FOI is an assignment-based authority under the Ministry of Defence. The core activities are research, method and technology development, as well as studies for the use of defence and security. The organization employs around 1350 people of whom around 950 are researchers. This makes FOI the largest research institute in Sweden. FOI provides its customers with leading expertise in a large number of fields such as security-policy studies and analyses in defence and security, assessment of different types of threats, systems for control and management of crises, protection against and management of hazardous substances, IT-security an the potential of new sensors.

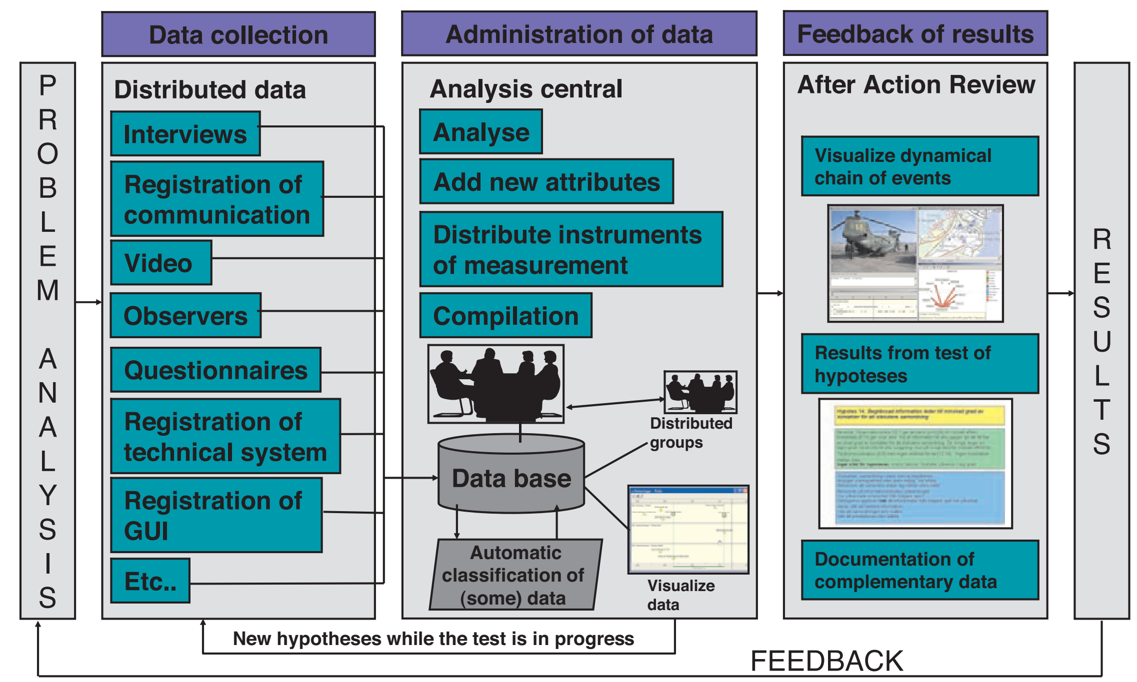
FOI
Defence Research Agency
Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

Methodological tools and procedures for experimentation in C2 system development - Concept development and experimentation in theory and practice

PER WIKBERG, PÄR ANDERS ALBINSSON, DENNIS ANDERSSON, TORBJÖRN DANIELSSON, HELENA HOLMSTRÖM, MATTIAS JOHANSSON, MIRKO THORSTENSSON, MARIA-ELENA WULFF



FOI-R-- 1773 --SE Scientific report
ISSN 1650-1942 December 2005

Command and Control Systems

Per Wikberg, Pär Anders Albinsson, Dennis Andersson, Torbjörn Danielsson, Helena Holmström, Mattias Johansson, Mirko Thorstensson, Maria-Elena Wulff

Methodological tools and procedures for experimentation in C2 system development - Concept development and experimentation in theory and practice

FOI-R--1773--SE
ISSN 1650-1942

Command and Control Systems
Scientific report.

December 2005

Issuing organization	Report number, ISRN	Report type
FOI – Swedish Defence Research Agency Command and Control Systems P.O. Box 1165 SE-581 11 Linköping	FOI-R--1773--SE	Scientific report
	Research area code 4. C4ISTAR	
	Month year December 2005	**Project no.** E-7080
	Sub area code 49 Interdisciplinary Projects regarding C4ISTAR	
	Sub area code 2	
Author/s (editor/s) Per Wikberg Per Anders Albinsson Dennis Andersson Torbjörn Danielsson Helena Holmström Mattias Johansson Mirko Thorstensson Maria-Elena Wulff	**Project manager** Per Wikberg	
	Approved by	
	Sponsoring agency	
	Scientifically and technically responsible Per Wikberg	
Report title Methodological tools and procedures for experimentation in C2 system development - Concept development and experimentation in theory and practice		

Abstract

This report is the concluding documentation of the FOI-project Evolva which has been carried out between 2003 and 2005 with an objective to produce long-term knowledge about development and evaluation methodology serving to improve the command and control system of the armed forces, considering operational and technical as well as economical requirements. The project has had an inter-disciplinary constitution and the work has focused on developing a domain relevant ”practice” based on the hypothetical-deductive approach to scientific investigation. The two major goals of the project have been to generate a body of research and to create an education package on research methods for officers. Important areas of work includes: 1) modeling as an approach to define experimental settings based on vaguely defined problems; 2) efficient data-collection techniques by exploiting network technologies; 3) alternatives to ”formalized laboratory experiments” and “field exercises” as test environments; 4) alternatives to the written reports to transfer results to clients. The report is intended to be a position paper describing our approach to experimentation in order to facilitate future cooperation on this issue. The report describes our theoretical perspective. The practical applications of these theoretical principles are illustrated by a number of cases.

Keywords

Experimentation, Command and control, Method

Further bibliographic information	Language English
ISSN 1650-1942	**Pages** 79 p.
	Price acc. to pricelist

Utgivare

FOI - Totalförsvarets forskningsinstitut

Ledningssystem
Box 1165
581 11 Linköping

Rapportnummer, ISRN
FOI-R--1773--SE

Klassificering
Vetenskaplig rapport

Forskningsområde
4. Ledning, informationsteknik och sensorer

Månad, år
December 2005

Projektnummer
E-7080

Delområde
49 Breda projekt inom ledning, informationsteknik och sensorer

Delområde 2

Författare/redaktör

Per Wikberg
Mirko Thorstensson
Per Anders Albinsson
Maria-Elena Wulff
Dennis Andersson
Torbjörn Danielsson
Helena Holmström
Mattias Johansson

Projektledare
Per Wikberg

Godkänd av

Uppdragsgivare/kundbeteckning

Tekniskt och/eller vetenskapligt ansvarig
Per Wikberg

Rapportens titel

Metoder för utveckling och värdering av ledningssystem

Sammanfattning

Denna rapport är den sammanfattande dokumentationen av FOI-projektet Evolva som genomfördes under perioden 2003-2005. Syftet med projektet var att bygga upp långsiktig kunskap om utvecklings- och värderingsmetodik bidragande till att utveckla Försvarsmaktens ledningssystem med hänsyn till såväl operativa och tekniska som ekonomiska krav. I projektet ingick personal från flera olika kompetensområden. Arbetet fokuserade på att utveckla en domänspecifik praktisk tillämpning av vetenskaplig metod enligt den hypotetisk-deduktiva metoden. De två övergripande målen med projektet var en väldokumenterad forskningsbas samt kursmaterial för utbildning av officerare avseende utvärdering försöksmetodik. Viktiga arbetsområden för projektet har varit: 1) Modellering för att definiera försöksdesign. 2) Effektivisering av tekniker för datainsamling. 3) Alternativ till laboratorieexperiment och fältövningar som försöksmiljö för att studera ledning. 4) Alternativa sätt att kommunicera försöksresultat. Rapporten är avsedd som ett ”position paper” med syftet att kunna användas vid samarbete kring experiment. Rapporten beskriver vår teoretiska ansats och praktisk tillämpning illustreras med hjälp av ett antal case.

Nyckelord

Experiment och försök, Ledning, Metod

Övriga bibliografiska uppgifter

Språk Engelska

ISSN 1650-1942

Antal sidor: 79 s.

Distribution enligt missiv

Pris: Enligt prislista

Table of Contents

Preface

Evolva has been a three year research and development project at the Swedish Defence Research Agency (FOI), funded by the Swedish Armed Forces. The project has been a part of FoT area 4, C4ISR, and was initiated in January 2003 and ended in December 2005. The total budget has been ten million Swedish kronor. The ambition has been to develop knowledge on tools and procedures for development and evaluations of command and control systems. Thus, the focus of the project has been on methodological issues rather than generating knowledge on domains associated with command and control such as situation awareness or decision making.

We have employed an interdisciplinary approach with diverse competences, such as behavioral sciences, and technology and military operational analysis, to achieve the objectives of the project. Personnel from three departments at two of FOI's divisions (Command and Control Systems and Defence Analysis) have participated. The following people have been members of the project group: Per Wikberg (project manager), Pär-Anders Albinsson, Dennis Andersson, Torbjörn Danielsson, Mattias Johansson, Helena Holmström, Mirko Thorstensson, and Maria Elena Wulff. In addition, Johan Stjernberger, Mia Löw, Liselotte Dahlén, and Malin Östensson also have been members of the project team part of the time.

In total, the project has arranged and/or participated in twelve experimentation studies. Besides this report, a total of fifteen scientific references have been produced by the project, three are FOI-reports and the rest are international publications. In addition, five memos and one article in FOI's magazine "Framsyn" have been published. Another product of the project is a university course on research methods, tailored for officers working in development projects. The first batch of students completed the course in June 2005. Parts of the work of Evolva are also in included in a licentiate thesis (Albinsson, 2004) which was presented at Linköping University in December 2004.

We would like to thank the following people and organizations, for contributing to Evolva's work: Bo Strangert and Jonas Wikman at the Communication Research Unit at Umeå University, Johan Lundin at the Swedish National Defence College, members of the Swedish Armed Forces' programs LedSystM and LedSystP, the staff at the Swedish NBC Defence Centre and Norrland's Dragoon Regiment K4, and The Singapore Armed Forces Centre for Military Experimentation.

Per Wikberg
Project Manager

1 Introduction

1.1 Background

Experimentation (Alberts & Hayes, 2002) is an important feature of the Swedish Armed Forces' transformation from being invasion focused to a flexible network-based defense founded on battle groups with an enhanced capability to participate in international operations (Swedish Defense Forces, 2002, 2005). Demonstrations and tests of new technology and methods are used to resolve uncertainties, test technical (and financial) feasibility, and illustrate alternative courses of action. The results are analyzed and assessed through study activities before deciding on continued development activities.

The transformation process was launched in 2002 with a number of interrelated programs controlled by the Armed Forces Head Quarter. The technological implementation process is lead by the Swedish Defence Material Administration (FMV) in the program 'LedSystT'. Another important sub-program in the process is 'LedsystM', affiliated to the joint Forces Command (OPIL), which is responsible of development of command and control (C2) methods. 'LedSystP', is the third major sub-program aiming at developing the human resource management in integration with the development towards network-based defense. The first phase of this transformation process will end in 2006.

Evolva has been carried out between 2003 and 2005. Consequently, the LedSyst program has had a strong influence on our work. The ambition of this project has been to develop knowledge on tools and procedures to support this and forthcoming experimentation processes. The project has had an inter-disciplinary constitution with a mix of competences from behavioral science, engineering, and military-operation analysis.

1.2 The need of experimentation

As Alberts & Hayes notes, experimentation is an area of emerging competence rather than an area with a well-developed theoretical foundation and a mature practice. Is there any real need of developing such competence? Is not the "traditional way" of learning by doing in exercises without having a "scientific approach" good enough?

Historically, investments in military systems have often been of long term nature and thus research and system development had a perspective of decades in advance. Competence on implemented systems developed over time based on officers specializing on these systems and developing routine procedures. However, the last decades have shown that the lifecycle of systems tend to be drastically shortened. Such an example is the digital command and control systems developed and used by the different armed services during the nineties. Many of these systems are optimized for a cold war scenario with Sweden acting as a neutral state with a limited need of systems compatible with other nations C2-systems. Today the development focuses on a common system for all the services of the Armed Forces where Sweden acts as one of many nations in joint international operations. The consequence of this is that even though the systems are relatively newly developed and implemented, they can only be re-used partially.

We argue that the new defense implies a move away from the traditional static structures and towards a situation were constant change is the normal state. During the three years this project was undertaken, the transformation process of the Swedish Defence Forces has changed focus. Instead of developing “demonstrators” in order to enhance strategic recapture of defense ability, the focus has changed to an ambition of combat ready battle groups for international operations. Rapid changes in tactics, tasks, technology, and organization create a demand of an enhanced ability to develop technical and organizational solutions. Consequently, competence can not just be based on routine. The only way to adapt to change is to try out solutions to handle these changes. Competence on change, the core of experimentation, will be a crucial factor in the process of establishing and maintaining these combat ready battle groups. Research and development must become a more integrated and natural part of the Defence Forces every day activities.

Developing competence on experimentation is thus of central importance. The narrowed time for experimentation on new solutions accentuate the Albert & Hayes (2002) notion that *“The single most important consideration for those responsible for experimentation design, whether single experiments or campaigns, is to ensure current expertise is available to support the plan. Almost without exception, this will mean doctoral level (Ph.D.) training and a decade or more of experience in actually designing and conducting experiments. (…) “Bad experiments, which cloak weak or false knowledge in an aura of science, will make mischief, not music. At a minimum they will slow the process of understanding the impact of innovation on transformation (by forcing other research and experimentation to learn and demonstrate that they are wrong) and cost money (by encouraging investments that will not pay off as expected). At a maximum, bad experiments will lead to flawed mission capability packages (MCPs) that fail in the field”. (Albert & Hayes, 2002 pp 6-9).*

1.3 The aim and scope of the project and this report

The overall objective of the project was to “*produce long-term knowledge about development and evaluation methodology, serving to improve the command and control system of the armed forces, considering operational and technical as well as economical requirements.”*

The two major aims of the project have been:

- To generate a well documented body of research by publishing results from the studies on methods conducted in the project. The purpose has been quality assurance of tools and procedures.
- To create a training and education package on research methods for officers working in development projects. The purpose has been to make it possible to transfer knowledge about research methods to individuals responsible for experimentation in the defense forces.

Consequently, the scope has not been to accumulate knowledge on certain domains of command and control such as decision making or situation awareness. Again, this project has been a project on research methodology. However, as the project aims at developing test methodology, it has been found advisable to vary the problem formulations between tests. Problems studied include access to real-time information, automated sensor-information processing, delegation of command and control responsibility, and comparison of different ways to present a situational picture.

