for new principles and organizational changes.

If, for instance, a problem in observability was found, the first step was to see how this could affect the other elements. It is not uncommon that improved observability facilitates prioritization of goals or development of mental models. The second step was to analyse if it was technologically or organizationally possible to increase observability or which further development would be needed to address the identified problem.

The basis for the *evaluation* was a fully functional prototype. It was used operationally in the traffic control centre in Norrköping by four different traffic controllers during their usual shifts. During the evaluation period, STEG was in operation for more than seven hundred hours. It covered an area with single and double track lines. The period included some severe disturbances caused by infrastructure reconstruction.

During the evaluation, the operators were observed and interviewed, and they participated in collegial verbalization sessions (see [30] for a description of collegial verbalization using the STEG prototype and for a comparison between the two verbalization methods). The evaluation aimed for a description of the new work situation in terms of GMOC. This allowed for comparing the initial situation and for evaluating whether the new systems and concepts improved the work situation related to the GMOC elements. Another goal was to explore if the new environment had generated a need for further improvement or if an important aspect had been overlooked.

11.2 Train Traffic Control with STEG

An early conclusion of the analysis was the need for a shift in the traffic controllers' strategy. In the traditional systems, they applied *control by exception*; that is, conflicts are identified and handled 'last minute'. The traditional control systems focused on the present traffic situation, without facilitating projection and planning ahead. *Control by awareness* was proposed as a new strategy, where quality of the traffic plan is improved by planning ahead of time [54]. STEG was developed together with Trafikverket to achieve this shift in the traffic controllers' control strategy. The system includes an automated traffic plan execution, which supports users and encourages them to spend their cognitive resources on the planning process, not on manual plan execution.

Figure 11.1 shows a small part of the STEG user interface. Its main view is a time–distance graph, inspired by the traditional printed graph. Here the traffic controllers can use the mouse to change trajectories of trains in the plan. Using the scroll wheel, they can move a trajectory closer to real time, meaning to put it forward or 'prepone' a train. In the same way, they can postpone a train. They can add stops to a train and configure track usage at a station with one of very few context menus. It is also possible to adjust the planned speed for a train, that is, to change the slope of a trajectory. The plan scrolls in real time and identifies the commands that need to be executed. In the following, the interface elements shown in Figure 11.1 are described in terms of goals, models, observability, and controllability, as well as the automatic plan execution function.

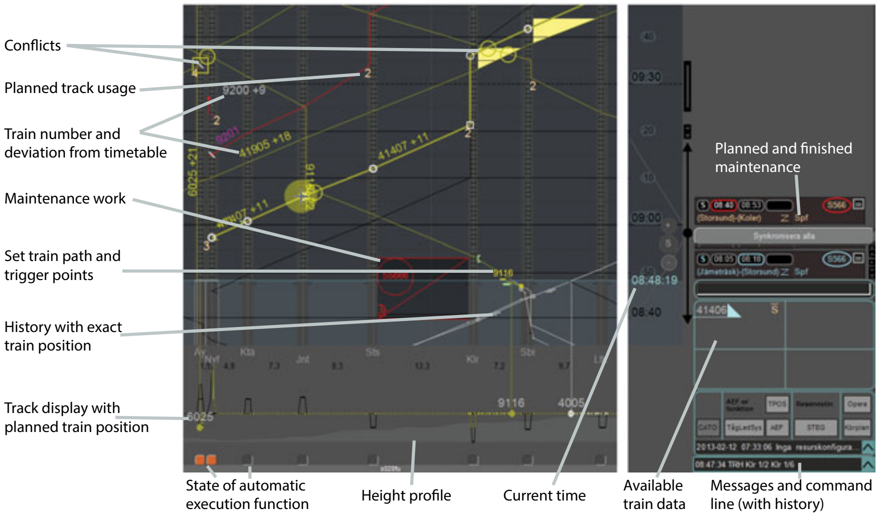


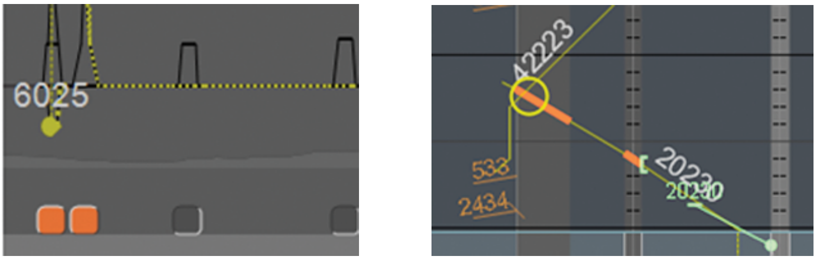
Figure 11.1. A part of the STEG user interface, including descriptions of the different elements.

11.2.1 Automatic Plan Execution

A central point in the development of STEG has been the discussion of automation. Traditional automation in railway systems has often proved problematic (see, e.g., [9]). Especially in scenarios with disturbances, when the traffic controllers need most support and controllability, automation tends to act in ways conflicting with the controllers' plans. Therefore the main principle for the automation in STEG is to never change the traffic plan. Additionally, the concepts presented in Chapter 6 were used as guidelines. This means that the automation should be non-autonomous, as transparent as possible, controllable, and observable, that is, clearly showing its configuration and status.

The result is an automatic plan execution function that only executes the plan exactly as defined by the traffic controller via STEG. This execution is done by sending requests for train paths to the infrastructure. To give the human operators as much freedom as possible, these requests are sent just when the execution is needed, that is, with the necessary margins for the infrastructure to respond to the request. Depending on the infrastructure hardware, this usually allows changes in the plan up to two minutes before execution.

Figure 11.2 shows interface elements of the execution function in STEG. The function can be disabled for certain train numbers, single stations (see the filled orange squares in Figure 11.2a), or the whole area at once. STEG displays the train paths already set for a train (see the green '[' symbol and line in Figure 11.2b). This can be paths set by automation or commands manually typed by the controller. If manual action is needed, for example, because the automation is disabled at a station, this is shown by a thick orange line (see Figure 11.2b).



(a) Interface elements to trigger the automatic execution.

(b) Hints in the graph indicating the need for manual plan execution (orange parts) and the currently set train path (light-green parts).

Figure 11.2. Elements of the STEG user interface related to the automatic execution of the plan.

In terms of Parasuraman, Sheridan, and Wickens's model for types and levels of automation [72], the automatic plan execution can be classified as level 7 (automatically executing an action and informing the user) for action implementation. However, this is only true for the final step of transferring the plan to infrastructure, not for conflict resolution in the plan. This exemplifies the difficulties in practice of assigning a general classification to a system in such a model, because the exact definition of the different levels and stages is debatable and context dependent [22].

11.2.2 Interface Elements Supporting Goals

Supporting goals means that the operators should be supported in formation of concrete goals, identification of conflicts between goals, and assigning their prioritization. This can happen explicitly, for example, by highlighting events that might have a negative influence on goal achievement, or implicitly by supporting the traffic controllers in their tasks and thus in achieving their goals. The STEG interface supports the following traffic controller goals: efficient traffic control, that is, increased capacity utilization; reduced number of stops; safety; punctuality; control of workload; and accuracy.

Most of these goals are supported implicitly. Implicitly supporting a goal is often the same as explicitly supporting controllability; this is logical, as a high-level goal is usually to achieve control. Most goals can be reached by applying control by awareness. The whole STEG interface is designed towards this strategy. The main display area shows the plan for the past few minutes and at least an hour ahead (the interval can be freely adjusted; it is also possible to review the traffic flow during the past twenty-four hours to check for uncommon events), meaning that planning ahead is given the most display area. The automatic plan execution function handles the transmission of the plan to the traffic control system, thus the traffic controllers have more mental capacity available for planning. The large planning area and display of the exact train position in the history allow the traffic controllers to see more exact train runs and to test the effect of a change in the plan on the future development of the traffic process. With an overview of the plan ahead and the automatic execution function 'remembering' their decisions, traffic controllers have a greater possibility to control their workload, as they can decide to resolve future conflicts in the present or at a later point.

STEG itself does not include any functionality that is directly connected to safety; it is built on top of the existing safety systems, most importantly ATP. From the signalling system, it gets the safety classified information of a train's position based on it passing the border to the adjacent block and the end of the set train path. Beyond this, STEG indicates when the plan contains unsafe conditions – these are maintenance work, other traffic not covered by ATP, and track or line conflicts. The latter covers situations where more than one train is planned to use the same block at the same time, for example, when meeting on a single track line, trains catching up, or (wrong) track selection at a station or double track line. Parts of a plan that contain conflicts will not be executed.

To some extent, STEG also supports prioritization of goals. Some parts, such as conflicts, are visualized with a higher contrast to the background, directing the operators' attention. However, because priorities of goals are adaptive and highly dependent on the situation, we used this effect to a minimum. The different contrast is rather used as a classification of importance of information, where important information means that which is necessary to evaluate the importance and fulfilment of goals.

11.2.3 Interface Elements Supporting Models

There are two main ways to support (mental) models in interface design. These are to support the development of the models themselves or to provide information that is relevant for gaining situation awareness (via a mental model) and to make decisions on further actions.

An important factor for development of models is, as already discussed, the relation between control actions and produced feedback. In STEG, train traffic is controlled by a plan and its automated execution. Two important types of feedback are the state of the execution (i.e., if a command, sent manually or from the automatic plan execution, has been executed and the infrastructure has reacted) and the relation of the real traffic situation to the plan (i.e., if a train has reached the position that was planned). Figure 11.2 b) shows an example of feedback on control commands. It shows the result of a command, meaning the range until a train path has been set. This feedback is important for learning and understanding the functionality of the automatic execution function.

The relation between plan and reality can be observed by information on the trains' positions in the history in STEG. A deviation from the plan might imply that there is something wrong, possibly with the infrastructure or the train. If this deviation is observed continuously, this might mean that there is a permanent problem, possibly in the timetable. Precise feedback on the train position will also allow development of a model of the driver's driving style. Today, driving close to the speed limit is a typical (though not the only) strategy of train drivers. However, eco-driving is becoming more important, so drivers are tending towards changes in their driving strategies. More precise feedback on train movement will show that drivers are not only driving close to the speed limit. STEG supports the display of GPS data and thus facilitates development of a more adequate model.

There are many more sources of feedback in STEG; examples are the display of a train's delay (feedback from the train traffic process), the indication of disabled automation (feedback from the automatic execution function), or indication of conflicts (feedback from interaction with the STEG interface).

It is our goal to display information (i.e., provide observability) that supports an understanding of the situation as precisely as possible and generation of actions as adequately as possible. If STEG indicates a conflict, it should be, given the traffic controllers' general understanding of train traffic, as simple as possible to understand the reasons for the current plan producing this conflict. At the same time, if information display is designed to facilitate models, mental workload can be reduced. An example is the track layout in STEG. It is displayed to correct scale, thus the traffic controller can judge the distance between two stations directly, whereas this information is seldom available in traditional track diagrams (see Figure 11.3). In STEG, the traffic controller can estimate the passing time of a train through a block directly (using, e.g., a linear model as simple as ‘short distance on screen equals short distance in reality’). In the traditional system, the traffic controller needs experience or local knowledge to judge the distance.

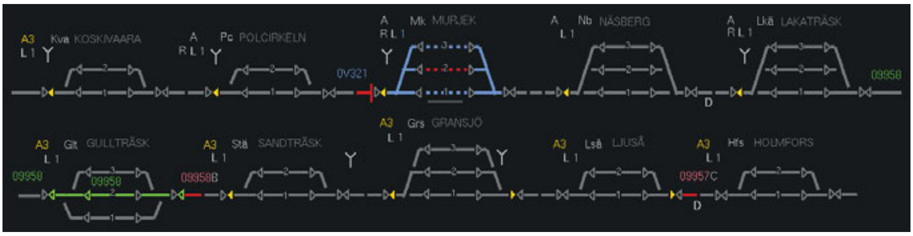


Figure 11.3. Track diagram in the traditional traffic control systems do often not contain information about distances between stations.

11.2.4 Interface Elements Supporting Observability

Conforming with our recommendations from Paper II, STEG is built on the principle to display all information that is needed for efficient train traffic control, rather than limiting the amount of information, following the rationale of avoiding information overload. Instead, information, or rather cognitive overload, is avoided by visual sorting, grouping, and prioritizing information and by reducing the time needed to search for specific data. As a first step, we integrate the information that is traditionally provided by different systems into one view. The STEG main view displays, for example, the track diagram, train positions, train delays, detailed information on a train and on maintenance work, route settings, track selections, timetable information, and current planning.

The main focus is on the time–distance graph. It contains historical data and the plan for the next twenty-four hours. It is a proven tool and has been used on paper for decades. Digital versions are used in simulations and planning tasks such as timetabling or creation of a daily traffic plan, but in real-time traffic control, track displays are most common. Despite this, the time–distance graph is a great visualization of train traffic. With a bit of training, one can immediately see where and when a train is scheduled to stop, where trains meet, and how fast a train is planned to go. It creates patterns that, for instance, reveal two trains following the same schedule (e.g., commuter trains going every thirty minutes), track segments with low speed limits, or regions (in terms of time and area) with high or low capacity utilization. Planning in this view has several advantages, most important of which for us is the focus on the time frame beyond real time, in comparison to the view limited to real time in traditional track diagrams.

Especially the more exact data that can be obtained from trainborne GPS units improves observability. With this information, the traffic controllers are no longer dependent on imprecise track diagrams. They can see exactly where a train is located and how fast it actually moves. This allows the traffic controllers to monitor trains' deviation from the plan and to re-evaluate the plan.

Interaction with STEG is designed to minimize the use of context menus and dialogues. As most information is displayed in the main view, context menus and dialogues are seldom needed to display more information. Another way to reduce the need for dialogues is by displaying less important information in the column on the right-hand side of the STEG display. Here one area is reserved for detailed information about the selected train. As the information is always at the same position, it is easy to locate. Context menus are sometimes needed for data input, for example, to assign a track at a station to a train. The reason for avoiding context menus and dialogues is to prevent situations where they hide important information.

11.2.5 Interface Elements Supporting Controllability

Traffic controllers reported improved controllability of the traffic process, as the interface supports planning ahead and predicting future developments in the situation. Substantial impact on controllability comes from the automatic plan execution. As expected, it frees additional mental resources, allowing the traffic controllers to plan ahead. They have more time to react to upcoming conflicts and thus are able to find more optimal solutions. It is the key element supporting the shift of the traffic controllers' strategy to control by awareness. Controllability and observability over the automatic plan execution are given as discussed earlier (see also Figure 11.2).

The traffic controllers especially valued the increased accuracy with which they could reply to requests by train drivers. This is mainly an effect of the simplicity of updating the digital traffic plan compared to changing the traditional graphs on paper. This leads to more accurate updates and gives traffic controllers the feeling of really being in charge and knowing what is going to happen. The traffic controllers are also supported in managing their workload. By observing and changing the traffic plan, the traffic controllers are able to identify and handle critical situations in advance. Thus they are less often confronted with situations where many actions have to be taken in a short time. As a result, controllability of the process is increased.

11.3 Evaluation of the First STEG System

The development and deployment of STEG in Norrköping has been successful. The traffic controllers were so satisfied with the results that it was decided to keep the system in operation – which was not planned at the beginning of the project. Trafikverket even decided to deploy the system at a complete control centre (see Chapter 13).

The automatic execution seems to be a key factor in the acceptance of STEG and for adaptation of the new traffic control principle, control by replanning [54]. Having it take care of plan execution in real time allows the traffic controllers to focus more on the future plan. At the same time, the automatic plan execution takes over an important part of the traditional work of the traffic controller. It is thus important that the automatic execution be reliable and that it clearly inform about possible problems. This is necessary so that the traffic controllers can trust it and really focus on the plan.

Goal prioritization with support of the interface has shown to be difficult. It is definitively possible to direct the operators' attention towards parts in the plan that might affect goal achievement negatively. However, it is hard, or even impossible, for a system to evaluate which events might have the most critical influence. In the end, too much visual stress on certain events can even mislead the operators: as an example taken from STEG, too many conflicts at one point will lead to an accumulation of conflict symbols with high contrast. On one hand, this clutters the display and might make it hard to actually understand the current situation. On the other hand, this attracts attention, even though the reason for the conflicts might be at a different point in the plan.

The time–distance graph is an excellent visualization. However, it does have its limitations. The biggest problem is its limitation to display lines with several tracks. As train traffic is commonly organized in a way that in double-track networks, each track is reserved for a certain direction (e.g., northbound or southbound trains), double tracks can be managed (this scenario was included in the STEG test area). One or two additional tracks seem possible, for example. with use of different colours or shades, but this has only been implemented in a graphical prototype. It is impossible to manage larger stations with several platforms and complex layouts using time–distance graphs. Our group is currently involved in developing a new control system for Stockholm Central Station, which is based on a time–resource graph. Different platforms, track segments, and points are the resources.

A final point for discussion is controllability. It remained a weak spot. Several improvements in STEG, especially the new control strategy, made it easier to handle the traffic process and caused traffic controllers to feel more in charge. However, the actual controllability of the traffic process has not considerably improved. This has led to further exploration of the traffic process and to collaboration between traffic controllers and train drivers, as presented in the following chapters.

12. Supporting Collaboration between Train Drivers and Traffic Controllers

During evaluation of the STEG prototypes, it became quite obvious that collaboration in railway traffic in general is important but hardly supported by the current systems and organizations. Consequently, we analysed the collaboration between the different actors. We expected to find ways to make the collaboration more efficient and effective, leading to a better work environment and more potential for optimization of traffic control. Details from our analysis of collaboration are published in Paper IV. Paper III describes the developed concepts to deal with the discovered problems in collaboration.

12.1 Analysis of Collaboration

In our earlier studies, we used GMOC to analyse and structure the work situation in relation to the traffic process and the available support systems. Questions were of the nature, "How can traffic controllers directly influence the traffic situation?" In this study, we focused on the collaboration among and between train drivers and traffic controllers. This is meant to adapt the view of GMOC, as explained in Section 5.4. Typical questions were as follows:

- How are goals correlated or conflicting?
- How do operators estimate their colleagues' performance?
- How can colleagues help an operator to achieve additional observability or controllability of the traffic situation?

We started with the results from earlier studies. Existing recordings and documentation of observations and interviews were analysed from this new perspective. To further investigate our findings, we conducted additional interviews and observations. We interviewed one traffic controller who was experienced in STEG and involved in its development in Norrköping on his personal view on collaboration and how STEG has influenced collaboration. We observed two traffic controllers during half a shift with subsequent interviews. They were controlling a line that was recently upgraded to ERTMS (European Rail Traffic Management System). This allowed them to see a train's position and speed in more detail; that is, they could see if a train was located at the beginning, in the middle, or at the end of a block (a block was basically divided into three subsections), and they could see if a train had stopped or was moving. At last, we joined a traffic controller in northern Sweden who was using a traditional traffic control system. Via an interconnected blue-tooth headset, we were able to monitor (but not to record) all communication over the course of one morning. The communication involved that particular traffic controller and mostly train drivers in a typical traffic situation.

From this analysis, we derived new concepts that were subsequently integrated into the systems. Because of a tight schedule, we were only able to conduct two whole-day workshops together with the system developers in the later course of the project. One workshop included one traffic controller experienced in STEG; three train drivers experienced in CATO participated in the other workshop. The workshops had as a goal to discuss the new concepts and ways to integrate them into the interfaces. Evaluation of concepts and interfaces was part of the deployment.

12.2 Results of the Analysis

Our studies revealed that increased collaboration bears good potential to improve train traffic. To some extent, operators collaborated even in traditional systems, however, particularly tedious communication via telephone (see the red dotted lines in Figure 10.1, p. 93) hinders efficient collaboration. Communication is especially needed to improve observability and controllability.

In the traditional systems, traffic controllers can only see positions (based on blocks) and delays of trains. Experienced traffic controllers have mental models helping them to predict a neighbour's plan relatively well, but it is not possible to know exactly how a colleague plans to lead the train towards one's own control area. Commenting on their collaboration with train drivers, a traffic controller said, "They are our only eyes on the tracks." The traffic controllers actually value information that helps them to estimate the situation at the tracks, both for safety reasons and for support in predicting train movements.

Another quote, common for the way train drivers are seen, is, "They just hit the gas until they reach a stop signal, and then they complain." This reveals several things. First, it shows that there exist prejudices as the predominant models between train drivers and traffic controllers – similarly, train drivers had the impression that they are often forgotten and that traffic controllers are not really able to imagine the situation at the tracks. Second, it shows the lack of controllability traffic controllers have, as the only effective way to control the traffic is the signalling.

Train drivers can contact other train drivers or the control centre of their railway undertakings to get information about the current situation, but most important in relation to the surrounding traffic, to plan and optimize their journey, would be information from the traffic controllers. We very rarely observed that train drivers were contacted and informed about, for example, possible waiting time. And if so, this happened reactively. This means they had to contact the traffic control centre when the train had already stopped or when a larger disruption was right ahead. Conversely, we did not commonly observe that train drivers contacted the traffic controllers to inform about events that could delay a train (e.g., loss of an engine). In one of our discussions with the train drivers, they stated that they sometimes avoid this, because it might mean that the traffic controller would prioritize them even less. But it turned out that train drivers had other ways to control the traffic: they knew that a certain behaviour, namely, triggering mechanisms in signal boxes, would give either them or a meeting train an 'advantage'. They sometimes use this for what they believe is best for them or the traffic as a whole.

In relation to goals, we could identify potential conflicts on an organizational level. In certain situations, the goal of railway undertakings is to reduce costs by more energy efficient train driving. This could conflict with the infrastructure manager's goal to reduce delays. Additionally, all actors had models describing the behaviour of others. For example, traffic controllers tried to interpret and predict the train drivers' behaviour, while train drivers tried to predict actions of traffic controllers and other drivers, sometimes using this information to change their driving strategies to gain an advantage over another train or to contribute to a better traffic flow. But because of the limited observability, these models often lead to incorrect (in case of insufficient information) or biased (in case of models built on prejudices) predictions. Combined with different goals, sub-optimal results are foreseeable.

We can conclude the following:

1. The traffic controllers had a too low controllability on the traffic process.
2. All actors had deficient observability leading to construction of insufficient models in form of prejudices about their colleagues.
3. Information about the traffic process is widespread and not available for all actors with sufficient accuracy and actuality. The process lacks an efficient way to communicate up-to-date information.

12.3 Developed Concepts

Following the results, we developed a shared real-time traffic plan as a concept to overcome deficiencies in collaboration and communication.

12.3.1 Introducing a Real-Time Traffic Plan

A closer look at the information available to the actors in train traffic operation revealed that their access to different kinds of information about the traffic process and plan depended substantially on their organization and role. A large portion of the information is spread out, only available at certain organizations, to certain roles, and even only to certain actors.

Most important is the access to a traffic plan. We identified large discrepancies between richness and actuality of information in the different plans available to different actors. Furthermore, each actor built his or her own interpretation of the available information, forming a personal traffic plan. If these plans differ considerably, inconsistent actions are a result, leading to a sub-optimal traffic situation.

These inconsistent plans are a result of lacks in observability, but especially interesting is their impact on controllability. Traffic controllers are supposed to direct trains in an optimal way. Adjustments to the original timetable are inevitable. Train drivers are supposed to follow the traffic controllers' plans. However, as soon as they enter the train cab, they have most often no means to observe further changes. They will only realize that something has changed when they approach a restricting signal.

We propose to maintain a digital version of the plan created by the traffic controller, a real-time traffic plan (RTTP, see Paper III and Figure 10.1), as a solution. This RTTP is stored at the traffic control system, where it can be accessed by the traffic controllers and other actors and systems that need this information. The concept of an RTTP has been implemented by Trafikverket for the traffic control area Boden, where it is in live operation, and by partners in the On-Time project, where it can be demonstrated together with the traffic simulator Hermes.

In its current implementation in Boden, the traffic controllers can change the RTTP of their areas of authority. STEG visualizes this part of the RTTP, allows interaction, and writes changes back to the control system. From there it can be distributed to other actors and executed with commands via the traffic control system to the infrastructure. Figure 12.1 shows a possible architecture. It includes a driver advisory system (see Section 10.2.4), which is necessary to improve controllability of train traffic, as will be explained in the next section.

12.3.2 Improved Controllability through Target Points

With the RTTP in place, the next task was to support efficient and preferably automated communication between train drivers and traffic controllers. We used *target points* (e.g., [3]). For our purpose, their atomic form includes a train number, a position, a time, and a speed. This means that they tell a train driver to pass (or stop at) a position at a certain speed and time. Additionally, explanatory context information for such a target point can be added, such as ongoing track work or a meeting with another train.