The aim of this report is to serve as a position paper describing our approach to experimentation in order to facilitate future cooperation on experimentation. The text is intended to give individuals engaged in other experimentation enterprises an outline of our theoretical and practical point of departure to tackle military experimentation. Consequently, the report is not a handbook on scientific method.

1.4 Practical considerations

The core activity has been to gain practical experience on experimentation. Therefore, a relatively large number of studies have been undertaken. To accomplish this, it has been necessary to seek cooperation with other projects and organizations. Pooling of resources has allowed us to carry out more studies. As a consequence, the role of the project has varied between different studies. However, a common feature of all studies has been an explicit aim to explore the areas of work defined for the project:

- Modeling of experimental design.
- Test environment.
- Data collection.
- Feedback of results.

The cooperating projects have normally had another purpose with the studies. A disadvantage with this circumstance has been that we have not been able to conduct a coherent experimental campaign (Alberts & Hays, 2002, 2005).

1.5 Cooperation

Over time, the project has cooperated with a wide range of projects and organizations:

LedSystM, which is responsible for development of command and control (C2) methods, has been a valuable client for our work. As a part of the demonstrator program, we have been able to test some of our ideas on research methodology in a series of experiments on command and control methods- and principles;

LedSystP has been the project that made it possible to initiate the course on research methods which was one of the aims of our project;

the Swedish Army's Ranger Battalion, former Norrlands Dragoon Regiment, has given us opportunities to use real units in order to evaluate the possibility to use commercial software as low cost simulators as test environment;

the Swedish NBC Defence Centre in an experiment with the national NBC task force;

the Singapore Armed Forces' Centre for Military Experimentation together with the 8th Singapore Armour Brigade (8 SAB), 40th Battalion of the Singapore Armour Regiment (40 SAR), and the Singapore Armed Forces' TRADOC in an experiment prepared and conducted in the winter of 2005;

the Swedish War College. Cooperation during the Singapore experiment, experimentation in the command and control laboratory and participation in workshops and in seminars;

finally, several projects conducted at FOI: MARULK, GRU, LINK, KoLäge, NBC Demo, Commercial game technology for distributed learning and FoRMA. These projects have worked on issues related to the Armed Forces transformation process. More information is available on FOI's website, www.foi.se.

1.6 Project activities

Twelve different experiments have been undertaken, most of them in cooperation with other projects and organizations.

A university course on research methods tailored for officers working in development projects have been initiated and carried out. A second course is planned for 2006.

Two workshops and three seminars have been conducted during the project.

Besides this report a total of 14 scientific references have been produced by the project. Three of these are FOI-reports and the rest are international publications. In addition, five memos and one article in FOI's magazine "Framsyn" have been published. A list of the projects publications is included in chapter 2.

1.7 Comment to the reader

This is a report to summarize three years of work on developing procedures for experimentation. Consequently, the report is a position paper describing our approach in order to facilitate future cooperation on experimentation. Our intended reader is an individual with knowledge or interest in scientific methods and its application to experimentation. To make our approach explicit for our intended reader, we have briefly described our theoretical perspective in each section. To illustrate our practical application of the theoretical principles, a number of cases are included in the report. Readers more interested in practical application might focus on these sections.

This report is not a handbook on scientific method. For that purpose, we recommend Strangert 2005, written in Swedish, or Graziano & Raulin (2004). For project managers who need a guide to the rationale of experimentation in a military context, we recommend Alberts & Hayes books on experimentation (2002) and experimental campaigns (2005).

Chapter 2 of this report is an account on the two major goals of this project, a well documented research base, and a course on research method. Chapter 3 is a brief description of the methodological approach we recommend. The following four chapters present and discuss the four different areas of work identified for the project. A discussion in the final chapter concludes the report.

2 Project goals

The two major goals of the project have been:

- to create an education package on research methods for officers working in development projects;
- to generate a well documented body of research by publishing results from the studies on methods conducted in the project.

2.1 Education package

How should results from experimentation be brought back to the clients? This is an important issue for all research projects. A basic idea of commissioned research is that the results should be delivered in a way that enables the clients to put the recommendations to work in a practical sense.

Our assumption was that this meant turning the results and experiences from our own research into an education package including theory and instructions for evaluation of command and control in real tasks and missions, exercises, demonstrations, and experimental studies. The target group was officers working in development projects within the Swedish Armed Forces. The aim was to create not only a ‘handbook of methodology’, but also to develop and customize practical and pedagogical principles. The reason was that, in order for a student to make real use of the knowledge acquired in such a course, the knowledge has to be concrete and practical. The people doing the hands-on work in military development projects need to understand, as well as be able to apply, the principles behind concepts like validity, reliability, measurement, etc. Such capabilities are necessary for carrying out and understanding qualified experimental and developmental work, and for communicating needs and results to the assigners.

During 2004-2005, Evolva, in collaboration with the Communication Research Unit at Umeå University, arranged a ‘trial run’ for a one-year distance education course on research methods. The course was partly commissioned and funded by the Swedish Armed Forces human resource development program LedSystP. Eight officers participated in this trial run. During the academic year 2005-2006, the ambition is to admit twelve students.

The goals of the course were to:

- develop personal skills and methodological competence to carry out experimentation in development projects;
- contribute to quality assurance of the studies carried out by the students;
- establish general and sustainable competence about validation and verification of results.

The course was divided into two parts: *“Scientific procedures in the design of development projects”* and *“Research methods in a naturalistic context”*. During the course, basic concepts and principles of research methodology were covered. The concept of the education package was based on the idea that each participant had their own real study assignment (i.e. a development task or project which was a part of their regular work), to which they could apply the methodological concepts and principles that were taught on the course. The course required an average of eight hours work per week for one year. On completion the students were given ten university credit points.

The course outline was based on personal supervision throughout the development project. Supervision was given using a web based educational platform and regular meetings. The students' examination included planning, conducting, and documenting their respective experimental studies, as well as presenting their work in a poster session and a written report.
Combining and integrating practical development projects and education make it possible to achieve good results at a relatively low cost. It also does not require the students to spend much time away from their regular work. In order to achieve this synergy, it is very important that the experimentation activities performed as a part of the course are relevant to the students' real development tasks in their regular work. Another significant and positive effect of this integrating approach is that the development projects, in which the students carried out their studies, are provided with extra competence through the supervision on research design and methodology.

A more thorough documentation of the course, along with an evaluation of the trial run of the course, is found in Wikberg, Strangert, & Danielsson (2005a & 2005b). For the compendium on research methodology which was produced and used on the course, see Strangert (2005). Both these publications are in Swedish. Besides the compendium, the folowing references were used at the course:

Alberts, D., & Hayes, R. (2002). *Codes of best practice for experimentation*. The Command and Control Research Program, USA: CCRP publication series.

Alberts, D., & Hayes, R. (2005). *Campaigns of Experimentation: Pathways to Innovation and Transformation*. The Command and Control Research Program, USA: CCRP publication series.

Graziano, A. M., & Raulin, M. L. (2004). *Research methods. A process of inquiry.* 3rd ed. New York: Longman.

Sandberg, L., & Sandberg, R. (2002). *Elementär statistik.* Stockholm: Studentlitteratur.

Silverman, D. (2001). *Interpreting qualitative data. Methods for analyzing talk, text and interaction.* 2nd ed. London: Sage.

2.2 Documented body of research

Generally, the studies that Evolva has arranged and/or participated in have had multiple aims. One aim has always been to satisfy the client's needs, in terms of exploring the identified problem area and presenting results related to the research questions. Another aim, which is the one that the project has focused on, has been the methodological problems of conducting experimentation. Examples of such problems are: "How should feedback of the results be given to the clients?"; "What are the conditions for using virtual computer game settings for command and control research purposes?"; and "How do people with different roles and backgrounds interact in a modeling session aiming to create a research design?".

One of the project's goals has been to generate a well documented body of research on methods and tools for developing military command and control systems. In practice, this has involved an ambition to present our results in international publications and at international conferences, in order to attain scientific quality assurance. A list of publications is presented in appendix 1.

3 Methodological approach

3.1 Background

Considering the high costs of developing military systems, the experimentation process should aim at a high degree of scientific rigor and professionalism. The experiment and development process — the road from conceptual thinking to implemented practice — is not without its challenges. We have addressed some of these challenges in our work, for example:

Vaguely defined research problems. The research problem is often relatively vague and not theoretically well developed. The clients' requirement on a planned experiment is usually described on a pragmatic level, such as ”we want to learn more about the service concept” or ”we would like to compare push and pull as principles for distribution of information”.

Time pressure and limited resources. Our experience is that the time available for preparing, conducting, and documenting an experiment is very limited. In system development, and consequently in research supporting development, time is money.

Different demands from the involved parties. The same experiment is often used for several different, sometimes conflicting, purposes. Furthermore, the complexity of command and control often makes it necessary to involve a cross section of domain competences. Consequently, different demands need to be unified in the experimental design.

Our work has focused on developing a ”practice” based on a scientific approach to conduct experimentation in the context of the Defence Forces' progressive development of command and control. Such a practice should take into account circumstances such as those listed above and still maintaining a scientific rigor. The ambition is to be able to empirically test the organization's ”best guesses” with a limited amount of resources in time, money, training, and technical aids still gaining knowledge from experimentation.

3.2 The scientific method

The golden standard of generating knowledge is of course the scientific method. Perhaps one of the most basic features of the scientific method is the distinction between two basic processes of research; discovery and justification (Reichenbach, 1938).

Discovery refers to the process of coming up with ideas that might be of interest for testing. One famous example is the case of Fleming who discovered that bacteria had died in contaminated petri dishes. This had most certainly happened before but it was Fleming that realized that it might be of significance. Consequently, the hypothesis that the mould, *Penicillium Notatum*, could kill bacteria was tested. Fleming was awarded the 1945 Nobel prize in medicine for this discovery. The discovery process is basically a creative process and normally we do not refer to this process when we talk about the “scientific method”. However, this generation of hypotheses is of central importance for the development of knowledge. Insights, ideas, and assumptions must be formalized into hypotheses.

Justification refers to the process of empirically testing the hypotheses formulated in the discovery phase. Most people refer to the procedures undertaken in this process when they talk about the “scientific method”. The justification process might be viewed as a combination of:

1) Rationalism. Rules of scientific reasoning, including the formalized rules of logic and, more or less, accepted principles such as the principle of falsification (Popper, 1959); and
2) Empirism. Rules of which methods are acceptable in order to observe reality.

The mainstream application of these rules is the hypothetical-deductive method as described by Hempel (1952, 1965, & 1966) and Popper (1959). This approach is based on formulating and testing "best guesses", or hypotheses, to answers of identified problems. The hypothetical-deductive method might be summarized in the following rules (Mårtensson & Nilstun, 1988):

A) The hypothesis H states that a certain condition will have the consequence C. For example, a new C2 is assumed to lead to a more efficient use of resources. If an empirical investigation reveals that this is not the case then the hypothesis is rejected as false. The conclusion is deductive, as the hypothesis is obviously not true in all cases.

B) If an investigation finds that empirical data supports the hypothesis, the hypothesis is tentatively accepted as true. The conclusion is only made on a preliminary basis. It might be rejected later as the conclusion is inductive. The investigation has only shown that the hypothesis is true in this case.

C) In practice, the hypothesis H is tested in a certain context containing a set of circumstances $Z_1 \ldots Z_n$. If H is rejected, the cause might be found among the set of Z.

The scientific process is thus a process where ideas are formulated as hypotheses which then are tested empirically. Ideally, "good ideas" will be accepted while "bad ideas" and false assumptions will be rejected. The general conception is, however, that one single test is normally not enough to accept a hypothesis. Replication is a basic feature of science (Gratziano & Raulin, 1996). The hypothesis needs to be tested in different conditions, with different methods and by different researchers. As evidence accumulates the hypothesis will be accepted as being in accordance with the truth.

It must be mentioned that there are well established approaches to science which differs from the traditional approach described above. A discussion on this matter is, however, beyond the scope of this report.

The starting point to our work has been the mainstream application to science described above. The organizations' "best guesses" about what they think would be proper solutions on identified problems should be empirically tested or as Staw (1983) puts it "… experimentally undertake the seemingly best course of action" (p. 426). Such an approach means a strive towards an "experimenting organization" (Campbell, 1969, 1974; Staw, 1983).

3.3 Military experimentation

As mentioned in section 3.1 hypotheses are tested in a certain context containing a set of circumstances which might affect the result. Military experimentation is normally conducted in a much more complex setting compared to the settings used in traditional laboratory experiments. We argue that there are basically two research traditions that might be applicable in evaluating and changing complex systems and behaviors in a natural context. Both traditions, multi-factor case studies and action research, take advantage of natural variance and uncertainty instead of reducing or excluding it (Strangert, 2005).

Case studies. Multi-factor case studies assume that the studied system, irrespective of its complexity, has a reasonably simple structure which is possible to study and evaluate. Instead of using the "randomized assignment to treatments" and "laboratory control" models as research design, phenomena are studied in their real context. Measurement normally involves an array of different variables, thus using these as a set of indicators of the behavior of the phenomena in the studied context. As replication of studies might be a problem, generalization of findings is based on theoretical considerations (Yin, 1984, 1993). Examples of organizational development using case studies come from law, business, medicine, and public policy (Llewellyn, 1948; Stein, 1952; Towl, 1969, Windsor & Greanias, 1983). In a practical sense, case studies might be appropriate when different solutions are contrasted to each other such as comparing two command and control systems. A more thorough discussion and analysis of different action research programs is found in Yin (1984, 1993). An example of a case study conducted in Evolva is presented in Wikberg, Andersson, Berggren, Hedström, Lindoff, Rencrantz, Thorstensson, and Holmström (2004). The case study is also described in section 5.4.2 as case 5b.

Action Research. Action research is similar to experiments as it tries to manipulate the studied system based on theoretical considerations. In the 1940's, a tradition of research evolved which tried to solve real organizational problems and improve conditions by using carefully designed programs and interventions. Lewin (1952) described organizations as dynamic systems only possible to understand by using interventions; *"if you want to study an organization try to change it*". Lewin also stressed the importance of having a model, ideally with predictive power, of how the system works; *"there is nothing as practical as a good theory"*. Without a model the intervention would be a random process, neither efficient nor ethical. The 'action-research model' is an iterative sequence of actions: theorizing, intervening, gathering data on the effects of the intervention, and then checking the theory prior to the next intervention. Whether or not the predicted consequences occur becomes a test of the initial theory. Other researchers have also influenced this school of 'planned change'; Edgar Schein and Chris Argyris are perhaps the most well known. Examples of organizational development using action research as a cornerstone can be found in the mining industry (Gavin, 1984), health care (Shani & Eberhardt, 1987), or banking (Santalainen & Hunt, 1988). Action research might be appropriate in the successive development of a single solution; for example a new command and control process. A more thorough discussion and analysis of different action research programs is found in Weisbord (1987) or in French & Bell (1999). An example of a study based on an action-research approach is presented in section 7.4.1 as case 7a.