STEG creates and adjusts target points automatically. Most of the time, they should be invisible to the traffic controller. They are only shown for a few seconds after their creation, for example, when a meeting has been re-planned. During this time, but also continuously based on train movement, they are evaluated by the DAS or a centralized server for the DAS. Only if a

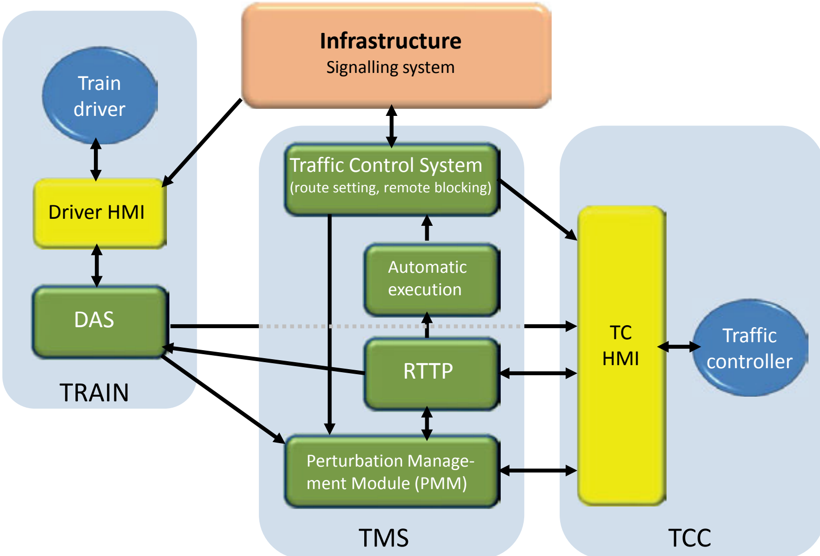


Figure 12.1. Architecture of systems for train traffic control with a real-time traffic plan. This view is based on the On-Time project, including a module for automatic perturbation management.

target point is or becomes unrealistic is this indicated to the traffic controller. Reasons for this can be trains that accumulated too large delays or meetings that are planned too tightly.

The aim is that target points allow the train drivers to follow the RTTP as accurately as possible. Our hypothesis is that this will lead to improved traffic controllability. This will reduce the necessity of re-planning actions and improve the ability to plan tight (but feasible) meetings, which will in turn improve the possibilities to recover from delays or to increase capacity.

During our sessions with the traffic controllers, some of them estimated that far more than half of the changes they have to apply to their traffic plans are necessary, because trains do not follow their current plan (but the original timetable/daily traffic plan). We expect that displaying the RTTP via a DAS will not only allow the train drivers to apply more optimal driving strategies but will also reduce the traffic controllers' workload needed to update the plan. Instead, additional mental resources can be used for re-planning, which in turn further supports the new control strategy of control by awareness.

13. Deployment of STEG and CATO

For Trafikverket, the next step was to deploy STEG at the control centre in Boden. Every workstation should be equipped with STEG, meaning that it would be used to control a large part of the railway network in northern Sweden. This includes the iron ore line, consisting of long, single-track sections, where long and heavy iron ore trains are operating close to the capacity limits. The mining company LKAB, operating most iron ore trains, was willing to equip its trains with a DAS, to be developed by a private company, Transrail. The idea behind this DAS, called CATO (Computer Aided Train Operation), was to support eco-driving and optimization of meetings between loaded and empty iron ore trains [99]. The system should give advice to the drivers that should help to avoid stops of the heavier and slower, loaded trains. The systems would thus implement the basics of a shared RTTP.

13.1 Our Involvement in the Deployment of STEG and CATO

Based on the operational system in Norrköping, STEG was deployed in Boden. First at a single station, for the purpose of demonstration and for the traffic controllers to become familiarized, and later at the entire control centre. Overall, this process was conducted by Trafikverket. Our task was to monitor its progress and to evaluate changes in the systems and work strategies. For this purpose, we planned and conducted three visits at the control centre. The first visit was planned at the beginning of the process, when only one station was deployed; the second after the completed deployment; and the third more than a year after the second visit, to see how STEG and the new concepts had been adopted. All visits were planned for one or two complete working days, including observations, interviews, and discussions at the workplace.

During the deployment period, we had access to screen captures, ‘STEG films’, from all workstations. With these, we could follow the progress and identify problems in usage of STEG. We also participated in regular telephone conferences, where progress was discussed. As the films indicated problems in the way the new system was understood and used, we conducted an additional training session between the first two visits. Technical issues that we identified have been reported to the system development group at Trafikverket.

On the basis of our analysis of collaboration and experience from the Train project [55], we supported the development of CATO. Similar to the visits at the control centre, we also visited LKAB, operating the iron ore trains equipped with CATO, in Kiruna three times: first at the beginning of the deployment, when only one train was equipped with CATO; the second time when CATO had been installed on all trains; and the third time in coordination with the third visit of the control centre, to see how the concepts had been adopted.

13.2 Deployment of STEG

The development, deployment, and evaluation of STEG in Norrköping were very successful [45]. During the deployment in Boden, not all pitfalls had been avoided. A couple of the problems that occurred had been reported earlier [44].

13.2.1 Problems and Observations

An important difference, and source for many problems, was the fact that STEG was developed with and for the traffic controllers in Norrköping, while the deployment in Boden was more equal to the deployment of a ‘commercial, off-the-shelf’ system.

In this context, the different traffic situations at the two control centres have not been thoroughly considered. Two examples of the resulting problems are missing features and some cumbersome tasks. In northern Sweden, the weather conditions during winter are more extreme than in the south. Therefore the traffic controllers are used to commands that allow them to test points, for example, to see if they are frozen or blocked by snow, and to initiate heating. These important options were not available via STEG. Therefore the traffic controllers in Boden regularly had to initiate commands via the old systems, which contradicted the concept of automatic execution introduced by STEG.

The freight traffic, especially on the iron ore line, is characterized by many changes to the original timetable. It is common for trains to be cancelled, added, considerably delayed, or even early. The inefficiencies in some replanning functions that have already been criticized in Norrköping were even more disturbing in Boden. These problems should have been identified and solved in the deployment process.

Instead, new features have been introduced. In Norrköping, STEG was only installed at one workstation and without the possibility to share an RTTP with train drivers. In Boden, STEG was installed at several workstations, with the possibility to exchange information with the workstations controlling adjacent areas and trains equipped with CATO. These extensions have not been developed in a proper UCSD process and have been deployed without sufficient usability evaluation. During our visits to Boden, we realized that the integration of target points was insufficient. Not all unreachable target points were indicated to the traffic controllers; that is, they did not always have the chance to observe problems with target points and thus did not adjust the plan. This led to unnecessary irritation of train drivers. Another problem was that it was unclear why target points were not reachable – all the traffic controller could see was the same small symbol for each unreachable target point. An explicit visualization of the reason or at least a hint towards possibilities for solutions would be needed.

Another problem was the implementation of the automatic plan execution. The traffic control system in Boden comes from a different supplier, and it turned out that implementation of the execution function was not as straightforward as in Norrköping. Consequently, the function was not available or not working as desired at several stations. Therefore the traffic controllers did not trust the system and eventually disabled it completely. The consequences were severe. On one hand, this resulted in sharing of incorrect data with train drivers. On the other hand, the main idea of the automation was to relieve the traffic controllers of manual execution, instead allowing them to use their mental resources for planning ahead. Without the automatic execution, traffic controllers were not able to sufficiently shift their planning strategies. As a result, they mainly experienced STEG as yet another system they had to use without sufficient benefit.

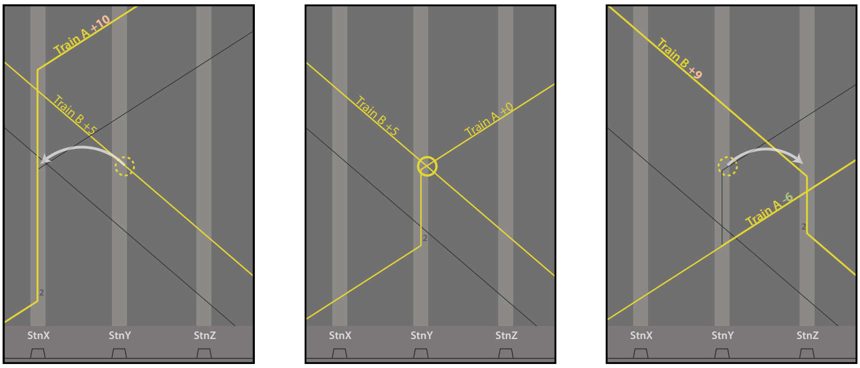
On top of that, Trafikverket was and is undergoing substantial organizational changes, starting with its formation from the fusion of Vägverket, the Swedish administration for road traffic, and Banverket, the Swedish administration of railway traffic. This fusion initiated a number of organizational changes, with larger restructuring of operational traffic control since autumn 2013. During this process, competencies have been removed from the traffic control centre in Boden, and rumours about it being closed worried the traffic controllers.

13.2.2 Suggestions for Improvement

Even though the organizational issues created unfavourable preconditions for the deployment of STEG, the observed problems that were related to the system itself were severe. We can divide the problems into three categories:

1. deficiencies in the original version of STEG
2. technical issues
3. additional features introduced without a sufficient design process

The main flaw in STEG itself was lack of efficiency of its interface in certain situations. We suggest support functions. These can be relatively simple, for example, helping the traffic controllers to 'clean up' a plan after removing a train (by eliminating obsolete stops after removal) or helping to resolve conflicts, which can be done following simple, logical rules (see Figure 13.1).



(a) 'Conflict' is dragged to StnX. The meeting is moved, giving priority to Train B.

(b) Original situation: Train B is delayed, thus the meeting at StnY is not possible as originally planned.

(c) 'Conflict' is dragged to StnZ. The meeting is moved, giving priority to Train A.

Figure 13.1. Today, STEG only offers the functionality to simply extend a meeting. At the iron ore line, it is common to move meetings to different stations. The figures show, how conflict resolution of this kind can be simplified.

For the future, we are also open to including more ambitious automation and decision support that would help traffic controllers optimize their plans or evaluate different solutions. However, to be accepted by the traffic controllers, this automation has to be experienced as supporting or expanding their control. Possibilities means of achieving this are through strict adherence to given and adjustable rules (e.g., priorities on trains), well-timed presentation of solutions (giving time for adjustments), and presentation of decision-relevant variables.

The technical issues related to the automatic execution function were unfortunate. Ideally, it would have been highest priority from the beginning to correct these issues. I support the conclusion that the deployment has not been done in a user-centred way, meaning that the users and their needs were not sufficiently present in the process [44]. Based on the observed problems, I define my general expectations towards a user-centred deployment process as follows:

1. reveal the impact of the original systems' deficiencies in a different work situation
2. show the importance of technical problems in relation to the acceptance of new systems and the included concepts
3. ensure that additional features are developed in a user-centred way

13.3 Obstacles for Incorporation of our Concepts

The original plan was to evaluate our concepts for collaboration between train drivers and traffic controllers by evaluation of STEG and CATO. However, our visit to the iron ore line, after both systems had been used for a while, revealed a couple of problems. As discussed, the deployment of STEG in Boden did not meet our expectations, and even the deployment of CATO met with some challenges. This section takes a step back and fathoms the consequences for incorporation of the new concepts.

The introduction of STEG and CATO meant large changes to the work situation of train drivers and traffic controllers. Two roles that used to be quite distant do now have much more possibilities to collaborate, but at the same time their respective performance is more visible to each other than ever. Interestingly, this was never a concern in our discussions with train drivers and traffic controllers. In fact, when all concepts and systems are working as desired, we expect this to have a positive effect. If the performance of each is more observable, following GMOC, we expect the actors to develop better models of each other, which can replace the prejudices we have observed (see Chapter 12).

However, to be more concrete, such large changes need many prerequisites to be fulfilled, before the organization fully adapts. Positive was the availability of STEG at the complete traffic control centre and LKAB's push to equip all its trains with CATO. This means that a sufficient number of trains in a sufficiently large control area is equipped with systems, supporting a shared RTTP. Disadvantageous was the deployment process of STEG. Additionally, technical problems in both systems limited the users' trust in them.

At the control centre, the technical problems were the described issues with the automatic execution function of the RTTP and the less efficient STEG interface. This led to a halfhearted usage of STEG, resulting in insufficiently planned target points. Conversely, these insufficient target points irritated the train drivers, as CATO, following them, computed and displayed theoretically correct but practically incorrect advice. Combined with other peculiarities of CATO, train drivers started to pay decreasing attention to CATO. Even more counterproductive were some situations that actually rewarded not following CATO.

Take a meeting of two trains as an example. The traffic controller plans that train A has to stop, while train B should pass. If train driver A follows the plan, but the driver of train B decides to drive faster, train B might arrive before train A was able to stop, and in the worst case both trains have to stop. However, if both drivers drive faster than the plan, the meeting might happen as planned, but ahead of time. This means that both train drivers gain some amount of time at this meeting. Consequently, the traffic controller has to adjust the RTTP, which might lead to a situation in which both trains will have to stop (unnecessarily) at a later situation. However, this is delayed feedback and is harder to relate to the cause than the immediate positive feedback of saving some time. Hence train drivers did not care as much about CATO as it was desired they would. As a result, they did not adjust their driving strategies sufficiently.

13.4 Countermeasures

This leads to a vicious circle: train drivers care less about the advice as they get wrong advice or positive feedback on not following advice, while traffic controllers care less about accurate planning as the system is not supportive enough and they see that the drivers do not follow their plans anyway.

In the current situation, this circle can only be broken slowly. A positive fact is that, despite the mentioned problems, many involved operators still believe that the idea behind the concepts is good. Currently we are working together with LKAB, Trafikverket, and Transrail to take countermeasures. The first action is to fix the main technical problems; of highest priority is the automatic execution function. If this is reliable and used, the target points transmitted to the trains will be precise. Additionally, the interaction with STEG should be made more efficient to allow for the traffic controllers to concentrate on re-planning.

With the proposed changes, STEG will, it is hoped, be experienced as supportive, as was seen in Norrköping. Additionally, the target points received by the train drivers should be exact and finally allow them really to follow the RTTP. The next step, though, will be to regain the train drivers' trust. This will be easier if they see that the target points are correct and that following actually leads to benefits.

Another problem that we have observed is an insufficient understanding of target points. Train drivers as well as traffic controllers only have inadequate mental models. In particular, the reasons for un-reachability are hard to observe in both interfaces (STEG and CATO). Improvements in visualization and clear feedback to the train driver about the consequences of not following speed advice are necessary.

Besides proper support from the systems, one key component in the process of eventually incorporating our concepts and correct usage of the systems has to be continuous instruction and discussion about the systems, concepts, and their usage. Additional on-site visits and combined tests, where we accompany train drivers and traffic controllers simultaneously, seem to be a promising approach.

14. Creating Driver Advisory Systems for Additional Traffic Information

As our influence on the development of CATO was limited, we decided to start another project to develop concepts for a future information system for train drivers. Our motivation follows from the conclusion in Section 10.2.4. We want to develop a DAS that gives the train drivers additional information about the surrounding traffic situation. The goal is to allow the train drivers to plan their journey more optimal in a global perspective in terms of timeliness and economic driving. The train drivers' decisions should be guided by the traffic controllers' plan. By this, the traffic controllers' controllability of the traffic should increase. This chapter is based on Paper V but includes later results as well.

14.1 Train Driver Studies

This project is not finished, yet. We have conducted the first phase and have produced concepts and initial designs for a driver advisory system, mainly based on different types of passenger traffic. Figure 14.1 shows the plan for analysis and initial design. In the end, train drivers representing all different kinds of train traffic in Sweden should participate, including all kinds of passenger traffic (SJ), local commuter trains (Stockholmståg), and freight traffic (Green Cargo).

The plan starts with a period of interviews and observations to understand the train drivers' work and identify the properties of the different traffic types. Three to four workshops are then conducted in separate groups to deepen the gained understandings. Afterwards the analysis is finalized by a joint work group, including train drivers from all traffic types, aiming to create the foundation for a design prototype. This group should meet for another three to four workshops, following the vision seminar process.

So far, we have conducted studies and workshops with train drivers from SJ. During the first phase, we joined four train drivers from SJ on different regional trains, following four complete working shifts. We completed nine trips, lasting around two hours each, with train drivers from SJ. Usually a trip started at an SJ office, where the train drivers receive their schedule, and ended at another office, where we had time for interviews. The interviews were semi-structured, starting with general, open questions, asking the train

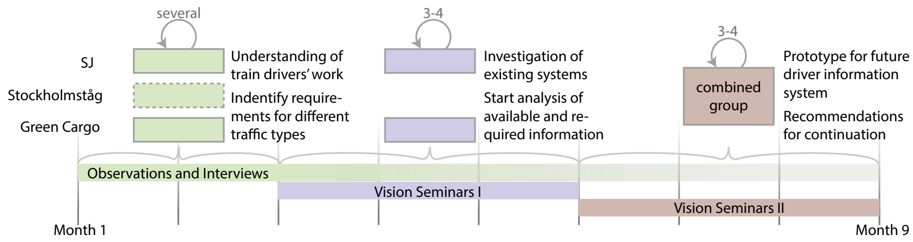


Figure 14.1. Plan for analysis and initial design of a driver advisory system.

drivers to describe their work. Later, the interviews were, if necessary, guided towards more concrete topics related to the traffic process, GMOC elements, and questions that arose during the observations. We recorded audio and took notes at the driver's cab and during the interviews. Afterwards we listened through the recordings and noted and transcribed parts that seemed relevant. These protocols were then analysed and coded by relating specific statements to GMOC elements.

In the same way, we followed four train drivers on the iron ore line in northern Sweden. They were operating iron ore trains for LKAB (the largest Swedish mining corporation) but were employed by Green Cargo. We attended five trips, of which one working shift (including two trips; from Kiruna to Björkliden and return) and two interviews were directly related to this project, while the other studies were related to the development of STEG and CATO, discussed in the previous chapter. The interviews lasted around two hours each. One interview was conducted together with one experienced (over thirty years) and one inexperienced (less than five years) train driver on the iron ore line, the second with a train driver from Green Cargo, operating common freight trains. Additionally, we were fortunate enough to discuss advisory systems with a larger group of iron ore train drivers, in a slightly heated atmosphere before a union meeting.

Based on the collected data, we conducted four monthly half-day workshops with train drivers from SJ in Stockholm. As the drivers of the iron ore trains are based in Kiruna, at a distance of approximately 1.000 km from Stockholm, they could not attend these workshops. The attending train drivers from SJ belong to an expert group, including two drivers, one experienced, one less experienced, for each type of train operated by SJ. At the final workshop, we presented and discussed a sketch for a DAS. After the workshop, we had one more meeting with one of the train drivers, who was involved in an internal SJ project to develop the DAS 'TrAppen'. We discussed the results and drew conclusions for further development.

14.2 Main Elements of the New DAS

Both our project partners, SJ and LKAB, had eco-driving as a main motivation for developing a DAS. A central element, though, is the speed advice or motion profile, which is very obvious in CATO (see Figure 14.2). As a preliminary result from the described studies and workshops, this section contains the themes that are interesting in the design of a future DAS going beyond simple speed advice. Even though the development of CATO and the work process described in the previous section were separated, the conclusions from the development of CATO were influential for this project. After all, both projects implemented the same concept – sharing of a plan – but with different motivations.

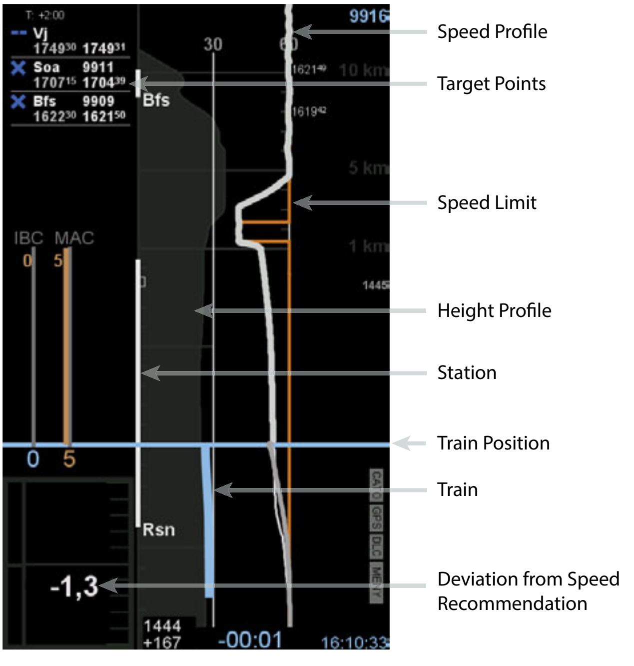


Figure 14.2. Screenshot from CATO describing some main interface elements.

14.2.1 Target Points to Display Changes in the Plan

Even though the motion profile is the central element in many DAS, it is the information lying behind this advice that is even more important. Part of this are the target points. As discussed in Section 12.3.2, they are the main instrument in STEG to share the traffic controllers' plan with the train drivers. As the traffic controllers have the best overview of the traffic situation, they can optimize the plan from a global perspective. Therefore the drivers' main goal should be to optimize his or her journey within the boundaries of the target points.

To work from a purely technical perspective, these points should at least contain information about time, position, and speed (i.e., where to be, when, at what speed). However, to be easy to understand and accepted by the train drivers, more information should be provided. If these points are the only way to inform the drivers about the surrounding traffic situation, they should at least also inform about the reason for the target point, for example, a meeting with train number 9909 (see the third target point in Figure 14.2). Train drivers can identify trains by their numbers. They can see which trains are iron ore trains, loaded, and equipped with CATO. With such information, train drivers are able to understand a large number of situations and motivations for target points. For example, loaded trains should not have to stop, as it is costly in terms of time, energy, and wear to stop and re-start the heavy, loaded iron ore trains. Consequently, these trains have priority at meetings.

When using and displaying target points, it is important to indicate if they can be complied with. There are two reasons for infeasible target points. One is an inappropriate plan. This should, in the ordinary case, only occur for short periods, when the controller is re-planning. The other reason is train drivers who do not follow the plan. This can happen if technical issues cause them to, or if they drive so slowly that they will be unable to catch up with the plan, or if they drive so quickly that they can only reach too early.

In theory, the reason for infeasible target points should be relatively clear – train drivers should know if they have followed the advice from their DAS. However, our experience from CATO shows that this, with the target points as displayed in CATO, is not always the case. The explanation is not straightforward, as many factors influence the acceptance of CATO's recommendation and the advice received from STEG. Section 13.2 has discussed some of the problems we saw during the deployment of CATO.

So, if train drivers cause infeasibility of a target point by not following it, there is either something wrong with the target point or the train drivers do not care about it. If we rule out technical issues, the question is rather how to make the target point visually clear and worthwhile to follow.

Still, we can conclude that displaying target points is a feasible way to inform the drivers about changes in the plan. However, to be really effective, this display should be easy to understand, and the target points should be clearly motivated by the traffic situation. The target points have to be reliable and reachable, and information on the context has to be correct.

When a traffic controller has to make many changes in the traffic plan to resolve a conflict, target points will change frequently. A fear was that these might disturb the train drivers. Instead, the train drivers in our studies were more irritated by last-minute changes and target points that were marked as unreachable for a long time or without any clear reason.

14.2.2 Motion Profile

The motion profile or speed advice is often used to support eco-driving. We prefer a two-tier system. The first tier would comprise advice that is visually easy to perceive. This should be the main advice and require a very limited amount of cognitive resources to attend to, in order to keep distraction at a minimum. The second tier is a motion profile for an adjustable range ahead of the train, giving the drivers the possibility to plan ahead.

One important prerequisite for this advice is that it must be reasonable: target points should be kept and speed limits must not be violated, but at the same time the average speed must not be too low, otherwise this increases wear on material and irritates the train drivers. Additionally, it should be understandable to be accepted by the drivers. We see an integrated view of motion profile, height profile, target points, and speed limits as a good solution to support the speed advice: speed limits mark the hard boundaries for the motion profile, target points set the range for optimization, and the height profile bears potential for energy saving.

Our observations from CATO have shown that inexperienced train drivers appreciated the displayed advice and found it helpful. Experienced drivers, in contrast, seemed to follow it naturally when trying to apply an energy-efficient driving style and sometimes found it disturbing. These effects were even stronger when a recommendation for train controls, in the form of requests for traction or brake power, were added: novices liked them, as they understood how the speed curve could be followed; experienced drivers regarded them both as impossible to follow and as questioning their expertise. Viewing this kind of recommendation should be optional.

14.2.3 Observability of the Traffic Situation

During almost all observations of train drivers, we experienced situations in which they constructed hypotheses about the surrounding traffic. They wondered if another train was delayed, when they did not meet it where they usually did, or when they had to stop at a signal where they usually do not. They wondered how to adjust their driving style to avoid stops or disturbing other trains when they were delayed themselves. Some drivers checked information available from Trafikverket for the current delay and rough position of another train. This information was better than nothing, but similar and even less expressive, meaning with lower precision and a certain delay, than the information traffic controllers can observe in their traditional systems.