Undeniably, there are great similarities between the approaches, a series of case studies might be considered as an action-research approach and theoretical considerations are of course also important in action research. Both traditions allows for an experimental as well as an observational method. Our main point is that these two research traditions have addressed the theoretical and practical problems present in military experimentation. As the studies undertaken in the project have all comprised relatively complex sets of factors and circumstances, we have generally relied on these two research traditions. We will not expound the theoretical framework of these approaches further in this report. Instead we refer to the suggested references for further studies for those who might be interested.

3.4 The focus of the project

As mentioned, our work has focused on developing a "practice" for experiment and development - the road from conceptual thinking to implemented practice – based on the scientific procedure described above. Much of the scientific process is theoretically and practically well developed and has not been the subject of our work. For example, methods of data analysis are already a well developed domain and different command and control measures are explored in other projects. Instead, we have limited our work to some aspects of the scientific procedure which we considered to be less well developed for experimentation on command and control:

- Modeling of experimental design.
- Test environment.
- Data collection.
- Feed back of results.

3.4.1 Modeling of experimental design

As mentioned in section 3.2, it is important to make assumptions and basic facts explicit before any empirical investigation. This ”problem analysis” corresponds to the ”discovery” phase of scientific investigation. In terms of the scientific method, this means to formalize these assumptions as hypotheses (or “best guesses”) possible to test empirically.

The traditional way to undertake this problem analysis in science is by studying relevant literature on the subject. However, this approach effectively excludes the clients from the problem analysis. It is the clients’ “best guesses” that should be tested and not the analysts’. In many cases there are also a lack of relevant research on command and control, especially if the problem is interdisciplinary. Finally, the literature-based problem analysis is time consuming. We argue that relevant literature should support, not be the core of, the problem analysis. We suggest an approach for problem analysis based on modeling where the relevant clients and domain competences are engaged. Our theoretical approach and experience of modeling of experimental design is presented in chapter 4.

3.4.2 Test environments

Every empirical investigation has to be undertaken in some context, i.e. a test environment. Choosing test environments when developing new command and control systems is an important issue. One basic feature of tests and experiments is to gain control of variables and factors by excluding, as far as possible, external influences outside the scope of the experiment. However, when developing new organizational solutions, it is often necessary to conduct experimentation in as real a setting as possible, making it impossible to control extraneous variables. We have especially explored the possibilities to use complements to laboratory experiments and exercises as test environments. In chapter 4, we present our theoretical approach and experience to balance the contradictory demands on realism and control of the test environment.

3.4.3 Data collection

Many methods for data collections are not applicable in applied commissioned research. For example, the time factor of analyzing video recordings is in between 10 to 1000, i.e. one hour of video takes between 10 to 1000 hours to administrate and analyze. We have tried to make data collection in a setting of distributed command and control systems as efficient as possible by using different technical solutions. To achieve this we have used MIND, which is briefly described in section 3.5, as a framework. Our theoretical approach and experience of data collection is presented in chapter 6.

3.4.4 Feedback of results

One challenge is how empirical findings from tests and experiments should be managed and used. The normal procedure within science to present results from experiments is to write reports. That procedure is quite time consuming and is thus often too slow to keep pace with the development process. Considering that decisions on how to proceed with a development project often need to be taken within a shorter time span, it must be possible to deliver the analysis of a test much quicker and more efficient.

A related challenge is to have those who "own the problem", clients, engaged in the process of experimentation. Presumably, the active engagement in the process of designing, conducting, and documenting tests and experiments is the most efficient way to establish a relevant feedback of results. In chapter 7, we present our theoretical approach and experience of feed back of results from experimentation.

3.5 The technical approach

In our work we have used and developed a set of methods and tools, denominated the MIND system in order to make experimentation more efficient. MIND is a framework for computer supported reconstruction and exploration of distributed courses of events. The MIND framework was originally developed to support training, but in the scope of this project and some related projects MIND has been adapted and used to support feedback in an experimentation context.

The system supports data collection from a large amount of data sources. The captured dataset can then be explored through operations such as navigation, exploration, filtering, and synchronization. The MIND framework also provides data-presentation views that aim to make the data as informative as possible, considering the goals of the analyses. It is possible to synchronize and connect all data to a timeline. The key feature of the MIND framework is the ability to build replayable computer models of courses of events.

Figure 3:1. gives an example of how data from different sources can be processed and presented using MIND. The visualization shows the flow of events (unit movements and positions on geographical maps), snapshots of certain interesting events or actions (annotated photographs, videos, and textual observations), and communication (from radio networks or computerized communication systems). The visualization serves as a basis for exploration of the chain of events and thus the replayable model can serve as a documentation of the exercise.

Figure 3:1. A multimedia representation of a distributed tactical helicopter operation presented in the MIND framework.

Traditionally, the MIND framework has focused on presenting and managing 'objective' process data. Supporting feedback of experimentation also necessitates the use of subjective data, such as participant questionnaires and retrospective interviews, and the possibility to connect these data to the MIND mission history. Therefore, we have tried widening the MIND framework to more efficiently incorporate subjective measures. Another issue has been to develop procedures to support hypothesis-driven experimentation.

4 Modeling as a tool to define evaluation measures

4.1 Background.

As we have argued in chapter 3, the initial modeling of the system to be studied is a crucial task in empirical evaluations. Development of complex systems should follow the same principle as science. Ideas and assumptions underlying efforts of changes of organizations and systems should be made explicit and tested empirically as early as possible in a developmental cycle. The alternative, that changes in an organizational system will be based on trial and error or collection of anecdotes, is of course possible but progress will be unsure, inefficient and relatively slow (Alberts & Hayes, 2005).

Normally modeling is referred to as a methodological tool which is used to structure phenomenon for a specific purpose. Thus, modeling can be used for an initial problem analysis. The procedure consist of some sort of group interview were the discussion is summarized by a graphical representation, a model with notations. One key issue is that it often required that clients, i.e. the "owners" of the problem, and domain experts are involved in the modeling process.

Modeling can thereby be said to have the purpose to give an initial outline of a certain problem area in order to create a foundation for the development of theories, structure variable values in order to construct simulations, be a part of the design process, be a problem identifying process etc.

In this context we have had the perspective that the purpose of the modeling is to explicitly translate the client's assumptions of and approach to the problem into problem statements and hypotheses. An example of client's is those responsible for the development of new command and control systems. The core in our modeling approach is that the modeling should produce a base for data collection about the stated problem i.e. it should include a definition concerning how the measure should be conducted. The result from this problem analysis, or modeling, represents the definition of the research design.

Methodological procedures includes extracting ideas from different kind of experts and users of a system and formalize these into a hypothetical model, in order to relatively fast be able to empirically test ideas and assumptions of importance for a developmental project. Seen in the context of science the use of modeling corresponds to the discovery phase.

4.2 Models as representations of empirical processes.

Modeling is the construction of a model, based on data from the systems analysis. There are several definitions of the term 'systems analysis', but a definition usually involves some kind of procedure (more or less formal) for collecting and organizing data about an empirical phenomenon. There is a variety of systems analysis techniques and approaches such as 'task analysis' (Annett et al., 1971; Drury et al., 1987), 'job analysis' (Harvey, 1991), 'content analysis' (Kolbe, 1991; Weber, 1990), 'action analysis' (Singleton, 1979), and 'cognitive systems engineering' (Hollnagel & Woods, 1983; Rasmussen, Pejtersen & Goodstein, 1994). Despite the fact that these techniques differ somewhat when it comes to perspectives and procedures, they are rather similar. They are related to a scientific style of analytically approaching a certain phenomenon, in order to treat or analyze reality as a systematically connected set of elements (Gasparski, 1991).

The following section is a brief summary of some theoretical bases of modeling. Readers primarily interested in practical applications of modeling of experiment design could move on to section 4.3. A basic prerequisite for using modeling to define design of empirical tests is that the resulting models are valid representations of reality. Krantz, Luce, Suppes and Tversky (1971) specify the necessary conditions for representing empirical systems in a scientific sense with a numerical measurement model through two theorems: 1. The representation theorem. 2. The uniqueness theorem.

4.2.1 The representation theorem.

If a specific structure of relations in an empirical system is measurable it is possible to make a homomorphic reproduction of that empirical system into a formal and numerical system (Figure 4:1).

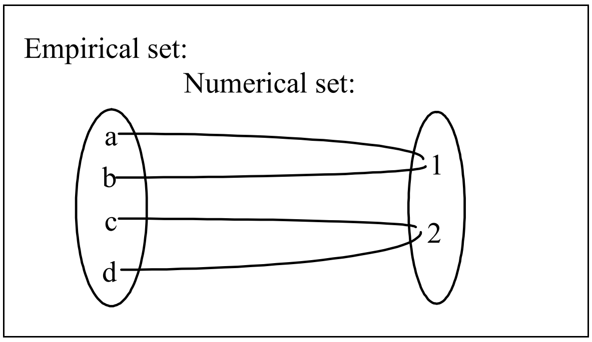


Figure 4:1. Illustration of the representation theorem. The two empirical objects ‘a’ and ‘b’ are represented by the formal and numerical entity ‘1’.

Several different specific empirical objects can be represented according to the theorem by one single numerical value or class. Several different objects can thus be devoted a summarizing variable value or category in order to express a certain quality which the different terms are judged to have in common. For example, a large number of individuals may be categorized into a fewer number of categories, for example ‘expert’ and ‘non-expert’. Consequently, a large empirical variety of relations and qualities may be reduced to a much simpler model with only a few variables.

It is possible to represent reality in a number of ways. A military unit can be represented numerically according to several principles; level of protection, level of training, command structure, etc. Thus, representing the empirical objects is not enough.

4.2.2 The uniqueness theorem.

The representation theorem calls for the relational structure in a chosen representation system to correspond to the relational structure in the empirical system. This puts a limit to the possible transformations between different representation models. If the empirical relation ‘R’ in Figure 4 implies that ‘b’ is twice the size of ‘a’ this means that only the numerical relation ‘r’ can represent this while relation ‘f’ is less suitable. If the empirical relation ‘R’ only implies that ‘b’ is of greater size than ‘a’, then it is possible to use both relation ‘r’ and ‘f’ according to the same reasoning. The uniqueness theorem therefore states that every empirical relation has a certain degree of uniqueness in permitted ways of representation in a numerical model.

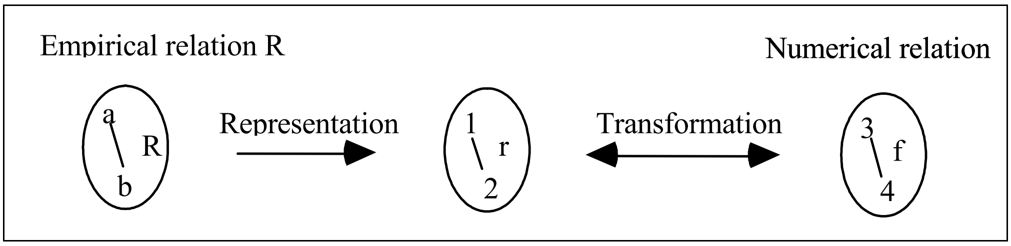


Figure 4:2. Illustration of the uniqueness theorem.

The two theorems are part of measurement theory and numerical representation, but the same reasoning is also relevant for other forms of representation (Wikberg, 1997). Thus, it is also possible to map empirical relational structures into formal qualitative relational structures. For example, any ethnic language can be seen as a natural representation system to represent thoughts, happenings, processes, etc. The language as a representation system has different objects (words) and rules for relations (grammar) which are intended to represent and describe reality. The ethnic language can be translated to other ethnical languages as well as to formal predicate logic. If the empirical relation has a high degree of uniqueness, the number of possible transformations decreases. Based on this assumption, the distinction between quantitative and qualitative research is somewhat blurred as any given reality can be represented by a large or even infinitive number of models (Silverman, 2001).

As a consequence of the theorems above the model must be empirically specified to enable empirical evaluation, i.e. there has to be a definition of the set of empirical elements and the relations between them that corresponds to the model's conceptual terms. Consider the example of a causal model (where A causes B) as shown in figure 4:3.

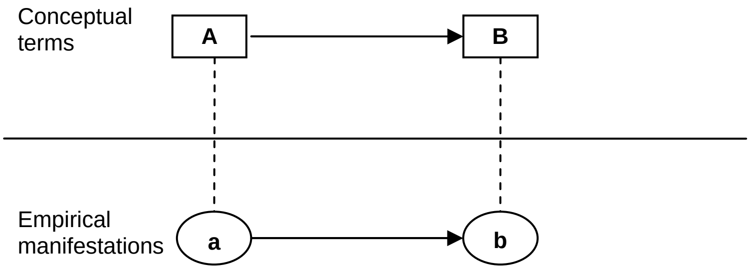


Figure 4:3. Example of a causal model

In order to measure and evaluate this model (figure 4:3), the following conditions have to be met (Wikberg, 1997):

- There must be a set of elements, conceptual terms, explicitly defined:
  (1) A is defined as 'A' . For example Commander is defined as 'the individual responsible of the operation'.
  (2) B is defined as 'B'

- There must be a set of defined formal rules for relations between the elements which determine the possible combinations:
  A affects B (i.e. A is the most significant factor, of several other possible factors, affecting B)

- There must be a set of defined empirical elements and relations corresponding to the model:
  (1) ‘a’ is an empirical manifestation of 'A' (i.e. ‘a’ is the best, or most suitable, measure of 'A' within the context of the model)
  (2) ‘b’ is an empirical manifestation of 'B' (i.e. ‘b’ is the best, or most suitable, measure of 'B' within the context of the model)

The example in figure 4:3 is a causal model. Other kinds of relations, such as mean-end or order, are of course possible.

4.3 Our approach of using modeling to define test design.

In our approach to use modeling of test design we have normally used two major types of models. The first type is a business model which describes the organization, technology or process to be used as a framework for the experimentation. The second type is a measurement model which defines the relevant factors to be measured, i.e. an operational definition of the actual research question.