Because train drivers are so interested in the surrounding traffic, we also developed prototypes that include a track diagram (see Figure 14.3) or even a time–distance graph (see Figure 14.4). The latter should represent a DAS that shows an RTTP similar to STEG. This view should be optional for experienced drivers who do not need or want to see a motion profile. Unfortunately, we

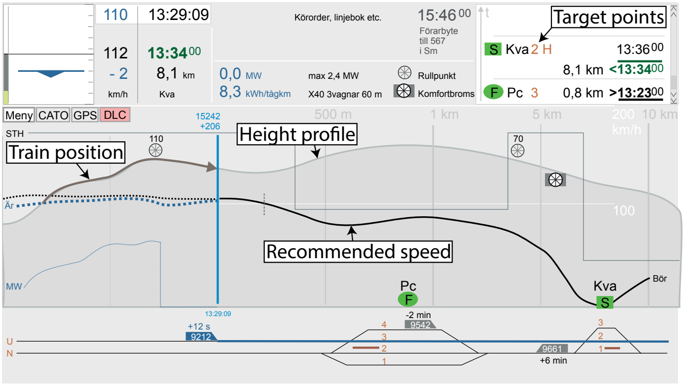


Figure 14.3. Extended prototype, based on our design for CATO and collaboration with SJ.

have not been able to discuss these prototypes with train drivers in detail yet. Compared to CATO, these DAS contain more details and are visually harder to attend to. Further research is necessary to find a more feasible means of visualization.

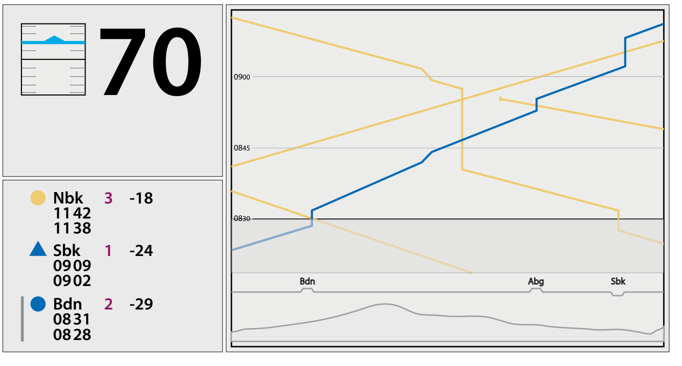


Figure 14.4. Very early DAS prototype including a time–distance graph of the traffic plan.

14.2.4 Improved Communication

From RouteLint, we know that (C-)DAS can lead to improved communication between train drivers and traffic controllers [4]. In our own studies, both groups claimed that they wish for better communication, meaning, for the drivers, that controllers reply more quickly to their calls and give more detailed information about the surrounding situation, and for controllers, that they can avoid unnecessary calls, such as questions about red signals or schedules of other trains.

In general, a DAS should provide information that makes communication more efficient. In CATO, train drivers can see their exact position, and this helps them and the traffic controllers to locate problems along the line. Additional information about the surrounding traffic can help to reduce the need for communication, too, for example, when this information explains reasons for unplanned stops.

We would even like to go further and introduce automatic transmission of information from a train to the traffic control system, and vice versa. Already today, CATO transmits the GPS position, reachability of target points, and shortest possible runtimes to STEG. Additionally, information like maximal possible acceleration could be transmitted. On the other side, a C-DAS could allow for sending a couple of standard messages between train drivers and traffic controllers. Some train drivers told us that they would like to inform the traffic controllers manually if they believe that a target point is planned too optimistically or pessimistically. Other standard messages could be to report expected delays in departure, for example, when the train driver expects unusually many passengers entering or leaving at a certain station; information about technical problems, for example, loss or recovery of a power car; or just a request for a call.

14.3 Discussion

As we have seen in most of our studies, train drivers are often interested in the surrounding traffic, and they have formed extensive mental models or knowledge around this. They are even aware of regular freight trains and can explain their special purposes. We observed this during sessions with the train drivers at the iron ore line in northern Sweden as well as with the drivers of passenger trains in southern and middle Sweden.

We see target points as key to informing drivers about changes in the traffic plan. In general, drivers have a positive attitude towards DAS and target points. Despite this, experiences with CATO have shown that the advice given by the system is not commonly followed. We see the main causes for this disregard in the problems discussed in Section 13.2.

We argue for a speed recommendation that optimizes energy consumption under good conditions. On one hand, it is impossible to predict exactly all (local) influences on a train; on the other hand, it is more difficult for the drivers to understand the advice if the calculation behind it is too complicated. Instead, drivers should be able to trust the advice under optimal conditions, motivating them to try to follow the advice more exactly from time to time to see if there exist alternatives to their driving style that might improve energy efficiency.

A point that we have not yet sufficiently investigated is the workload generated by systems as presented in Figures 14.3 and 14.4. There exists work concluding that simple speed advice in numerical form has less influence on workload than advice in the form of a timetable [60]. It would be interesting to study the influence of a DAS as presented in this chapter on workload and relevant aspects of situation awareness. We have observed several situations in which train drivers paid attention to tasks other than controlling the train, observing the environment through the windshield, and observing the train's instruments. So maybe the question is *when* it is safe to spend cognitive resources on a DAS rather than *how much* resources are spent.

Summary

This part combined the first research question of this thesis – investigating the utility of GMOC in describing human work in complex, dynamic scenarios – with the second research question: the development of systems and concepts for future train traffic control. To do so, we utilized the GMOC model in several studies related to train traffic control.

Key points of this part are as follows:

- GMOC supported us in analysing the complex organizational structure of train traffic operation.
- We have successfully applied methods in user-centred design processes built around GMOC in different projects regarding train traffic control and train driving.
- GMOC helped us to identify shortcomings in the human operators' work environment.
- The interface elements in STEG can be motivated with GMOC.
- A user-centred deployment process is necessary as well.
- Train traffic control in Sweden differs from control in other countries. It is highly centralized, and only one role is responsible for dispatching and signalling/maintaining and executing the traffic plan.
- New systems can improve train traffic control as they allow the operators to concentrate on planning ahead instead of handling exceptions.
- For next-generation train systems, a real-time traffic plan should be implemented and available to each actor. This will have a large impact on the traffic controllers' controllability of the traffic and the train drivers' awareness of the overall traffic situation.
- Driver advisory systems are often designed from a certain perspective.
- Connected DAS are needed to increase railway capacity, controllability (from the traffic controllers' perspective), and observability (from the train drivers' perspective).
- The introduction of new concepts on this scale is slow and prone to errors. Close collaboration with the users, an elaborated deployment process, continuous training, and endurance will be necessary.

Part IV: Conclusions

The final part of this thesis summarizes the findings of the presented work and discusses them in relation to the research goals described in the introduction. First, the overall results from our studies in the area of train traffic control are evaluated. This evaluation shows the significance of our results. It thus is the basis for discussion of the role that the GMOC model played in our studies. Finally, main conclusions, contributions towards the research objectives, and possibilities for future work are discussed.

15. Evaluation of Concepts and Systems for Future Train Traffic Control

This chapter evaluates the central concepts developed during the studies described in Chapters 11–14. The goal is to see how promising the developed concepts and systems for future train traffic control are:

- Control by awareness
- Automatic plan execution/non-autonomous automation
- The real-time traffic plan (RTTP)
- Sharing of the RTTP with the train drivers via target points

These concepts are implemented in the traffic control system STEG and the driver advisory system CATO. Therefore the evaluation of the concepts overlaps with the evaluation of STEG and CATO.

15.1 Control by Awareness

The first evaluation of the concept control by awareness was done in a laboratory environment using a functional prototype of STEG [54]. The results were quite positive. Traffic controllers seemed to become familiar with the new control system quickly and did like the new control strategy. A concern was that a real-world scenario would be much more complex; therefore further evaluation was necessary.

This has been done using an operative, deployed version of STEG [45, 83]. These tests showed acceptance and efficiency of the new control strategy but also revealed its sensitivity to practical and technical problems [83]. Still, the traffic controllers using STEG were very optimistic. They felt more in control and able to plan more accurately. Main reasons 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. They were supported in maintaining a plan reaching one to two hours ahead and had to make fewer last-minute changes [45] – the last two points are exactly the goals of control by awareness. We thus can conclude that STEG is successful in supporting control by awareness. Acceptance and rapid adoption of the strategy as well as the positive comments from the traffic controllers let us conclude that STEG and control by awareness improve the traffic controllers' work environment.

15.2 Automatic Plan Execution and Non-autonomous Automation

Control by awareness and the automatic plan execution function are closely related. Thus this kind of non-autonomous automation (compare Chapter 6) provided by the automatic plan execution function was another aspect of the evaluations discussed earlier [54, 45, 83]. The two main advantages of this kind of automation are the relief that train traffic controllers do not have to remember and execute their control decisions regarding the underlying traffic control system and that this kind of automation follows the operators' plan; that is, it acts in line with the plan without changing it.

Traffic controllers using the traditional systems commonly report automation surprises and the urge to turn off automation [53, 9]. During the evaluation of STEG, we did not observe this behaviour when using the automatic plan execution function. The plan execution function instead was accepted by the train traffic controllers and reduced their urge to take over manual control [45].

Supported by our observations during the deployment of STEG in Boden, we can confirm that the automatic plan execution function is critical for the adoption of control by awareness. We can further conclude that this kind of non-autonomous automation is judged as being more helpful by the traffic controllers than traditional systems.

15.3 The Real-Time Traffic Plan – RTTP

The RTTP is a central concept in the new principles for train traffic control. There exists only one common RTTP, which is the valid plan for all included trains. It is the train traffic controllers' responsibility to maintain it by updating it whenever required, for example, when disruptions cause conflicts that must be resolved. Existence of a common RTTP gives the opportunity to make it known and available to all different actors, such as other traffic controllers, railway companies, and train drivers. It is important to understand that the planning activities, documented in the RTTP, are not just done with an eye towards each individual controller's goal but for contributing to better plans for the traffic system as a whole.

The RTTP is also a necessary prerequisite for the automatic plan execution function. It is the RTTP that is executed, exactly as it is specified, when it comes close to the present time. In STEG, the RTTP exists as a data model behind the visualization in the interface. It is these data that are changed by interactions with the time–distance graph in STEG. These data are stored in the traffic control system and can be accessed by the automatic execution function to identify commands that in time have to be sent to the infrastructure, that is, the signalling system. In multiuser systems, that is, when several STEG clients are used to control a complete traffic area, the data are refreshed between the traffic control system and the different clients. Each client is only allowed to change the plan for a distinct region, even if the plan as such covers the whole traffic system. If the exit of a train from one region is delayed, the adjacent traffic controller can observe this and take appropriate measures.

15.4 Sharing of the RTTP

The new aspect of the RTTP presented in this work is its utilization for communication between train drivers and traffic controllers. Target points are used to transfer the traffic plan to the train driver. Our hypotheses in relation to sharing of the traffic plan are as follows:

1. Traffic controllers achieve better controllability of the traffic process.
2. Train drivers gain observability of the surrounding traffic.
3. The competence of the train drivers is supported in a way that does enhance – not reduce – their driving skills.
4. The number of unnecessary stops is decreased.
5. Train drivers are able to plan their ride more efficiently and thus save energy while achieving punctuality.
6. Eventually, the capacity on the line increases.

The first three points can only be evaluated qualitatively, for instance, with interviews, whereas it would be possible to evaluate the following three points quantitatively.

In practice, the evaluation turned out to be complicated. Even though the sharing of the RTTP is implemented, proving its technical feasibility, several factors such as insufficiencies in the deployment process and technical as well as organizational difficulties have led to a reluctant attitude towards the new systems (see Chapter 13). In consequence, instead of being able to conduct an evaluation, the challenges were more basic. We identified the insufficiencies in the deployment process, the system usage, and the systems themselves and proposed actions opposing these insufficiencies. As a result, we could only evaluate the new concept of communicating the real-time traffic plan in a controlled test run, leading to new action points, ideas, and hypotheses about ways to improve acceptance and efficiency of the solutions.

There is some evidence that supports our concepts, such as a generally positive attitude during the last controlled test run and measurements done by Transrail [99]. The presented measurements cover energy consumption and the average punctuality of trains equipped with CATO, and the re-planning actions of a train traffic controller. According to these data, the average punctuality increased by around ten percentage points. However, relevance of this value is questionable, as punctuality is influenced by many factors, such as weather conditions, construction work, and the properties (e.g., robustness) of the traffic plan. One would have to compare traffic under similar conditions to generate meaningful data. This is the main reason why we did not evaluate our concepts based on such criteria. However, the other two values are much more interesting. These are taken from selected train runs with drivers following the advice and thus the target points shown by CATO. The results are energy savings of around 15 to 25 per cent and drastically reduced re-planning events, that is, changes in the traffic plan done by the controller. The latter is especially interesting: traffic controllers have mentioned that more than half of their re-planning actions are caused by train drivers not following their traffic plans. The measurements confirm this statement. We can thus conclude that a shared traffic plan can lead to a clear reduction in required re-planning.

15.5 Discussion

The central observation during the deployment of STEG in Boden was the hesitant adoption of the system's new control strategy, control by awareness. As this is one of the basic concepts behind STEG and important for improving train traffic control, we investigated this point further. As described in Chapter 13, we found that implementation of the automatic plan execution in Boden had shown to be much more difficult than in Norrköpping. It did not provide the needed functionality and was not working reliably. Therefore STEG was often used with this function disabled. This finding confirms our assumption that the automatic execution function is a key concept for adoption of control by awareness as a strategy.

The main flaw in STEG itself is its lack of efficiency in certain kinds of re-planning activities. This flaw was discovered already during evaluation of the STEG prototype in Norrköping and shown to be problematic in accommodating last-minute changes [45]. Still, it had smaller impact on the traffic controllers' work, and STEG was evaluated very positively by its users despite this problem. The different traffic situation in northern Sweden, however, increased the impact of this flaw on the traffic controllers' daily work in Boden. This, combined with the described problems of the plan execution function, made use of STEG clearly less efficiently than in Norrköping. With a correctly working automatic plan execution function that supports the special needs of the operators in Boden, additional improvements in the efficiency of the interaction in STEG, and additional training, I would expect a much better result in the deployment of STEG in Boden.

Another interpretation of the studies of STEG in Boden is that automation that does change the traffic plan should not be excluded categorically. There are at least two reasons to still consider it. First, it can be used to simplify interaction with STEG; second, it can optimize the plan according to boundaries given by the traffic controller. The first option would basically reduce the number of manual actions needed to apply changes in the traffic plan. According to the model proposed by Parasuraman et al. [72], this would, as well as the automatic plan execution, be automation of action implementation. This means it would not interfere with the traffic controllers' decision processes. The second option could interfere. Therefore further research is needed to figure out how such automation can be integrated into the operators' decision processes in the most supporting way. This option could, for instance, be implemented in a semi-automatic way as suggestions; this means that the traffic controllers could request and then accept or refuse an optimization.

15.6 Steps for Further Evaluation

Since the concept of the shared RTTP is still not fully adopted on the iron ore line, further evaluation is necessary. Owing to the discussed problems, quantitative assessment of generic values such as punctuality is error-prone. Therefore quantitative evaluation seems only suitable for rather specific measurements. CATO allows measurement of a couple of values. One example is the number of recalculations of its speed profile. A low value is desired, while a high value indicates a sub-optimal ride. Using CATO, we can relate the recalculation either to re-planning by the traffic controller or to the train driver not following the speed recommendation. Combined with further measurements, we will be able to draw conclusions regarding, for example, technical reliability, attitudes towards the systems, driving styles, or re-planning strategies.

In any case, additional qualitative evaluation is necessary. Naturally, we recommend an assessment of the work environment according to the GMOC model. Did the expected improvements occur? Did the traffic controllers adapt the new control strategy? Do they see an improvement in the traffic flow (qualitative evaluation of values that are hard to assess quantitatively). Additionally, the qualitative measurements can be important to support or falsify hypotheses based on quantitative measurements, such as the one just discussed.

Another interesting point would be to evaluate designs for new driver advisory systems, as discussed in Chapter 14. It would be very interesting to connect prototypes of such systems to a train driver simulator and to assess certain effects of the driver advisory system, for example, effects on the driving style, performance, or mental workload of the train driver. Such a study would allow us to relate our recommendations to the simplistic systems evaluated in Large et al. [60].

16. The Development Process with GMOC

In Table 3.1, we have defined how we applied action research. Following this definition, the results presented in the previous chapter illustrate our problem-solving interest. Our central research interest is the exploration of GMOC as a model to support analysis of human control of complex, dynamic environments and development of future support systems.

We have taken five steps to pursue this research objective. The first step was to define GMOC. The definition can be found in Chapter 5. We then related the model to theories and concepts that are relevant in the area of human control of complex, dynamic processes (see Chapter 2). The third step was to provide a method to apply GMOC in system development. Chapter 9 presented a generic approach to applying the model in user-centred system design. Part III of this thesis presented several studies covering different stages of the development process. This was the fourth step. Following the action research process, GMOC was used and refined during all these studies. Similarly, we reflected on the used methods. This means that the methods presented in Chapters 9 and 8 and the description of GMOC in Chapter 5 reflect the latest iterations of model and methods. The fifth and last step in assessing the fulfilment of the research objective is to show the importance and value of the results and the efficiency of the process in a way that allows comparison of our model with other approaches.

16.1 Analysis of Complex Environments

The organization of train traffic operation is complex, and train traffic control itself is a complex, dynamic process. The combination of ethnographic methods guided by the elements of GMOC supported us in the analysis of this process. It allowed us to gain an overview of the process (see Figure 10.1.1) and to analyse the implications of the process for the roles of train drivers and traffic controllers. This gave additional depth to the analysis of their specific work environments. We were able to identify problems and the potential for improvement in the traffic controllers' work environment (see Chapter 11), in the train drivers' work environment (see Chapter 14), and in the collaboration between these two roles (see Chapter 12).

In particular, it was helpful to analyse the organization's influence on the operators' work environment according to the GMOC elements. We identified that the different layers in a socio-technical system relate to different influences on goals and mental models and on observability and controllability (see Figure 5.2). Applied to train traffic control, this meant supporting the means for collaboration, especially with regard to the efficient exchange of traffic plans and information.

The main purpose of the analysis utilizing GMOC lies in its support for understanding user–work interaction, meaning the intentions users have when interacting with their environments and the means they have to do so. Important here is the fact that we actually have a picture of all relevant aspects. In this thesis, we related GMOC to the most relevant aspects and concepts in human control, among these visualization, automation, control theory, dynamic decision making, and situation awareness. We have mentioned other research that builds on approaches or models similar to GMOC (e.g., [11, 66, 67, 58]). We can indeed see GMOC as a common denominator of these concepts and approaches. This supports our claim that we cover the relevant aspects of human control of complex, dynamic systems.

16.2 GMOC in User-Centred System Design

Development of new systems for train traffic control is a long process in which design decisions affect each other and therefore every detail has to be implemented and evaluated with care. GMOC has proven to be useful in understanding and organizing this complexity. Its utilization in analysis reveals problems in today's work environment and routines. During the design, these insights are valuable in creating new work environments and routines. GMOC motivates the design decisions and can reveal possible interference between new routines and the existing systems and organization. We can formulate hypotheses about how changes in the work environment will affect the operators' work and its outcome.

A central point is to close the gap between user and system. Our approach to system design allows the users to develop adequate mental models. In Chapter 5, we mentioned ways to achieve this. One important aspect is the relation between controllability and observability, or in other words, the feedback that the system produces in reaction to control actions. When analysing human work with GMOC, we can see which information the system produces, how the users interpret it, and how they respond. Here we can identify missing information or feedback that could make the users' control more effective or efficient. It is a key principle in visualization following GMOC to show all relevant information. The complexity of the process should be visualized, not hidden. Guidelines for efficient visualization are discussed in Paper II.

16.3 GMOC and Usability

At the beginning of this thesis, we defined the development of systems with improved usability as a central goal of HCI. Usability is defined by purpose, learnability, memorability, effectiveness, efficiency, and satisfaction. Chapter 11 presented the novel traffic control system STEG and analysed its user interface with regard to goals, models, observability, and controllability. The point to discuss here is the relation between the elements of GMOC and usability and how usability was established in STEG.

Section 10.3 has given a more concrete description of the criteria for more optimal train traffic control. The purpose of STEG is to support the human operators in maintaining more optimal control according to these criteria. This has been evaluated in Chapters 11 and 15.

Effectiveness is given, as traffic controllers using STEG affirmed that it allowed them to perform more optimal and precise planning. A main factor here was the improved observability. As STEG provided more exact information about the current position of a train and allowed a more exact projection of the train's future position, planning could be optimized. At the same time, it became more obvious how a train (driver) behaved. This allowed the traffic controllers to develop more adequate mental models (e.g., how does driver X usually behave, or how do timetable, track profile, and weather conditions usually affect the traffic flow at position Y) and to gain better situation awareness (e.g., the track at position X seems to be slippery today, or train Y seems to have technical problems).

Efficiency is given, as the train traffic controllers confirmed a reduced mental workload achieved by more efficient re-planning (excepting last-minute changes). Support of the automatic plan execution is obvious here, as it reduces the mental capacity needed for recalling control decisions at certain times and the effort required for manual plan execution. These are mainly aspects that improved controllability. Further improvement of controllability can be achieved by the concepts presented in Chapter 12. As we assume that adoption of these concepts will reduce the need for re-planning, further positive effects on efficiency are expected. A side effect of the reduction of required mental activity for plan execution is the availability of mental resources for optimization of the traffic plan, which has positive effects on effectiveness as well. It is not an objective 'per se' to make the work of the traffic controllers' easier, but to allow them to use their cognitive capacity and skills to improve the quality of their work.

We see the combination of effectiveness and efficiency as an important factor creating satisfaction. STEG clearly satisfied its users in Norrköping, who wanted to keep the prototype system in operation. They particularly liked that STEG allowed them to give much more exact estimations and information to the train drivers.

Learnability and memorability are effects that are related to users' ability to develop a mental model of the system. This is a central point in GMOC and has been discussed before. In the context of our studies, it is meaningful to distinguish between learnability of the task itself and of the system. The holistic view of GMOC covers both, but for assessment of STEG's usability, it is more meaningful to concentrate on its learnability. Our studies in Norrköping have shown that traffic controllers became familiar with STEG relatively quickly, and we did not observe problems in use in the long run. However, because we are dealing with a system for a complex control task that even introduces a new control strategy, a certain amount of training is required. Part of the problems we observed in Boden can be related to a lack of training.

We can conclude that utilization of GMOC in the development process supports usability in the final system. It seems that certain elements can be better related to certain usability criteria than others. However, a strong distinction does not seem to be particularly useful, as in the view of GMOC, the interplay between the elements is important as well. We cannot create a usable system if it has severe shortcomings in one of the elements.

16.4 Relation between STEG Deployment Problems in Boden and GMOC

Many of the positive results mentioned earlier emerged from our studies of STEG in Norrköping. Our studies in Boden partially show a different picture. Here the deployment of STEG was not as successful. As described in Chapters 13 and 15, we see that the main reason is a sub-optimal deployment process where users' needs and different conditions in the control environment have not been sufficiently considered. We therefore see the observed problems in adoption of the concepts incorporated in STEG not as a problem of STEG and its GMOC-based development process. Instead, the main issues have been technical. An example is the error-prone automatic execution function. Its difficulties in implementation derive from the different signalling system in Boden, which has turned out to be harder to connect to STEG.

In fact, we can see the experiences related to the deployment in Boden as a support for GMOC. These experiences show that it is indeed important that the control situation be seen as a whole. If one part is missing, effectiveness of a new system can be reduced considerably. Analogously, a lack in one GMOC element, controllability in this example, can have strong effects on the other elements. In this way, GMOC even helped us analyse and understand the problems that arose.

16.5 Discussion

GMOC and action research seem to share many properties. Both can support the study of complex environments where a strict methodology might not be adequate. At the same time, the lack of clear methods make them harder to apply and the results harder to generalize. Still, we were able to show that GMOC is applicable and valuable in the analysis and design of human control in complex, dynamic environments. We were also able to derive general principles for train traffic control and train operation, visualization of data in complex processes, and utilization of GMOC in user-centred system design. However, we were not able to perform a direct comparison between GMOC-based user-centred design and other methods, because our study was of large-scale development projects which are impossible to repeat using different methods. Therefore we have chosen to evaluate the development process based on the significance of the results.

GMOC was used from the beginning in the analysis of train traffic control and the work of traffic controllers and train drivers. It helped us to understand their work as well as to identify problems and limitations in their information and control systems. It helped in analysing the complex organization of train traffic control and revealed that collaboration could expand the operators' observability and controllability. This observation was an important argument for implementing and sharing a real-time traffic plan (see Sections 12.3.1 and 12.3.2 and Paper III).