4.3.1 Business modeling

The business model is intended to give an outline of participants and systems included in the experimentation, the relation between them and the systems input and output. As shown in figure 4.3 it is of central importance that the manifestations, operationalisation, of the models elements are defined.

The main purpose is to specify the context in which the experimentation is conducted. The example illustrated in figure 4:4 can, despite that the underlying process might be extremely complex, give a fairly simple outline of the structure of the overall process.

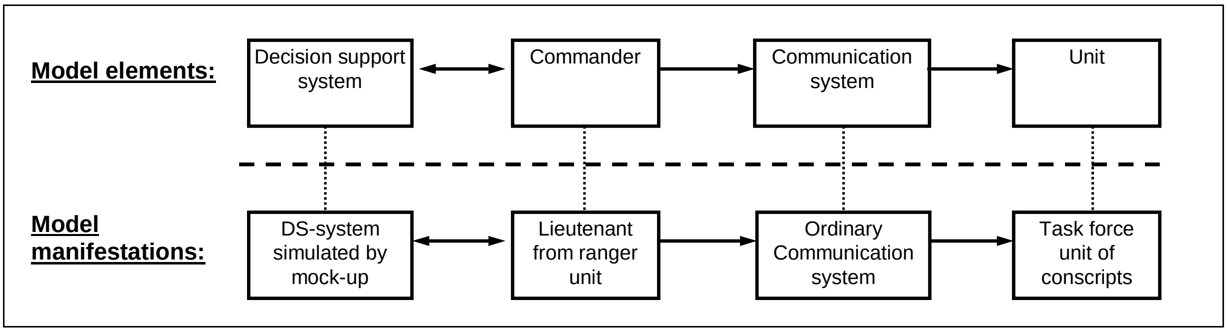


Figure 4:4. An example of a business model

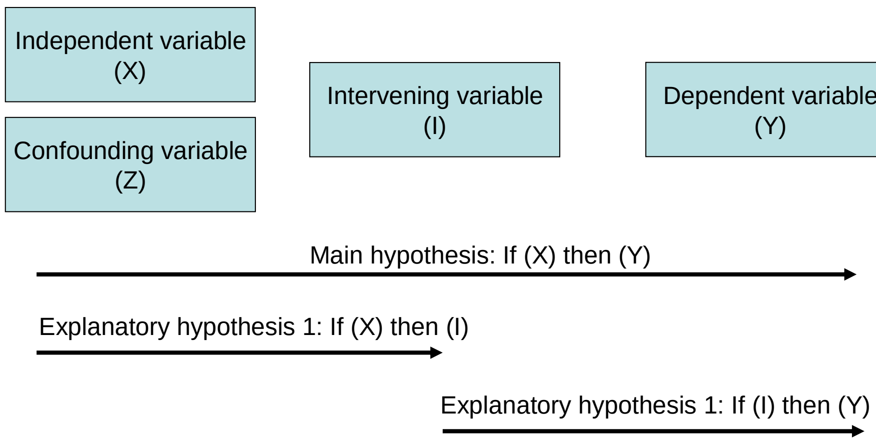
Suppose that the model in figure 4.4 has been defined to describe the fundamental activities in using a new decision support system. An experiment derived from the model in figure 4.4 should at least consist of a commander, a decision support system, a communication system and a task force unit. The business model thus becomes a communicative aid when the experimental situation is created. The decision support system, indicated in figure 4:4, might be in an early stage of a developmental project. When this is the case it might be necessary to create a simulated system by using some sort of mock-up, for example paper models of user interface, physical mock-ups of equipment or more advanced computer based simulations.

Preferably, the test environment should as much as possible be represented by real systems and real personnel. In the case illustrated in figure 4:4, the communication system can be the one used on a day to day basis and the task force unit might consist of conscripts. As commander, the use of one whom might use the system in the future might be preferable. An experimental situation might of course include other essential aspects not described in the model, for example a tactical situation or a staff supporting the commander. However, in this particular case this was not considered as important for experimentation. What to include in the business model, and thereby in the experimentation or simulation, is practically more or less arbitrary. A more thorough discussion on test environments is found in chapter 5.

4.3.2 Measurement modeling

A business model is normally not sufficient as foundation to specify experimental design. For example, a study on command and control might focus on the expected improvement of organizational performance caused by the introduction of a new decision support system. Presumably, it is possible to regard these more or less explicitly stated expectations as hypotheses. What is the client's "best guess" of what will happen?

To formalize these expectations, a measurement model must be defined. In the measurement model, relevant factors of the actual research question are defined. The purpose of this model is to define hypotheses and/or research questions. As in the case of the business model it is essential that the manifestation, operationalisation, of the elements of the model is defined. The definition of how the elements in the model are made manifest constitutes the basis for the construction of instruments of measurement. The measurement model is basically a hypothesis tree with a number of generic elements (Figure 4:5).



Independent variable (X). The independent variable is the variable or factor whose levels are selected in order to determine the effect of that variable on a dependent variable(s). Note that variable doesn't necessarily mean a quantitative variable such as hours of sleep fatigue. It is possible to have different organizational or technological solutions compared. One typical independent variable is a ”new” work method compared with an ”old”.

Dependent variable (Y). The dependent variable is the particular measures used to reflect the effect(s) of the independent variable. Which differences on performance will the "new" organizational or technological solution impose? As a result of the variation between conditions the dependent variables and are expected to vary.

Confounding variable (Z). It is normally not possible to create to exactly equivalent conditions in all aspects in experimentation. Any variable other than the independent variable that might affect the dependent variable and consequently confound the result must be controlled by being eliminated, minimized, held constant, explicitly randomized or by using statistical control.

Intervening variable (I). A reasonable assumption is that several processes might be of importance between the stimuli and response. Intervening variables reflects a hypothesized process to explain the relations between antecedent conditions (i.e. independent variable) and consequence events (i.e. dependent variables). (conditions and system outcome). A number of intervening variables might help explain why differences in effects between conditions occur.

Hypotheses. The model also presents the hypotheses to be tested. The main hypothesis is the (proposed) relationship between the independent and dependent variables. It is also possible to define explanatory hypotheses which describe the assumed relations between independent variable and the intervening variable and the intervening variables and the dependent variables.

Figure 4:5. A generic measurement model

Several variables of each type might be included in the experiment design. In experimentation on command and control the dependent variable is typically a mix of several indicators that are supposed to give a "performance profile" of the studied organization or system. In the same way several possible intervening variables might be defined in order to explain variance in the dependent variable.

An example of a measurement model taken from a study performed in the autumn of 2004 (Wikberg, Andersson, Berggren, Hedström, Lindoff, Rencrantz, Thorstensson, & Holmström, 2004) is shown in figure 4:6. In this example two different setting of task environments were compared. One is a real physical environment. The other is an environment based on computer simulation using commercial PC-software. The initial research problem was to examine whether it was possible to use simulated environments on the soldier level in combination with real C2-systems to create an experimental setting to study command and control. The measurement model defines two major dependent variables: task force performance and command post performance that might vary between settings. Each of these variables consists of a set of sub-variables. Some confounding variables were also considered such as 'weather' and 'task force skill level' in analyzing the results. Consequently, these factors have also to be monitored. Finally some intervening variables where identified as possible explanations to any variation in the dependent variable. For example, is it differences in task dynamics or differences in communication pattern that explain any difference in performance between settings? This particular study is described later in this chapter in section 4.4.2. as case 4b.

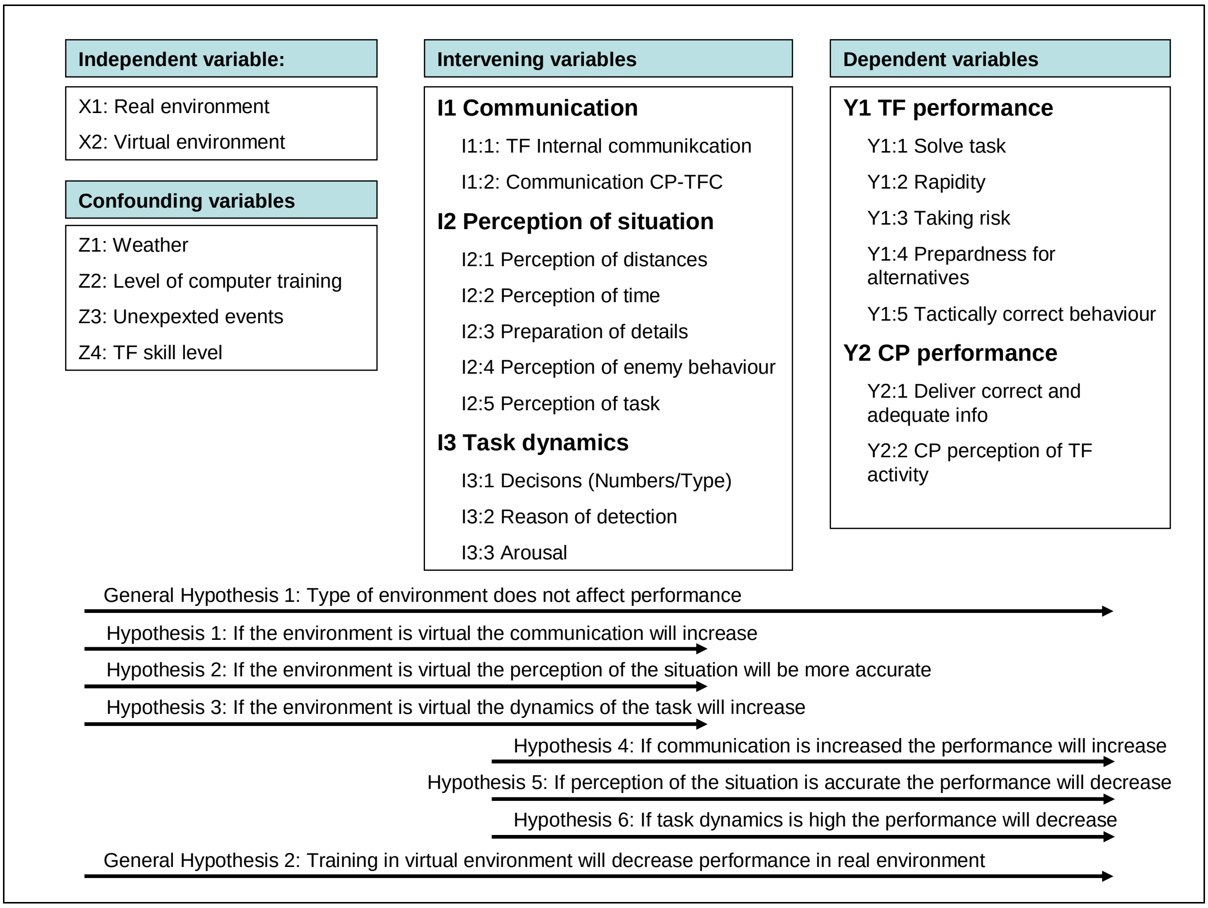


Figure 4:6. An example of a measurement model.

Data on each of the variables were collected from a variety of data sources. Data collection is somewhat beyond the scope of this chapter. In chapter 6 we will present more thoroughly our approach to data collection. However, again it must be emphasized that definition of instruments of measurement is an integrated part of the measurement modeling.

A final note on measure modeling is the importance of level of scale. Scale type is important because it defines which analyses are possible, or at least permitted, to carry out. As will be discussed in chapter 6 on data collection, the operation of measurement is a relation-preserving function which should translate an empirical structure into an abstract structure. The scale levels; nominal, ordinal, interval and ratio scales, defines the "precision" of measurement and thus also the precision of the translation. For example, it is not possible to conclude whether something is twice as good based on a measurement using an ordinal scale as an ordinal scale represents order between objects (and not interval or ratio between objects). This issue is briefly introduced in section 6.2.1. To anyone familiar with inferential statistics this is nothing new as it is a part of the basic of statistics. The pragmatic problem is rather to explain the restrictions of different measures to clients in the modeling phase. For example, the precision of measures might not be good enough to tell how long it takes to make a decision for a division staff.

4.4 Illustrative cases

Two cases are presented in order to illustrate the use of modeling to define experimentation design.

The first case, 4A, describes modeling of an experimental staff exercise. The modeling includes a business model with definition of exercise set up and a measurement model with variables to measure and corresponding hypotheses. The purpose of the study was to compare two possible information structures.

The second case, 4B, describes modeling of an explorative study. The purpose was to come up with a proposition of a new decision making model. The model was used to manage the discussion during two war gaming exercises. The aim was to generate hypotheses about future decision making processes.

4.4.1 Case 4A: Measurement Model on effect of information structure on C2.

A study was performed in the context of the Swedish Armed Forces' demonstrator program which aims at reshaping and adapting the defence from an invasion focused to a flexible network based defence (Swedish Armed Forces, 2002). The study was performed on two occasions and focused on how the C2 process is influenced by which information structure is chosen. The first study was performed in autumn of 2003, DEMO 03 H, and a replicate was conducted during May 2004, DEMO 04 V.

The basis of the experimental design were outlined in a business model and a measurement model in a series of group modeling sessions with participants from the Armed Forces and FOI. The starting point for the modeling session was a hypothetical chain of effects defined by LedSystM, the project responsible for development of command and control (C2) methods in the armed forces transformation process. The assumption was that "role based but shared situational information" creates an "a mutual situational awareness". The more mutual a situational awareness is the less effort is needed to achieve decentralized coordinated action. Thus, a mutual situational awareness in turn facilitates coordination and even self synchronization. Better opportunities of coordination will make it possible to achieve what is called "command and control superiority". The hypothetical chain of effects is shown in figure 4:7.

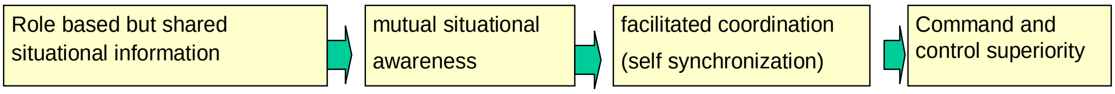


Figure 4:7. The hypothetical chain of effects used to model Demo 03 H and Demo 04 V.

The basic structure of this assumed relationship was modeled into a business model of the activities of a command and control system and into a measurement model describing the variables and hypotheses implied by the assumed the relationship. A series of modeling sessions were conducted with representatives from LedSystM and FOI. The modeling was documented in graphic form on a whiteboard under the supervision of an analyst from FOI.

The business model defines how the processes which were included in the hypothetical chain of effects should be realized in an exercise organization and which restrictions to put on the scenario according to the principles illustrated in figure 4:8.