When the new control principles and systems and the STEG user interface were specified, designed, and developed, this was also based on the GMOC model. The model, combined with the user-centred and ethnographic methods, helped us identify core requirements that otherwise would have been omitted or disregarded. We can thus conclude that utilization of GMOC in user-centred system design is quite useful.

17. Final Remarks

This chapter summarizes the results regarding system design with GMOC and the development of new concepts for train traffic control and train driving. It relates my personal contribution towards achievement of my research objectives and proposes directions for future work.

17.1 GMOC

This thesis presented GMOC, a model adopted from control theory and dynamic decision making. It describes human–work interaction, as it contains the four necessary elements for efficient human control of complex, dynamic systems: goals, (mental) models, observability, and controllability. As these elements are the four central prerequisites for efficient control, the model supports analysis and development of control environments (including systems, tasks, and organizations). GMOC structures analysis and evaluation as it supports identification of the four elements and of new environments, as it helps to identify problems and fields for improvement according to the four elements. As the four elements represent properties of the human operators (goals and models) and their interactions with the system (observability and controllability), GMOC bridges the gap between the human operator and the technical system. In its applicability during all phases of user-centred system design, it also bridges analysis, design, and development.

A main contribution of this thesis is its definition of the four elements and their interplay. Additionally, the model is placed in the context of HCI and system development. A method for using GMOC in user-centred system design is presented to make the model more accessible to other researchers and system developers. The method derives from our application of GMOC in a couple studies related to train traffic control. These studies have been presented and show that it is possible to analyse very complex work organizations and environments using GMOC. The analyses have led to designs that have been implemented and partly evaluated in productive use. The results are positive and have been described and discussed. Suggestions for visualization in complex, dynamic environments based on our experience with the application of GMOC are given in Paper II.

There is no strictly defined method based on GMOC, with a very clear description how to transfer information from one phase of the development process to the next. Instead, the presented method has evolved implicitly from our studies. However, as every project has its own context, possibilities, and limitations, it will be possible to apply different methods. A suggestion is to see GMOC in the first place as a ‘language’ describing human–work interaction. During system development, one has to select appropriate methods to fill this description with content and to transfer the description into concrete designs and systems.

17.2 Train Traffic Control

The Swedish railway systems, like many others all over the world, are controlled based on principles from past decades. Even though the control systems themselves have often been updated during this time, the characteristics of the work done by traffic controllers have not changed significantly. The technological development of rolling stock during the past decades is huge. It is impressive to see the progress in terms of comfort and technology between the old generation of iron ore locomotives (Dm3) and the new generation (IORE). However, when it comes to traffic information and collaboration with train traffic control, the work of the train driver has not developed at a similar pace.

The studies done during the work on this thesis have shown that this traditional operation of train traffic control and train driving has significant drawbacks. Introduction of new work principles and changes to the organization have great potential for improvement and increased effectiveness in terms of, for example, energy consumption and capacity utilization. The projects and studies that we have performed demonstrate that it is technologically possible to realize this potential.

A key factor in increasing effectiveness, and work satisfaction at the same time, is improving collaboration between the different actors in train traffic operation. This thesis focused on the collaboration between train drivers and traffic controllers. The main steps towards improve collaboration are making communication more efficient and increasing the amount of shared information. Effective communication will lead to higher satisfaction, and better depth of information will give all actors better opportunities to improve their own plans, avoid sub-optimization, and collaborate around a common plan. A key concept here is the shared RTTP, as presented in this thesis.

With this RTTP, we are able to develop information systems that allow the actors to share all relevant information. Train drivers will be made aware of changes in the traffic plan, which will allow them to adjust their driving strategies and to inform their passengers or customers (in case this is part of their responsibility). Traffic controllers will have a better picture of the situation and will be better able to give accurate estimations of arrival times and delays.

With appropriate support systems, actors can be more proactive. Train drivers can avoid unnecessary stops at red traffic signals. Traffic controllers could have better means to guide train drivers and thus control the traffic process. Supported by appropriate automation, such as the automatic plan execution function implemented in STEG, as discussed in this thesis, traffic controllers can shift their control strategies to control by awareness and thus utilize their cognitive capacity for planning ahead.

Several principles for automation have been discussed in this thesis. It is central that operators be able to understand the automation, meaning that they know what the automation will do, when, and why. This is easiest to achieve if automation is designed in a non-autonomous way, only following the operators' plans. If this is not feasible or efficient, designers and developers should ensure that automation can be directed so that it does not conflict with the operators' plans and that it is transparent so that the operators can understand and predict its actions. This means that a control system has to provide observability also regarding the embedded automation.

Another conclusion from our studies is that even small differences between traffic centres and work environments can lead to large problems during deployment of new systems. Even within one country, conditions vary considerably in terms of local organizations, (seasonal) weather and environmental conditions, different types of traffic and infrastructure, and even social differences. Some of these factors are represented by operators' local knowledge. This knowledge naturally exists in regional borders. We have shown possible ways to include such information in new systems. However, we have to assume that it is impossible to transfer all relevant local knowledge from the human operators into systems. This is an important factor to consider when developing new systems that will allow control from more distant locations than the current systems.

Finally, we have to emphasize that developing new systems in train traffic control and changing traditional work concepts and strategies is a long process. On one hand, developing new systems and concepts requires a long phase of analysis of the work situation; on the other hand, introducing the new systems takes time and should follow a user-centred model. End users need close guidance and a supporting organization to adopt new concepts and strategies.

17.3 Contribution

During my research, I described and extended the GMOC model. In Paper IV, I applied it to organizational scenarios. This extension was necessary during the related study (Chapter 12) but can only be seen as a first step towards an 'organizational' GMOC. Furthermore, further exploration of the 'model' element was tempting, in particular, analysing the relation between mental models and intuitive behaviour or tacit knowledge. However, the first research objective of this thesis is to make GMOC applicable in the analysis and development of complex, dynamic systems and to assess its value. For this, I have described a user-centred design method based on GMOC and defined its four elements in detail. Further explanation and development of the elements are beyond the scope of this thesis. By describing and evaluating several studies using GMOC, I see this first research objective to be achieved.

The second research objective was to develop future concepts for improvement of train traffic operation. My main contribution lies in the analysis of collaboration between traffic controllers and train drivers as presented in Chapter 12. Owing to the problems described in Chapter 13, the concepts presented in Chapter 12 were not completely adopted by the deadline for my studies. This had impact on their evaluation (see Chapter 15). The study presented in Chapter 14 has produced concepts for DAS that go beyond those in existing systems. During the work on this thesis, new concepts for train traffic control and driver advice have been developed and implemented. Even though the time frame did not allow for a full evaluation of all concepts, I see the second objective as achieved.

17.4 Future Work

Our research has shown many possibilities for future projects in the area of train traffic control and train operation. The evaluation of the developed concepts has to be completed. Prerequisites and possibilities for such evaluations have been described in Section 15. The STEG interface and the RTTP hold great potential for future projects. An interesting topic would be an optimization of the user interface, for example, to better support goal prioritization and to allow for more efficient interaction. In general, the possibilities of including algorithms for optimization, feedback on the controllers' solutions, and further automation should be explored. A project to develop a STEG system for complex stations is already in progress.

The study of DAS presented in Chapter 14 is just the beginning of the development of novel DAS. The presented designs have to be evaluated and iterated. It would be especially interesting to explore if the display of additional traffic information helps train drivers to develop more appropriate mental models and situation awareness regarding the surrounding traffic, and if this would lead to a more optimal driving strategy. These effects should be compared to more simplistic DAS. The impact of the suggested visualization of the driving advice on mental workload should be investigated as well. A point that was not discussed in this thesis was the relation of the DAS to ETCS, the proposed European Train Control System. The reason for this lack is the relatively slow adaptation of the standard combined with the limited possibilities for including additional information. However, in the future, with further spread of ETCS, it will be necessary to suggest a design that can be integrated into such systems.

During the exploration and description of GMOC, several limitations became clear. First, a more explicit description of how to transfer a GMOC description into a design would be helpful. A more extensive analysis of its relation to more complex HCI methods, such as cognitive work analysis, is needed to further explore advantages and disadvantages of GMOC. Finally, its application to additional areas, such as industrial process control, based on the description and methods presented in this thesis, would be very interesting.

References

[1] Mark Ackerman. The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility. *Human-Computer Interaction*, 15(2):179–203, 2000.

[2] Thomas Albrecht. Human factor challenges in the development of a driver advisory system for regional passenger trains. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, pages 129–138. Taylor & Francis, 2013.

[3] Thomas Albrecht, Christian Gassel, Anne Binder, and Jelle van Luipen. Dealing with operational constraints in energy efficient driving. In *IET Conference on Railway Traction Systems (RTS 2010)*. IET, April 2010.

[4] Thomas Albrecht, Jelle van Luipen, Ingo A. Hansen, and A. Weeda. Bessere Echtzeitinformationen für Triebfahrzeugführer und Fahrdienstleister (in German). *EI - Eisenbahningenieur*, (58):73–79, 2007.

[5] Arne W. Andersson, Ingmar Frej, Anders Gideon, Peter Hellström, and Bengt Sandblad. A systems analysis approach to modelling train traffic control. In *WCRR*, pages 16–19, 1997.

[6] David E. Avison, Francis Lau, Michael D. Myers, and Peter Axel Nielsen. Action research. *Communications of the ACM*, 42(1):94–97, January 1999.

[7] Lisanne Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, November 1983.

[8] Lisanne Bainbridge. Mental models in cognitive skill: The example of industrial process operation. In *Models in the Mind: Theory, Perspective, and Applications*, chapter 9, pages 119–143. Academic Press, 1992.

[9] Nora Balfe, John R. Wilson, Sarah Sharples, and Theresa Clarke. Development of design principles for automated systems in transport control. *Ergonomics*, 55(1):37–54, January 2012.

[10] Jeanette Blomberg, Mark Burrell, and Greg Guest. An ethnographic approach to design. In *The HumanComputer Interaction Handbook*, pages 964–986. Lawrence Erlbaum Associates, Mahwah, New Jersey; London, 2003.

[11] Berndt Brehmer. Dynamic decision making: human control of complex systems. *Acta psychologica*, 81(3):211–41, December 1992.

[12] Berndt Brehmer. Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science*, 6(1):73–93, January 2005.

[13] Berndt Brehmer and Dietrich Dörner. 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*, 9:171–184, 1993.

[14] Mary Brydon-Miller, Davydd Greenwood, and Patricia Maguire. Why Action Research? *Action Research*, 1(1):9–28, July 2003.

[15] Graham Button. Studies of Work in Human-Computer Interaction. In *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, pages 357–380. Morgan Kaufmann, 2003.

[16] Rebecca Charles, Nora Balfe, John R. Wilson, Sarah Sharples, and M. Carey. Using graphical support tools to encourage active planning at stations. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, pages 427–432. Taylor & Francis, 2013.

[17] Francesco Corman and Lingyun Meng. A review of online dynamic models and algorithms for railway traffic control. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, pages 128–133. IEEE, August 2013.

[18] Nastaran Dadashi, David Golightly, and Sarah Sharples. Requirements elicitation for disruption management support. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, pages 399–408. Taylor & Francis, 2013.

[19] J. C. F. de Winter and D. Dodou. Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work*, 16(1):1–11, August 2011.

[20] Sidney W.A. Dekker and Erik Hollnagel. Human factors and folk models. *Cognition, Technology & Work*, 6(2):79–86, May 2004.

[21] Sydney W.A. Dekker. On the epistemology and ethics of communicating a Cartesian consciousness. *Safety science*, 56:96–99, July 2013.

[22] Sydney W.A. Dekker and David D. Woods. MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, 4(4):240–244, November 2002.

[23] Dietrich Dörner. *Die Logik des Misslingens*. Rowohlt, 1989.

[24] Ward Edwards. Dynamic decision theory and probabilistic information processing. *Human Factors*, 4:59–74, 1962.

[25] Mica R. Endsley. Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, 1988.

[26] Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, March 1995.

[27] Mica R. Endsley. Automation and situation awareness. *Automation and human performance: Theory and applications*, pages 163–181, 1996.

[28] Mica R. Endsley. Situation Models: An Avenue to the Modeling of Mental Models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(1):61–64, July 2000.

[29] Mica R. Endsley and Daniel J. Garland. *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates, 2000.

[30] Mikael Erlandsson and Anders Jansson. Verbal reports and domain-specific knowledge: a comparison between collegial and retrospective verbalisation. *Cognition, Technology & Work*, 15(3):239–254, March 2012.

[31] J. Fénix, T. Graffagnino, J. C. Sagot, and C. Valot. Train rescheduling based on predefined track-slots. Technical report, Swiss federal railways (SBB), 2006.

[32] John M. Flach. Situation awareness: Proceed with caution. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):149–157, 1995.

[33] Dedre Gentner and Albert L. Stevens. *Mental models*. Erlbaum, 1 edition, 1983.

[34] David Golightly, Bengt Sandblad, Nastaran Dadashi, Arne W. Andersson, Simon Tschirner, and Sarah Sharples. A socio-technical comparison of rail traffic control between GB and Sweden. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, pages 367–376. Taylor & Francis, London, 2013.

[35] David Golightly, John R. Wilson, Emma Lowe, and Sarah Sharples. The role of situation awareness for understanding signalling and control in rail operations. *Theoretical Issues in Ergonomics Science*, 11(1-2):84–98, January 2010.

[36] Cleotilde Gonzalez, Polina Vanyukov, and Michael K. Martin. The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, 21(2):273–286, March 2005.

[37] Jan Gulliksen, Bengt Göransson, Inger Boivie, Stefan Blomkvist, Jenny Persson, and Åsa Cajander. Key principles for user-centred systems design. *Behaviour & Information Technology*, 22(6):397–409, November 2003.

[38] Frank G. Halasz and Thomas P. Moran. Mental Models and Problem Solving in Using a Calculator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '83, pages 212–216, New York, NY, USA, 1983. ACM.

[39] Peter A. Hancock. Automation: how much is too much? *Ergonomics*, 57(3):449–454, March 2014.

[40] Niklas Hardenborg. *Designing work and IT systems: A participatory process that supports usability and sustainability*. Acta Universitatis Upsaliensis, Uppsala, 2007.

[41] Steve Harrison, Deborah Tatar, and Phoebe Sengers. The three paradigms of HCI. *Alt. Chi. Session at the SIGCHI ...*, pages 1–18, 2007.

[42] Gert Heister, Jörg Kuhnke, Carsten Lindstedt, Roswitha Pomp, Thorsten Schaer, Thomas Schill, Stephan Schmidt, Norbert Wagner, and Wolfgang Weber. *Eisenbahnbetriebstechnologie (in German)*. Bahn Fachverlag, Heidelberg; Mainz, 2006.

[43] Peter Hellström. Problems in the integration of timetabling and train traffic control. Technical report, Department of Information Technology, Uppsala University, Uppsala, 2014.

[44] Gunnika Isaksson-Lutteman. *Future Train Traffic Control: Development and deployment of new principles and systems in train traffic control*. Licentiate thesis, Uppsala University, 2012.

[45] Gunnika Isaksson-Lutteman, Arvid Kauppi, Arne W. Andersson, Bengt Sandblad, and Mikael Erlandsson. Operative tests of a new system for train traffic control. In *Rail Human Factors around the World: Impacts on and of People for Successful Rail Operations*, pages 424–433. Taylor & Francis, 2009.

[46] Anders Jansson, Mikael Erlandsson, and Anton Axelsson. Collegial Verbalisation – The value of an independent observer: An ecological approach. *Manuscript submitted for journal publication*, 2015.
[47] Anders Jansson, Eva Olsson, and Mikael Erlandsson. Bridging the gap between analysis and design: improving existing driver interfaces with tools from the framework of cognitive work analysis. *Cognition, Technology & Work*, 8(1):41–49, November 2005.
[48] Anders Jansson, Patrik Stensson, Ida Bodin, Anton Axelsson, and Simon Tschirner. Authority and Level of Automation. In Masaaki Kurosu, editor, *Human-Computer Interaction. Applications and Services*, volume 8512 of *Lecture Notes in Computer Science*, pages 413–424. Springer International Publishing, Cham, 2014.
[49] Martin Joborn, Per Leander, Tomas Lidén, and Thomas Nordmark. Energy efficiency and on-time performance for heavy haul trains. In *International Heavy Haul Association Conference, Conference Proceedings*. LKAB, 2011.
[50] Daniel Kahneman. *Thinking, Fast and Slow*. Penguin Books, 2011.
[51] Daniel Kahneman and Gary Klein. Conditions for intuitive expertise: a failure to disagree. *The American psychologist*, 64(6):515–26, September 2009.
[52] R. Kalman. On the general theory of control systems. *IEEE Transactions on Automatic Control*, 4(3):110–110, 1959.
[53] Arvid Kauppi. *A Human–Computer Interaction Approach to Train Traffic Control*. Licentiate, Uppsala University, 2006.
[54] Arvid Kauppi, Johan Wikström, Bengt Sandblad, and Arne W. Andersson. Future train traffic control: control by re-planning. *Cognition, Technology & Work*, 8(1):50–56, December 2005.
[55] Lena Kecklund, M. Ingre, G. Kecklund, M. Söderström, T. Åkerstedt, E. Lindberg, Anders Jansson, Eva Olsson, Bengt Sandblad, and P. Almqvist. The TRAIN-project: Railway safety and the train driver information environment and work situation-A summary of the main results. In *Signalling Safety*, 2001.
[56] Gary Klein. Naturalistic decision making. *Human factors*, 50(3):456–60, June 2008.
[57] Gary Klein, Roberta Calderwood, and Anne Clinton-Cirocco. Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, 4(3):186–209, 2010.
[58] Gary Klein, David D. Woods, Jefferey M. Bradshaw, Robert R. Hoffman, and Paul J. Feltovich. Ten challenges for making automation a “team player” in joint human–agent activity. *IEEE Intelligent Systems*, 19(6):91–95, 2004.
[59] Mikael Laaksoharju. *Designing for Autonomy*. Phd thesis, Uppsala University, 2014.
[60] David R. Large, David Golightly, and Emma L. Taylor. The Effect of Driver Advisory Systems on Train Driver Workload and Performance. In *Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014, Southampton, UK, 7-10 April 2014*, page 335. CRC Press, 2014.
[61] Marco Lüthi. *Improving the efficiency of heavily used railway networks through integrated real-time rescheduling*. PhD thesis, ETH Zürich, 2009.

[62] Judy McKay and Peter Marshall. The dual imperatives of action research. *Information Technology & People*, 14(1):46–59, March 2001.

[63] Ronald W. McLeod, Guy H. Walker, and Neville Moray. Analysing and modelling train driver performance. *Applied ergonomics*, 36(6):671–80, November 2005.

[64] F. Mehta, C. Rößiger, and M. Montigel. Latent energy savings due to the innovative use of advisory speeds to avoid occupation conflicts. In *International Conference on Computer System Design and Operation in the Railway and other Transit Systems*, volume 114, pages 99–108, Southampton, August 2010. WIT Press.

[65] Neville P. Moray and Beverly M. Huey. *Human Factors Research and Nuclear Safety*. National Academies Press, Washington DC, 1988.

[66] Donald A. Norman. Cognitive engineering. *User centered system design*, pages 31–61, 1986.

[67] Donald A. Norman. The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 327(1241):585–593, April 1990.

[68] Eva Olsson and Anders Jansson. Work on the bridge – studies of officers on high-speed ferries, 2006.

[69] Eva Olsson, Niklas Johansson, Jan Gulliksen, and Bengt Sandblad. A participatory process supporting design of future work. *Ergonomics An Introduction*, pages 127–160, 2007.

[70] Judith Orasanu and Terry Connolly. The Reinvention of Decision Making. In *Decision making in action: Models and Methods*, pages 3–20. Ablex Publishing, 1993.

[71] Konstantinos Panou, Panos Tzieropoulos, and Daniel Emery. Railway driver advice systems: Evaluation of methods, tools and systems. *Journal of Rail Transport Planning & Management*, 3(4):150–162, November 2013.

[72] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3):286–297, May 2000.

[73] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2):140–160, 2008.

[74] Stephen J. Payne. Users' Mental Models: The Very Ideas. In *HCI Models, Theories, and Frameworks: Toward a Multidisciplinary Science*, pages 135–156. Morgan Kaufmann, 2003.

[75] Charles Perrow. *Normal Accidents: Living with High Risk Technologies*. Basic Books, New York, 1984.

[76] Laura Pickup, Nora Balfe, Emma Lowe, and John R Wilson. 'He's not from around here!' The significance of local knowledge. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, pages 357–366. Taylor & Francis, 2013.

[77] Karsten Rahn, Christian Bode, and Thomas Albrecht. Energy-efficient driving in the context of a communications-based train control system (CBTC). In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, pages 19–24. IEEE, August 2013.

[78] Xiaolu Rao, Markus Montigel, and Ulrich Weidmann. Potential railway benefits according to enhanced cooperation between traffic management and automatic train operation. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, pages 111–116. IEEE, August 2013.

[79] Amnon Rapoport. Research paradigms for studying dynamic decision behavior. In D. Wendt and C. Vlek, editors, *Utility, probability, and human decision making*, pages 349–369. Reidel, Dordrecht, 1975.

[80] Jens Rasmussen. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):257–266, May 1983.

[81] P. Reason. Choice and Quality in Action Research Practice. *Journal of Management Inquiry*, 15(2):187–203, June 2006.

[82] E. M. Roth, J. Multer, and T. Raslear. Shared Situation Awareness as a Contributor to High Reliability Performance in Railroad Operations. *Organization Studies*, 27(7):967–987, July 2006.

[83] Bengt Sandblad, Arne W. Andersson, Arvid Kauppi, and Gunnika Isaksson-Lutteman. Development and implementation of new principles and systems for train traffic control in Sweden. In B. Ning, C.A. Brebbia, and N. Tomii, editors, *Computers in Railways 12*, volume 114, pages 441–450. WIT Press, August 2010.

[84] N. B. Sarter, David D. Woods, and C.E. Billings. Automation surprises. In *Handbook of Human Factors & Ergonomics*, chapter 57, pages 1926–1943. Wiley, 2nd edition, 1997.

[85] Helen Sharp, Yvonne Rogers, and Jenny Preece. *Interaction design: beyond human-computer interaction*, volume 11. Wiley, Chichester, West Sussex, England, 2007.

[86] Herbert A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118, 1955.

[87] Neville A. Stanton and Paul M. Salmon. Is situation awareness all in the mind? *Theoretical Issues in Ergonomics Science*, 11(1-2):29–40, January 2010.

[88] Neville A. Stanton, R. Stewart, D. Harris, R. J. Houghton, C. Baber, R. McMaster, Paul M. Salmon, G. Hoyle, G. Walker, M.S. Young, M. Linsell, R. Dymott, and D. Green. Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics*, 49(12-13):1288–311, 2006.

[89] Patrik Stensson and Anders Jansson. Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts. *Ergonomics*, 57(3):455–470, March 2014.

[90] Lucy Suchman. *Plans and Situated Actions*. Cambridge University Press, 1987.

[91] Peter M. Todd and Gerd Gigerenzer. Précis of Simple heuristics that make us smart. *The Behavioral and brain sciences*, 23:727–741; discussion 742–780, 2000.

[92] Peter M. Todd and Gerd Gigerenzer. Putting naturalistic decision making into the adaptive toolbox. *Journal of Behavioral Decision Making*, 14(5):381–383, December 2001.

[93] Kim J. Vicente. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Lawrence Erlbaum Associates, 1999.

[94] Colin Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2012.

[95] Christopher D. Wickens. Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):397–403, June 2008.

[96] John R. Wilson and Beverley J. Norris. Human factors in support of a successful railway: a review. *Cognition, Technology & Work*, 8(1):4–14, December 2006.

[97] David D. Woods. Coping with complexity: the psychology of human behaviour in complex systems. In L P Goodstein, H B Andersen, and S E Olsen, editors, *Tasks, Errors, and Mental Models*, chapter Coping wit, pages 128–148. Taylor & Francis, Inc., Bristol, PA, USA, 1988.

[98] David D. Woods and Matthieu Branlat. Hollnagel's test: being 'in control' of highly interdependent multi-layered networked systems. *Cognition, Technology & Work*, 12(2):95–101, April 2010.

[99] Li Yang, Tomas Lidén, and Per Leander. Achieving energy-efficiency and on-time performance with Driver Advisory Systems. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, pages 13–18. IEEE, August 2013.

[100] Tao Zhang. *Using measures of situation awareness to characterize mental models in inductive reasoning tasks*. Phd thesis, North Carolina State University, 2009.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 1237

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-246719

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2015