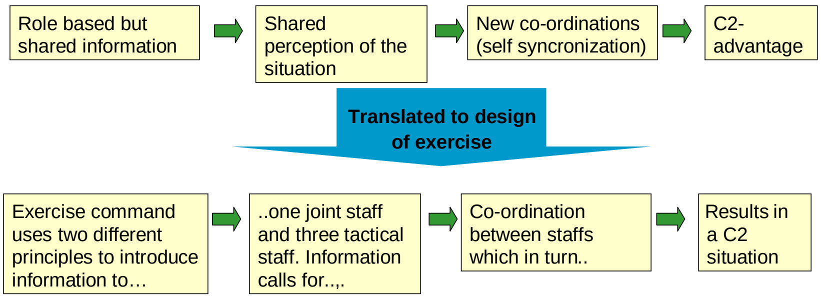


Figure 4:8. Translation of hypothetical chain of effects to experimental setting.

The business model formed the basis to organize the actual exercise setup. The test was performed as a series of staff exercises conducted over a period of three days. In one setting, the information structure was based on the principle that different staffs only had direct access to information based on their predefined organizational roles. Any additional information has to be requested from other staffs or actors. In the other setting, all actors and staffs had direct access to all information in the system. The final business model is shown in figure 4:9.

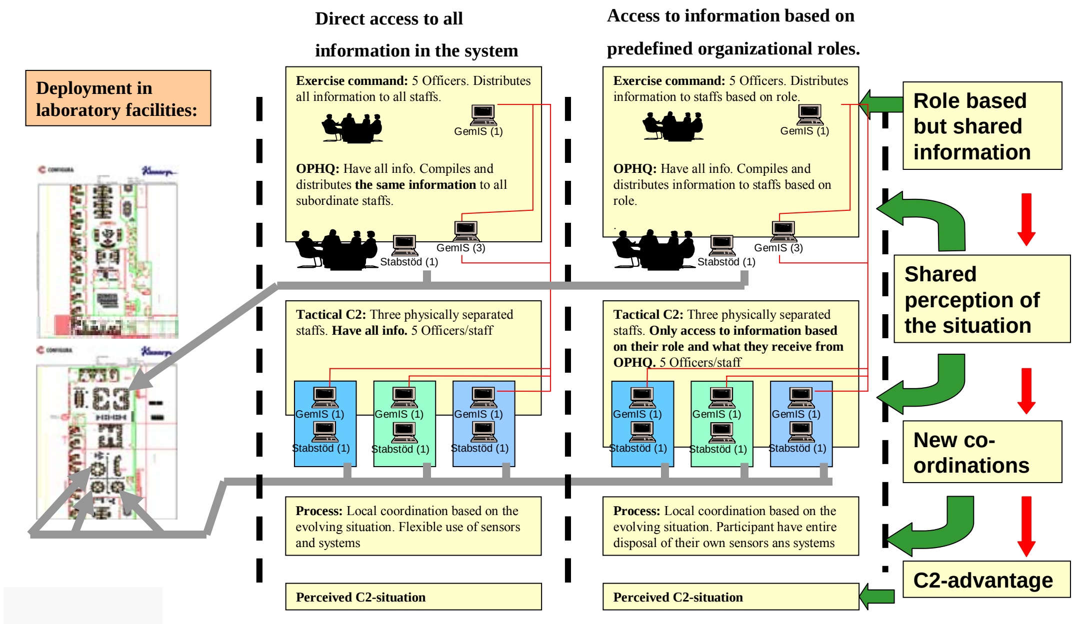


Figure 4:9. Business model of DEMO 03 H and DEMO 04 V

The measurement model was based on the same hypothetical cause and effect relationship between information structure and C2 process as the business model. As described in the business model in figure 4:9 two possible solutions for access to information were implemented limited and unlimited access to information. In the limited condition the lower staffs only had access to their own C2-system, i.e. lower staffs had access to different information databases and external information was only accessible to all staffs after it had been approved and manually implemented to the system. All staffs had to seek information via ISMARK to get the simulated picture. In the unlimited condition all the lower staffs had access to a mutual information database and had full access to the simulated picture via the Federation. These two solutions were treated as the independent variable in the test. Consequently, the research manipulation consisted of a variation in ‘information structure’, X1, between these settings.

The business model also defined relevant variables to measure. A graphical model illustrating the resulting measurement model is shown in figure 4:10. The model describes how information structures **(X1)** are related to a number of relevant factors in the C2 process (**Y1-Y4:** Perceived performance, Perceived situation, Perceived stress and critical activities). Intervening variables (I1-I6) refers to internal processes, perceived situation and coordination. The model also defines some confounding variables (**Z1-Z4:** external processes and characteristics of the participants).

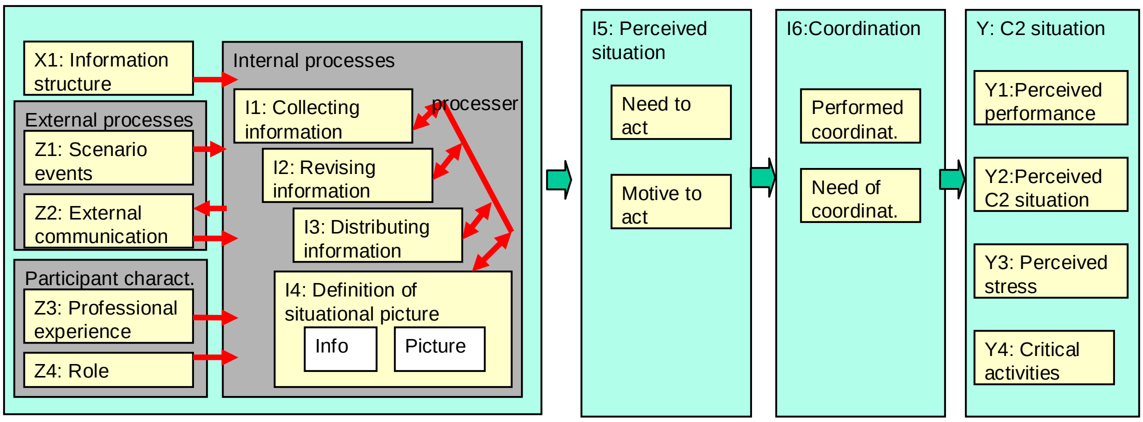


Figure 4:10. Measurement model of Demo 03 H and Demo 04 V

Based on the measurement model some hypotheses are possible to derive. The main hypothesis is the (proposed) relationship between the ‘information structure’ and ‘command and control situation’. Using the intervening variables a number of explanatory hypotheses were also possible to define. In this case the effect of the variation between settings was tested with nineteen different hypotheses.

An important component of the measurement model is instruments of measurements. The ambition was to use several different methods to measure each variable (triangulation). An example of a hypothesis from the study is: *“If access to information is limited **(X)** then the level of performed contacts between units to discuss coordination **(I6)** will decrease”*. This particular hypothesis, which was rejected, was measured with three different kinds of instruments of measurements:

1) Digital questionnaires, during and after the exercise, where the participants estimated their need of coordination and described which contacts they actually had taken to discuss coordination.
2) Observers documenting which contacts the units had actually taken in order to discuss coordination.

3) Logging of voice and text communication between units. All communication was then classified on a number of factors. One of the classes was if the communication included discussions on coordination between units.

The result of the study indicates that the participants were less inclined to question the reliability of the information when they had access to all information. The results further suggest that the participants were more satisfied with available information in the setting with access to all information and the situational picture they could create based on it. Accordingly, the participants also considered themselves to have a better situation awareness and better chance to exert command and control. Nevertheless, there was a larger disagreement between individuals within the same tactical staff regarding which factors that were the most central in the prevailing situation.

The conclusion from the study was that the results support a C2 system where all the actors have full access to all information available on the system databases. The somewhat confusing results of a larger discrepancy between individuals when the information is considered as being of better quality is judged to be a natural effect of a richer and more realistic assessment of the situation.

4.4.2. Case 4B: Modeling of future decision processes in C2

In a study (Wikberg, Strangert, Adolfsson, Holmström, 2005) performed in the fall of 2002, modeling was used to define alternative procedures for military decision making. The rationale for this study was that the realm of decision making in command and control today is wider than before and there is a general opinion that the traditional process of military planning is too slow. In addition, developing the new Swedish command and control systems, the ambition is to incorporate new perspectives on decision processes such as “self-synchronization” (Brehmer, 2002).

The study was conducted in two steps. The purpose of the initial phase was to define a basic theoretical model of decision making based on six theories of decision making from psychology and organization theory. Together with subject matter experts, the theories were compiled into one comprehensive model outlined like a ‘manual’ for a staff procedure. In the second phase, the model was used as a ’script’ to organize the discussion in two war gaming sessions.

Step 1 – Construction of a base for war gaming and for the description of the scenario. In the modeling sessions during the first step four persons participated: two from FOI, one from the Swedish Armed Forces and one from the Work and Organizational Psychology Unit at Umeå University. The participants contributed to the modeling with their experiences from each area of expertise. The starting point was four theoretical decision models from behavioral science and two from organization theory.

Theoretical perspectives on decision making in psychology

- Decision making as conditioned response on environment (Skinner, 1953, 1968)
- Decision making as rational choice between defined alternatives (Simon, 1957, Miller & Starr, 1967)
- Decision making as dynamic process control (Ashby, 1956, Brehmer & Svenmarck, 1994)
- Decision making as intuitive judgment based on expertise and recognition of contextual factors (Klein, 1993)

Theoretical perspectives on decision making in organization theory

- Decision making as negotiation in order to achieve acceptable solutions (Cyert & March, 1963)

- Decision making as organic streams of events (Cohen, March & Olsen, 1972)

The participants were gathered in four occasions. In the working process sheets of paper, representing the model components, were used. With the help of these sheets of paper a new model lay-out was created, edited and transformed into a digital format. An important part of the modeling was that the original features of the "original" model never were lost. Each and every model illustrates decision making in its own way and brought out different aspects of the decision maker, the task and the context. Great importance was attached to keeping each model characteristics and letting these characteristics leave its mark on the common model. It must be emphasized that the purpose never was to create a normative model for future decision making but to create a final product which could be considered a "script" for the successive parts of the study. The general features of this model are shown in figure 4:11. Note that figure 4:11 only shows the general features of the model. The model in detail is shown in figure 4:12.

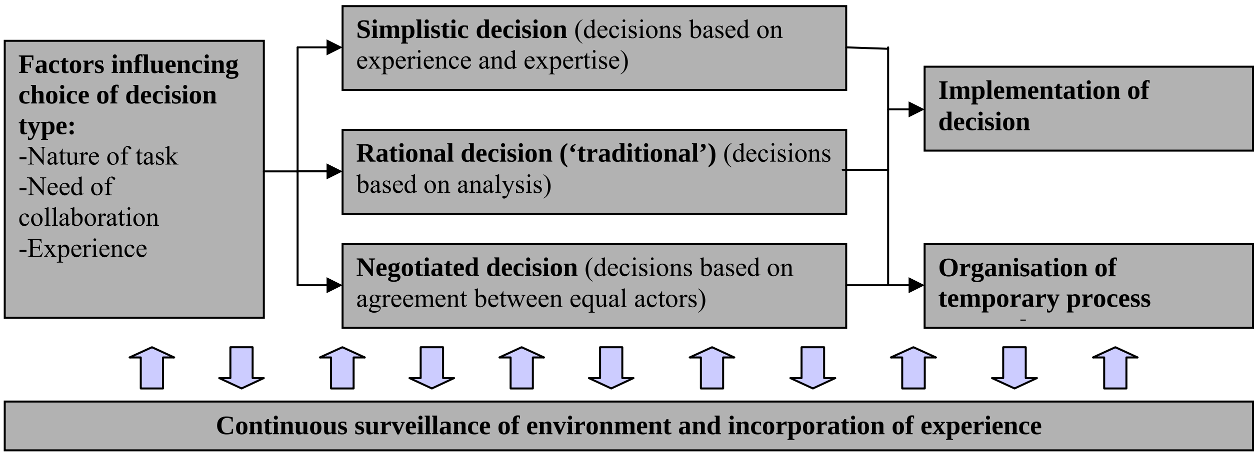


Figure 4:11. The general features of the suggested decision making model described in Wikberg, Adolfsson, Strangert, & Holmström (2005).

Step 2 – War gaming. In the second step of the study, the model was used as a script in two war gaming exercises focusing on command and control. Participants of the war gaming sessions were representatives from the Defence Forces and different departments of FOI. All participants were appointed roles which corresponded with their competence. The roles were either active, when the participant acted in the course of events, or passive, where the participant had a supporting role as domain expert (for example an expert in technology concerning military information and communication).

In the war game a tactical situation were presented. During the exercises participants described how they would act in the simulated situation. The process steps defined in the model were successively discussed based on the participants' different areas of expertise. The discussion with and amongst the participants concerning the evolving description of the scenario were documented successively by the war gaming manager. For this purpose a lap-top computer with a projector enhanced screen were used so that the participants could follow and comment what was noted. This documentation later became the foundation for the description of the course of events. The conclusion from the syudy was that further development of formalized decision methods are judged to be an important issue.. Focus of such an development should be adaptation to international operations.

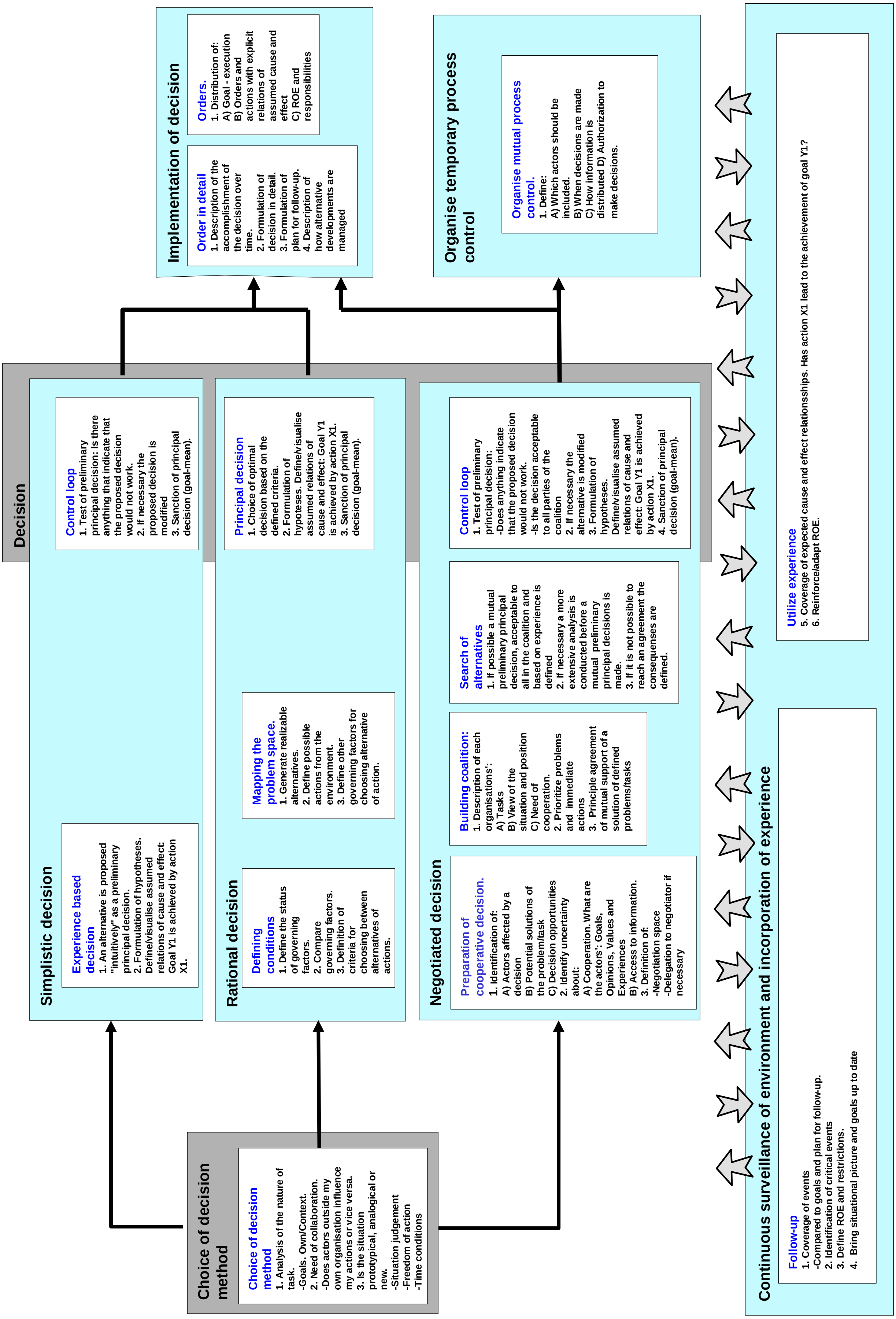


Figure 4:12. A model of decision processes (Wikberg, Strangert, Adolfsson, Holmström, 2005).

5 Test environments

5.1 Background

Empirical tests are always realized in some kind of practical context, i.e. a test environment. The term ‘test environment’ comprise not only the actual research setting, e.g. military exercises or laboratory environments, but also practical and theoretical restrictions due to the character of the experiment. For example, these restrictions can include limiting participants’ freedom of action in some ways or their access to information.

In designing a test environment, there are numerous factors that affect the process. Some of these factors can be controlled by the researchers or by the clients, while others cannot be regulated or adjusted. The research problem itself, the extent to which the problem area has been explored so far, as well as the resources available (such as money, technology and personnel) are examples of factors that are important in designing a test environment.

As mentioned in chapter 3, the objective is almost always to create a highly realistic test environment. Yet it is sometimes impractical, inappropriate or even impossible, to study certain phenomena in their natural setting, due to the complexity of the phenomena as well as the setting itself. Part of the solution is to adapt the setting in order to make things manageable, so that the phenomena of interest can be investigated in a systematic manner. This adaptation can involve creating simulations of entire (or parts of) processes, systems, exercise environments, courses of events or organizational aspects etc.

These issues have to be considered in the initial problem analysis conducted during modeling of experiment design as described in chapter 3. The following section is a brief summary of some theoretical aspects of test environment. Readers primarily interested in practical applications could move on to section 5.3.

5.2 Simulation and control in experimental design

Simulation is a tool to make different aspects of command and control practically testable. Dawson (1962, p.3) define simulation as ”…the construction and manipulation of an *operating model* - a physical or symbolic representation of all or some aspects of a process. An operating model in this case means that the simulation model can replicate the characteristics and the behaviors of the real system over time (Schultz & Sullivan, 1972). A simulation doesn’t necessarily need to be physically concrete. A mental simulation, where participants imagine a real situation, might also be included in the definition. Consequently, Harré, (2002, p. 54) defines simulation as a real or imaginary representation of an empirical system. Viewing simulations as representations means that it has the same theoretical foundation as modeling, presented in section 4.2.

Gist, Hopper & Daniels (1998) suggest two dimensions, test setting and research strategy, which together can be used to describe a ”space of experimental settings” in terms of simulation (figure 5:1).

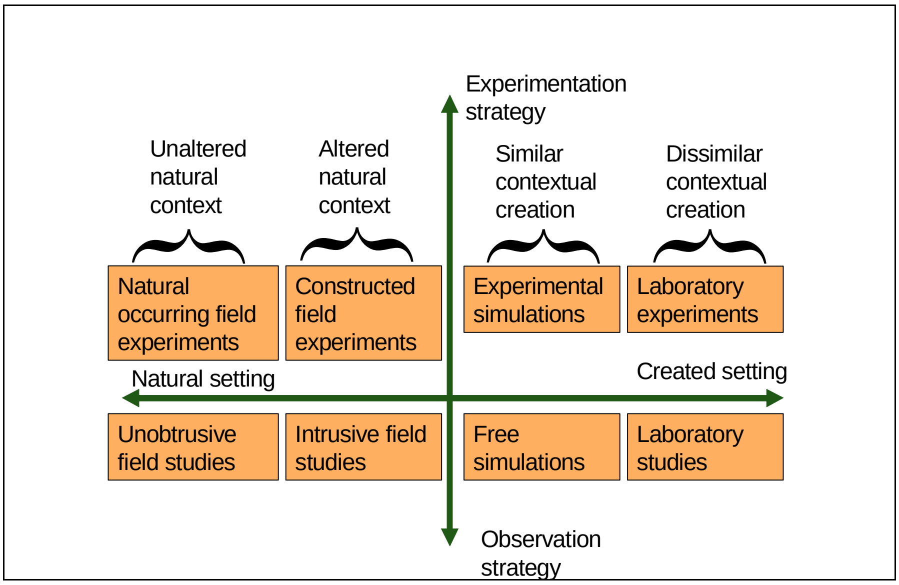


Figure 5:1. Eight types of games or organizational simulations along the two dimensions of research setting and research strategy (Gist et al., 1998).

The first dimension concerns the *setting*, i.e. the extent to which an experimental setting uses a ‘real’ environment or to which extent the setting is created. This dimension includes all possible research settings, from laboratories to actual field settings. In experimentation on military command and control, it spans from real, large scale operations to highly formalized laboratory experiments. In between these extremes there are exercises and war gaming with different degree of realism and simulation.

The second dimension deals with *research strategy*, i.e. the degree to which the researcher actively exerts control on the simulation. This dimension covers the spectrum of research strategies from the strict experiment with active manipulation of variables to passive observation studies.

As shown in figure 5.1 it is possible to identify eight types of games or organizational simulations. Gist et al. (1998) points out that in debating the merits of laboratory versus field research, the “middle ground”, i.e. organizational simulations in altered natural contexts and similar contextual settings, has been somewhat neglected in organizational literature. These research approaches permits certain experimental rigor which is difficult to obtain in the field, and offers contextual relevance, since measures are taken to simulate contextual elements of the natural, or real, setting.

We find this also true when it comes to studying command and control. Therefore, our work has focused on the “middle ground” of their model (figure 5.1). We have tried to find ways to maximize experimental control in complex simulations.

5.3 Our approach to developing test environments

Our experiences spans over a wide variety of settings and research strategies. We have conducted studies of command and control processes during field exercises and in C2- laboratories. Furthermore we have, in the different types of settings, used different degree of experimental control. An outline of some studies, classified in accordance with Gist et als’ model, is presented in figure 5:2. Some of these studies have been conducted within the scope of the project while some have been conducted in cooperation with other projects.

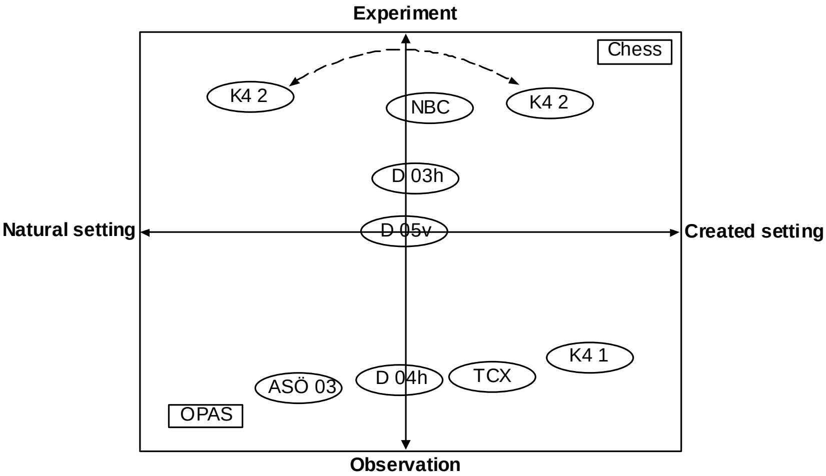


Figure 5:2. An overview of some of the conducted experiments related to the project. The studies is classified in accordance with the dimensions shown in figure 5.1

The studies listed in the figure have in general been conducted in a setting between "real" environments and laboratory setting. Some of the studies have been conducted as "staff-exercises", a context we have considered to be in between the extremes of the setting dimension. These staff exercises have in turn varied in level of simulation. Studies have also been conducted during field exercises, a setting we have considered to be nearer a real setting than staff exercises. Finally, computer simulated environments have also been used. They represent a higher degree of created setting compared to staff exercises.

Level of experimental control has also varied between studies. Often, it has only been possible to conduct the study using an observational strategy due to practical circumstances. However, in many cases the studies have been conducted exerting experimental control. As discussed earlier, the level of control has inevitable varied between studies. In some cases we have used a mix of observational and experimental strategy.

Note that the research traditions of action research and multi factor case studies (se section 3.3) are applicable in almost the whole "space of research settings" defined in figure 5:1 and 5:2. However, our experience is that the use of action research is often the most applicable approach in large scale experimentation when an experimental strategy is to be used. Multi factor case studies with an experimental strategy have the disadvantage that large scale exercises often have to be adjusted in relation to the planned execution. Changes in scenario and procedures might of course corrupt the experimental design. To adjust to this problem we recommend a combination of data collection in larger settings and smaller case studies as described in case 7A. If an observational strategy is used both case studies and action research is applicable.

Figure 5:2 presents the following experiments:

NBC. An experiment conducted at the Swedish NBC Defence Centre during an exercise with the national NBC protection task force. In the experiment the task force had to manage a situation with chemical warfare agents discharged over a major city during a staff exercise. Two conditions were compared with an experimental strategy. In the first condition the existing NBC C2 system was used. In the other condition a simulation of a future C2 system was used. This case is described in chapter 7 as case 7.B.

K4 1. An experiment, conducted at the Norrlands Dragoon Regiment, explored the possibility to use commercial computer games as simulations in command and control exercises. An observational strategy was used during the test. This case is described later in this chapter as case 5.A.

K4 2. In a sequel to the study described above, also conducted at the Norrlands Dragoon Regiment, two different task environments, field exercises and computer simulations, were compared. The comparison, which was based on an experimental strategy, focused on how the same task was solved in these different environments? This case is described later in this chapter as case 5.B.

Demo 03 H. A study performed in the context of the Swedish Armed Forces' demonstrator program. The study was performed on two occasions and focused on how the C2 process is influenced by information structure. Two conditions where compared using basically an experimental strategy. This case is described in chapter 4 as case 4.A.

Demo 04 H. An exercise held in the context of the Swedish Armed Forces' demonstrator program. The focus of the experiment was to further develop a procedure for planning under time pressure. The study was conducted as a war gaming exercise using an observational strategy.

Demo 05 V. An exercise held in the context of the Swedish Armed Forces' demonstrator program. The focus of the experiment was to further develop a procedure for planning under time pressure. The study was conducted as a staff exercise. A series of exercise runs were used for data collection. Some of these runs were studied using an observational strategy while others were studied with an experimental strategy. This case is described in chapter 7 as case 7.A.

TCX. A study carried out at the Singapore Center for Military Experimentation (SCME) with the purpose of exploring the effects of a decision model combining a naturalistic planning and decision-making model called the Knowledge Battle Procedure (KBP) and a C2 System for distributed planning called MissionMate (MM) with TeamSight (TS). The study was done as a part of the overall Swedish Armed Forces (SAF) and Singapore Armed Forces collaboration framework. The study was conducted as a formal staff exercise and an observational strategy was used.

ASÖ 03. During the Swedish Army's maneuver in the spring of 2003 a pilot study was performed in order to examine how distribution of responsibility was perceived in the chain of command of an army division. An observational strategy was used and the study is published in Andersson, Lindoff & Wikberg (2003).

5.4 Illustrative cases

Two cases are presented in order to illustrate the use of test environments as we have used it.

The first case, 5.A., focused on exploring the possibilities of using commercial game software in experimental simulation exercises. In this particular exercise, the aim was to evaluate the effects of real-time information access on command and control. The results from this study showed that the use of commercial game software for this purpose is possible.

The second case, 5.B., was a sequel to the previously mentioned case. One condition for using commercial PC-games for evaluation and training is that the behavior of units and commanders in the virtual and physical environment must have high correspondence. This study investigated this correspondence. A virtual three-dimensional copy of one square kilometer of the regiments exercise range was created and integrated in a commercial PC-game. Ranger task forces, supported by a rear command post, accomplished the same mission both in real and virtual environment and the performance of the task forces' and the commander's executions of the mission in the two different environments were compared.

5.4.1 Case 5A: Commercial software in a ranger command and control exercise

A computer game based simulation exercise was conducted at Norrland's Dragoon Regiment in May 2003. Commanders' use of real-time information from Unmanned Aerial Vehicles (UAV) was studied using commercial software to simulate the tactical context of command and control. Will commanders with access to detailed real time information interfere in operations on lower level, thereby loosing their overall perspective? The aim of the study was to explore the possibilities of using commercial game software in an exercise evaluating command and control.

At the regiment, ranger battalions are trained for combat and reconnaissance missions in a large variety of environments. The battalion command, rear command post, is dimensioned to lead ranger squads in several different directions and from a substantial distance. The task of the battalion command is to deliver qualified intelligence material to the commander in charge of the mission in the area. The battalion command post can, if needed, be deployed in close connection to the area of the mission. The principle of command and control used within the ranger battalion today is founded on communication over substantial distances using a High Frequency (HF) radio and a PC-based terminal for communicating data (PC DART). PC DART is a message based type of communication. Each task force has a PC DART client with software which enables the possibility to write, send and receive messages. In principle PC DART has the same function as a regular e-mail client.

Information transfer within a ranger battalion is limited compared to other military units since a ranger task force commander can not bring extensive equipment and analysis tools, i.e. the ranger must be able to carry his own combat load. Consequently, vast amounts of information from sensors such as UAVs, satellites and integrated helmet and display sight must be analyzed by a battalion staff deployed at distance from the target area. If the ranger battalion is equipped with, or supported by, such sensor systems, the rear command post will have better access to real-time information about the situation in the target zone than the task force commander. A number of hypotheses concerning effects of the staff's access to real time information on command and control methods were tested.

With support from military personnel from K4 a business model of command and control of ranger units using real time information was defined (Figure 5:3).

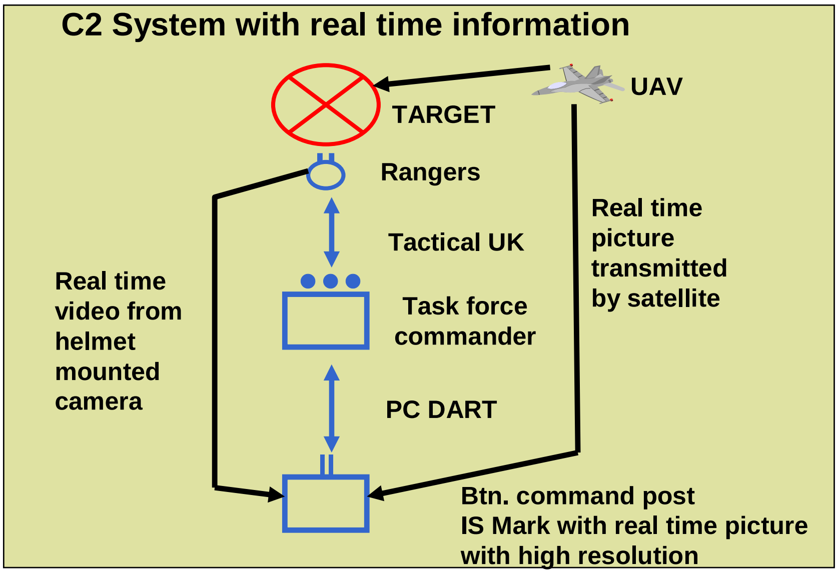


Figure 5:3. A business model of the command and control organization when real time information from a UAV is available. In the present case RADIO 180 were used as a substitute for PC DART as communication tool between task force commander and battalion commander.

The participants were four career officers and 24 conscripts from the ranger battalion. There were also two technical officers from the Swedish Army Technical School (ATS). The participants were divided into one ranger battalion staff and three task forces. Each of the three task forces consisted of one career officer and six to nine conscripts. The staff was composed of three career officers and two soldiers. The staff and the three task forces were placed in separate rooms.

The tasks were carried out in virtual commercial PC game environments. From one of the ranger units the battalion commander had access to real-time information from an UAV and from helmet assembled cameras. There was no such information from the other two ranger units. The outline of the exercise is illustrated in figure 5:4.

Figure 5:4. The command and control conditions represented in three different virtual environments

Three local networks (LAN) were organized. The staff used their regular tactical information system "IS MARK" while the task forces performed their missions in three different virtual environments. A commercial, non-modified, game was used in each LAN. The PC game enabled the soldiers to integrate their actions through their personal terminals. In two of the LAN the computer generated opponent provided by the software was used as enemy. In the third LAN the enemy was controlled by other soldiers.

Communication between the staff and the three task forces, i.e. LAN 1-3, were made through the ranger battalion's ordinary communication device PC DART. The battalion staff was able to communicate with all the ranger units. The ranger units on the other hand were limited to communication with the battalion staff and not with each other.

The tactical situation was followed up by the staff and registered in IS MARK on maps generated from the game. The different LAN was structured according to the following principles:

LAN 1 (Task force 1) was made out of eight computers and the game "Delta Force: Task Force Dagger" from Novalogic. The force was assisted by an UAV which circulated the area of interest maneuvered by the staff. The staff had a screen showing the image from the UAV. The computer screen of the commander of the task force was filmed and the image from the camera was directly available to the staff. The artificial intelligence served as the enemies

LAN 2 (Task force 2) used six Xboxes from Microsoft and the game "Ghost Recon" from Ubisoft. The artificial intelligence served as the enemies.

LAN 3 (Task force 3) consisted of eight computers where the game "Rogue Spear" from Red Storm Entertainment was used. The enemy was two additional soldiers.

The staff was in charge of the units and worked in accordance with their ordinary procedures. Before the exercise, the participants practiced in a tutorial generated by the game in order to get necessary basic gaming skills. Communication and tactic were also trained as well as handling of PC DART. Different scenarios were used for the pre training and the ordinary experimental exercise. Task force 2 and 3 executed the operations three times and stand-by force 1 did their operation two times. After each run, debriefing was carried out in accordance with normal procedures. Data was collected using observers, questionnaires and technical registration of communication. Afterwards, raw data and preliminary results generated from the data analysis was presented and discussed with all participants of the exercise during an After Action Review session.

The results indicate that access to real time information at higher levels of command doesn't necessarily lead to a change in command method from a mission oriented command tactics to a command guidance tactics. However, it seems that the battalion staff prioritized the unit with real time information on the behalf of the other units. The results underline the importance of systems to administer real time information as a vital component for the command and control of future ranger units. The assessment is that there are large possibilities to develop methods to use commercial game software for exercises, tests and tactical applications.

5.4.2 Case 5B: Simulated environments as experimental settings

A sequel to the study described above was carried out in the fall of 2004. The study, also conducted at the Norrland's Dragoon Regiment, investigated the correspondence between behavior of units and commanders in a virtual and a physical environment. A virtual three-dimensional copy of one square kilometer of the regiments exercise range was created and integrated in a commercial PC-game. Ranger task forces, supported by a rear command post, accomplished the same mission both in real and virtual environment. In both environments the rear command post had access to real-time information from a simulated UAV. The task forces' and the commander's execution of the mission in the two different environments were compared according to mission success, communication, situation awareness and the dynamic of the task. Data was gathered using observers, questionnaires and registration of radio communication. The same business model (Figure 5.3.) was used as in the first study to define research settings. The realization of the model in the two different settings is presented in Table 5:1.

Table 5:1. *Description of the elements in the business model and how these were realized in both settings.*

Model elements	Description
Goal	Objects, activities or area which constitutes the task force's mission. **Realized by:** The objective was to demolish a communication pylon. In the physical setting the pylon was located about 50 meters south of a buildings present in the mission area. The pylon was replicated in the virtual setting.
Soldiers	Rangers part of the task force with the task to strike at the objective. **Realized by:** Conscripts from the 41. Ranger platoon. In the physical setting weapons rigged for blank ammunition and blind explosives were used. The soldiers were equipped with equipment to indicate gun fire using laser indicators. In the virtual setting each soldier had access to a computer connected to the LAN and each of them controlled one player.
Task commander	The individual commanding the mission in the field. In missions concerning few units the task commander normally is in close connection to the soldiers at the objective. In missions with several units the location most suitable for the coordination of the mission is chosen. **Realized by:** Consisted of conscripts from 41. Ranger platoon. These were using the same equipment as the other soldiers in both settings.
Tactical radio	Within the task force the communication between units is based on a radio UHF system. **Realized by:** The ordinary radio system was used in both settings.
Rear command post	The battalion command located at a rear command post has the task of supporting the task commander. **Realized by:** Officers from the regiment with the rank of captains. That position was in accordance with the position they would have during a real operation. The rear command post in the virtual setting consisted of one officer located in the facility used as rear command post during the regiment's regular exercises. The rear command post in the physical setting was located in a hut not far from the physical mission area. Both rear command posts had same access to command and control support systems.
PC DART	A long range communication system based on HF radio and a PC-based terminal for communicating data. **Realized by:** In communication between mission and rear command post the Radio 180 was used in both the real and virtual setting.
UAV/Real time image via satellite	A sensor. In this particular case defined as a UAV able to register activities in the mission area in real time. **Realized by:**The real time information from the UAV in the virtual setting was an existing functionality of the used PC-game. The UAV-picture was presented on a screen on a terminal connected to the local network in the rear command post. The UAV could be controlled by the task force commander. The picture was not available to the task force. In the physical setting a remote controlled web-camera was mounted on a sky-lift placed in the mission area. In accordance with the virtual setting the commander in the physical setting could control the view of the simulated UAV. Furthermore the picture from the web-camera was not available to the task force. The position and altitude of the simulated UAV-function was as equal as possible in both settings.

Below, examples of UAV-pictures from the virtual (Figure 5:5) and the physical (Figure 5:6) setting are presented.

Figure 5:5. Example of a UAV-picture generated in the virtual setting

Figure 5:6. Example of a UAV-picture generated by means of a web-camera in the real setting

Results from the study reveals that the task forces' performance and behavior was more or less the same in the different conditions. Simulated environments are therefore sufficient for training and experiments of this kind of mission. However, it should be noted that communication within task forces was more frequent in the virtual environment. Furthermore, the communication between commander and task force was more frequent in real environment. A more through discussion of the result is presented in Wikberg, Andersson, Berggren, Hedström, Lindoff, Rencrantz, Thorstensson, & Holmström (2004).

6 Data collection

6.1 Background.

Experimentation always includes some kind of data collection as data is the information gathered which is assumed to represent the characteristic of the studied phenomena. As we have presented in chapter 3 we have focused our work on making the data collection process as efficient as possible.

Our definition of data collection, or more specifically, measurement is an operation to assign quantitative or qualitative values to empirical phenomena. Methods for data collection are major components of experiment design. Experimentation on complex systems normally calls for several different methods and data sources (Strangert, 2005). From a technical point of view the operation of measurement can be undertaken with a variety of methods. These methods includes, among other, observers, interviews, questionnaires, technical registrations, document analysis, documenting and analyzing artifacts such as tactical maps and other outputs etc.

As presented earlier, the need for huge amounts of data constitutes a resource problem. Administrating and analyzing data calls for resources in time and personnel. To enhance cooperation between analysts and clients the ambition is to have preliminary results presented to the participants/interested parties in close connection to the experiment.

6.2 Some foundations of measurement

The following section is a brief summary of some theoretical foundations of measurement. Readers primarily interested in practical applications could move on to section 6.3.

6.2.1 Measurement as representation

Formal measurement theory as it is known today is largely defined in "Foundations of Measurement Theory part I" by Krantz, Luce, Suppes & Tversky (1971). Consequently, the underlying perspective of measurement theory is the same as described in section 4.2 on modeling. Measurement as well as modeling is basically a matter of representation.

Krantz, Luce, Suppes & Tverskys' approach is based on the notion that empirical systems are viewed as non-numerical structures of relations with a set of non-numerical elements, N_1, N_2……N_n, and a set of relations, R_1, R_2……R_n, which constitutes a set N'. The representation of the empirical structure is an abstract structure with a set of (numerical) entities, n_1, n_2……n_n and a set of relations, r_1, r_2……r_n, which constitutes a set n'. A scale of measurement is then a relation-preserving function which translates N' to n'. The operation of measuring is thus viewed as the procedure of assigning numbers or symbols to empirical properties and relations in order to represent these in a formal system. Krantz et al. only discusses numerical representation but other types of representation are possible, for example those based on semantics or predicate logics (Flood & Carson, 1990). For example, subjective judgments can either be made by using a quantitative scale or by qualitative descriptions using natural language (Strangert, 2005). The theoretical discussion above on measurement is illustrated in figure 6.1.

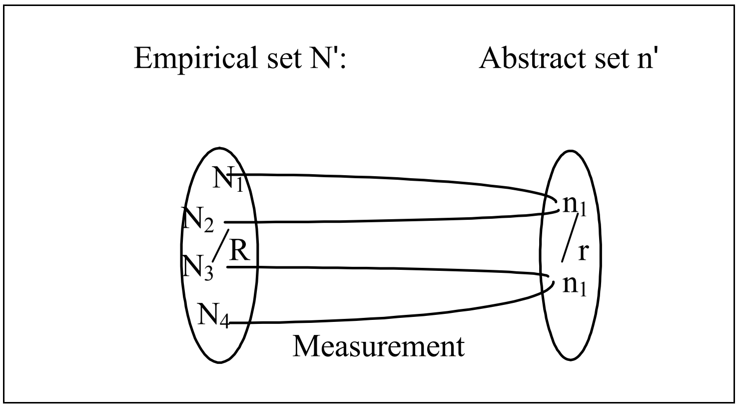


Figure 6:1. Measurement as a relation-preserving function, measurement theory Krantz, Luce, Suppes & Tversky (1971).

Another important notion in measurement theory is the distinction between four different scales of measurements (Stevens 1946, 1951). A nominal scale is the representation of (qualitative) differences between objects. Another word for nominal scale is classification. For example, mankind can be divided into two subcategories according to gender. An ordinal scale is the representation of order of objects according to some property, for example exam grades to rank study performance. An interval scale is the representation of intervals between objects according to some property, for example, the Celsius system to measure temperature. A ratio scale is the representation of ratios between object according to some property, for example the metric system to measure length.

Scale type is important because it defines which types of relations are possible to preserve on the measurement operation. As mentioned earlier a scale of measurement is a relation-preserving function which translates the empirical structure, N', to an abstract structure n'. A nominal scale translates empirical relations into an abstract representation based on the relations = and ≠, here expressed as (n', = or ≠). Thus, using such a scale will not permit representation of order, interval or ratio between objects. The corresponding translation for an ordinal scale is (n', > or <), an interval scale (n', + or -) and finally a ratio scale (n', * or ÷). In the same way not all relations are possible to represent using a certain measurement scale.

The type of scale used in the measurement operation limits which analyses are possible, or at least permitted, to carry out. For example, it is not possible to conclude whether something is twice as good based on a measurement using an ordinal scale, conclusions whether something is better is however permitted.

The practical consequence is that considerations on scale of measurement must be embedded in the procedure of modeling of experimental design. Modeling of experimental design is outlined in section 4.2.

6.2.2 Some problems associated with measurement.

Procedures and instruments for measurement might not be entirely reliable and valid. As a consequence the assigned numbers or symbols might not represent reality in a sufficient way. In general terms these problems are divided into (Graziano & Raulins, 2004):

Reliability means to which extent random errors have influenced in the measurement process. The experiment design must minimize circumstantial factors during the data collection.

Validity means absence of systematic errors in the measurement process. Different types of validity problems can be distinguished. A common classification is:

Construct validity refers to whether the instruments of measurement really produce data that represents the theoretical constructs. For example, there are several significant terms associated with command and control issues where there is no generally accepted approach for measurement. Situation awareness and decision superiority are examples of such terms.

Internal validity refers to whether factors, other than those central to a study, have influenced data. A study of how information structure influences situation awareness might be biased by extraneous factors present in the experiment setting. Examples of such factors are training effect and differences in experience among participants. The definition of extraneous variables in the modeling of experimental design is a manifestation of an ambition to avoid problem with internal validity.

External validity refers to whether data and results can be generalized to other settings and circumstances. As any experiment must be carried out in a specific setting it has to be considered whether there is anything about the experiment setting that makes it unique and/or different compared to a “real” situation.

The scientific approach means to make sure that the experiment is not charged with these problems. A feature of methodological tools and procedures is to gain control over the experimental setting. One such operation to gain control when studying complex systems is triangulation of data. This means that as many methods as possible should be used to measure one empirical variable. The rationale is that when several different subjective and objective indicators show the same patterns it indicates that the study is valid. When studying complex command and control systems the general approach is, if possible, to measure each variable from at least three different domains (Alberts & Hayes, 2002).

1. Physical domain. For example, the products each staff delivers or the activities they perform
2. Information domain. For example, which information is communicated by whom.
3. Cognitive domain. For example, the participants’ knowledge, opinions etc.

These domains could in turn be measured in several different ways, for example, by using both observers' and participants' judgments. Another example is shown in figure 4:6. The variable ’solve task’ was measured in three different ways: 1. The observer’s subjective judgment of how well the task had been solved. Data was collected by a pre-defined observer protocol. 2. The task force own perception of how well they had performed. Data was collected by a pre-defined questionnaire. 3. Whether certain actions where undertaken according to a predefined list. Put together, these indicators constitute a “profile” of organizational performance.

There is a vast amount of literature on data collection and issues related to reliability and validity and it will not be discussed further in this report. Our recommendation is to consult this literature for further references, for example Graziano, & Raulins (2004) or Blalock (1982).

6.3 Our approach to data collection.

A set of measures are used as “indicators” on organizational performance in the initial data analysis. The definition of these measures is a part of modeling the experiment design as described in chapter 4. Examples of some data collection methods are presented in table 6:1.

Table 6:1.*Examples of data collection registration methods*

Data collection method	**Significance**
Observation	Observers' manual recording of organizational activities. All observer reports specify time (hour and minutes) and place.
Recording of communication	Automatic digital recording of radio and telephone communication. All communication is labeled in terms of time.
Registration of positions	Automatic registration of geographic positions using satellite navigation equipment
Photography	Documentation of activities and situations using digital camera. All photographs must be manually recorded and labeled in terms of time.
Video recording	Video recording of activities and situations.
Event cards	When there is an evaluation situation and some predefined events are expected to happen to certain individuals, these are equipped with event cards. For example, extra personnel serving as casualty markers might be equipped with causality cards on which they make notes during the exercise (Thorstensson et al., 1999). They note the time for specific events in the medical attendance.
Collecting log files	When support systems are used, log files that contain interesting parameters are collected from these systems. This facilitates the study of system usage and availability, in the overall context.
Collection of (evolving) tactical documents	Documents belonging to orders, reports and briefings etc are a rich source of information. Copies of these (over time as they evolve) will contribute to the exploration of the information situation in the command posts and the field units.
Registration of table for recourses and activities	The rear echelon command posts often use tables for presenting situations and prognoses regarding important resources and key activities. Documenting these tables by regular documentation, photographs or other means available enables the analysis of the information situation and activities in the command post.
Questionnaires	Personnel's thoughts, opinions and feelings can be registered using questionnaires.
Interviews	Interviews can be used to get a picture of how personnel perceive certain matters.
Physiological measures	Some physiological measures, such as hormone levels or blood pressure, can be used to indicate for example fatigue and stress.

The last decade's development in information technology has had a profound impact on the possibilities to make the data collection process more efficient. The ambition in this and related work is and has been to eliminate as much manual work as possible. For example, digital questionnaires in combination with net work technology has made it possible to collect and import data to analytical tools such as SPSS (SPSS, 2005) more or less seamless.

Time pressure, the amount of data together with the strategy of using multiple indicators and multiple measures for every variable makes the administration of data a delicate matter. We have normally used a data matrix to organize data according to instruments of measurement and variables to measure. An example from an experiment carried out in the spring of 2005 at Singapore Center for Military Experimentation (SCME) is shown in figure 6:2. The purpose of the study (Cheah, Thunholm, Chew, Wikberg, Andersson, & Danielsson, 2005) was to explore the effects of a decision model combining a naturalistic planning and decision-making model called the Knowledge Battle Procedure (KBP) and a C2 System for distributed planning called MissionMate (MM) with TeamSight (TS). The study was done as a part of the overall Swedish Armed Forces (SAF) and Singapore Armed Forces collaboration framework. In the study a large numbers of variables were measured by several different instruments of measurements

	Other	Confounding variables		Dependent variables											
		Z1. Similarity of scenario	Z2: System failure	Y1 Team Creativity						Y2 Quality of decision		Y3 Completion of process		Y4 Communication pattern	
Instrument of measurement		=	=	Y1:1 Generated options	Y1:2 Idea sharing	Y1:3: Idea communication	Y1:4: Openess	Y1:5: Dominant member	Y1:6: Dominace from leader	Y2:1 Quality of critical decisions and plan.	Y2:2 Situation awareness	Y3:1 Reaction on critical evemnts	Y3:2 Completion of process	Y4:1 Communication pattern	Y4:2: Naturalistic approach
Observer protocols															
O1: On-line communication	4,5,6				1						2,3			1,2,3,4,5,6	
O2: Time-out staffs work					1,2	3,4	5,6	7,8	7,8				9		
O3: after run on KBP and scenario		2,3,4,5	1A,1B,1C												
O4: On line on process										3, (4)			1,2,4		
05: SME after run on quality										6A, 6B, 6C					
Questionnaries															
Q1: Time out questionarie					1,2	3,4	5,6A, 6B	7,8	7,8						
Q2: after run on quality, KBP and scenario		2,3,4,5	1A,1B,1C							6A,6B,6C			7,8,9		
Q3: Before exercise on background	1-9														
Q4: Time-out situation awareness											1A,1B,2A ,2B				
Q5: After planning questionnarie	6									1,2,3,4,5					
Q6: Mission Mate Questionnarie					3					1b	1a,2a, 2b		1b,1c,1d,1e		
System log															
L1 Communication					1						2,3			1,2,3,4,5,6	
L2: Products to be rated in 05															

Figure 6:2. Data matrix of an experiment carried out at Singapore Center for Military Experimentation .

Each column of the matrix represents a variable. Four “major” dependent variables were defined: ‘team creativity’, ‘quality of decision’, ‘completion of process’ and ‘communication pattern’. Each of these dependent variables consisted of a number of sub variables. A few confounding variables were defined but in this particular case no intervening variables were defined. Some of the measures defined in the measure model were later excluded. These variables are marked with a red cross.

Each row represents a specific instrument of measurement. Several different observer protocols and questionnaires were defined. Finally, some relevant system registrations were used as measurement instruments. In each cell the measure is defined more precisely. For example, a questionnaire might be used to collect data on several variables. In the cell are the specific questions of the relevant questionnaire indicated.

The major advantage of the data matrix is that it gives an overview of the data collection plan and is a guide line to keep track of the analysis if data. So far, the matrix has not been digitized and integrated in the MIND system.

The result from this initial analyses and interpretation of data provides a guideline for a further in depth analysis using a wider range of data. Even if the modeling is thorough and data collection on pre defined indicators is well elaborated, the complexity of the experimental setting often makes it impossible to predict every course of events and outcome. We have therefore taken a wider approach to data collection during the experimentation. It is our experience that complimentary data is always requested. The in depth analysis is then an exploration of large and diverse amounts of process data collected “parallel” to the pre defined indicators. Consequently, the need of efficient techniques and procedures to administer, structure and store data is an important issue. One ambition is to construct a wider data set consisting of a "*Mission History*", a time-synchronized multimedia model of the course of events is constructed. The construction of such multimedia models are based on the MIND-system described in section 3.5.

In the following section we briefly describe the collection and administration of data from different types of sources. We will also discuss how to practically deal with large amounts of process data, and provide a practical example from an experiment.

6.4 Case 6: Development of data collection techniques in experimentation, Demo 05 H.

Some of the principles to conduct experimentation on C2-processes outlined in Evolva are under the implementation at the Swedish Armed Forces C4I Development centre, LedUtvC. The work is undertaken by the project MARULK which aim is to develop a system to support analysis. Using the MIND system as the fundamental framework, the support system will provide:

- Data collection from a number of different sources.
- Time stamping of data to second precision.
- Data compilation in a common database.
- Data processing during ongoing experiment.
- Presentation and visualization of collected data, as multimedia representations of mission histories (Morin, 2002) and as explaining tested hypotheses (Albinsson, Morin & Thorstensson, 2004). Presentation will be possible a few hours after experimentation.
- Support for after-action reviews (Rankin, et al, 1995) and post-mission analysis.
- Support for including operator and domain-expert comments and interpretations in the database.

In the fall of 2005 the experiment Demo05H was executed at LedUtvC as a C2 exercise with 300 participants from all three branches of the Swedish Armed Forces organized in 22 staffs at different levels of command. Participants were connected in a joint local area network (LAN) simulating a network-centric warfare environment. During the experiment, MARULK in participation with other projects performed extended trials on data collection from multiple sources. Trials also included data compilation in a common database, data processing during experiments and connecting data to hypotheses.

6.4.1 Data from participants.

Digital questionnaires. A digital questionnaire system was introduced using web technique to increase speed, and performance in collecting, compiling, processing and storing data. All participants answered the questionnaires before the exercise and directly after each specific experimental run. Each questionnaire comprised 30-100 questions. The selected system did not meet expectations. A combination of technical and usability problems resulted in severely corrupted data such as duplicated registrations and loss of data.

Group interviews. Another data collection method was group interviews performed after each specific experimental run. Each group consisted of participants from different staffs, making up a special chain of command or a specific function. 20-30 different group interviews were held after each exercise. Documentation from interviews was made in MS word. Possibly, documentation could be done using a questionnaire tool.

6.4.2 Data from observers.

Digital questionnaires. Besides taking notes as a basic method, observers also answered questionnaires after each experimental run. The same digital questionnaire system was used as the one used by the participants. Of course, the content of the questionnaires for observers and participants differed.

Computerized observer tool. A computerized observer tool was also evaluated. The tool is based on earlier research on time-stamped structured reports (Thorstensson, 1997) and on how to support observers (Jenvald, Morin, Crissey & Thorstensson, 2002). Four experts observing four different command posts were used to test the prototype tool in the two final experimental runs. The number of reported observations was a total of 123 and 154 respectively, with a somewhat similar distribution between the different observers. Each observer tool was connected to the network and the reports were sent to a central database as soon as they were saved locally by the observer. This made it possible to display all observer reports in real time at different locations in LedUtvC and gave the personnel at the analysis centre (AC) the possibility to follow course of events at different command posts simultaneously, which was considered a great improvement for their work. All four experts appreciated the tool and regarded it as a significant support for making qualified reports from observations. Further development will be made in order to implement tools for communication

6.4.3 System data

Ground Truth. The tactical setting was simulated using different interconnected simulator systems. The tactical course of event was recorded by logging the HLA data for the federation. The spatial and temporal course of events for the approximately 800 entities participating in each experimental run was registered

Voice communication. Three different systems for voice communication were used during the exercise. All of them utilizing voice over IP (VoIP) in the joint LAN. None of the tested systems provided available means for logging communication. To overcome this problem software were developed to record everything that is said in selected microphones and everything that is heard in selected loudspeakers. These programs were installed on all approximately 300 computers in the LAN and were then used to register voice communication for 36 key participants. In each of the four experimental runs we recorded some 100.000 communication events. All communication events were time stamped and, if possible, connected to an identified sender and an identified receiver.

System utilization. Software was installed in all participating computers on the LAN to be able to register the screen of the operators. A screen shot was registered every fifth second making it possible to compile the data to a film strip. As the registration was time-stamped, it was possible to synchronize these film strips with the data voice recording of the selected 36 key participants and the recording of ground truth.

6.4.4 Mission histories

The possibility to synchronize data was used to compile “mission histories” for a specific time frame of the experiment. Selected sets of data formed mission histories making it possible to replay the selected course of events. An example of the visualization of such a mission history is shown in figure 6:3.

Figure 6:3. A screenshot from the MIND system, showing a Mission History from Demo05H.

It is possible to have predefined data collection plan for the compilation of such mission histories. It is also possible to compile a mission history in an “ad hoc” fashion based on the wide set of collected data. In one specific case during the experiment, there was a misunderstanding between two command posts at different levels of command. After the exercise the operators wanted to find out why this had happened and asked for help to reconstruct the communication from that part of the exercise. By exploring the course of events and communications from one specific “commander’s conference” it was possible support them in defining cause and effect relationships and use the collected data to improve future performance and decrease risk for similar misunderstandings.

6.4.5 Administration of hypotheses, data and conclusions

Tool to administer hypotheses. Initial studies to define a tool to enhance documentation of results from experimentation with a traceable link between hypotheses, data and conclusions were also undertaken. The idea is that the hypothesis tool will be used to organize and compile important data collected during experimentation. Hypotheses, data and conclusions will be linked to mission histories. It also becomes the entrance to the considerably wider set of data that has been generated during the experiment. The prototype of the tool is described in section 8.6 and shown in figure 8:1.

7 Feedback of test results

7.1 Background

As we have mentioned in chapter 3, it is important to integrate the clients into the experiment process. One way of doing this is by developing the methods of presenting results from experiments. However, our perspective has been to find new ways of communicating test results to development work as a complement to traditional methods, i.e. written reports. One ambition has been to have preliminary results displayed to the participants/interested parties in as close a connection to the experiment event as possible. Such an After Action Review (AAR) gives the opportunity to discuss and react on the results directly. In this chapter we present our experience of developing new procedures and approaches of providing feedback of results from experimentation.

7.2 A generic model for feedback

A control model derived from system theory (Katz & Kahn, 1978) includes a cyclical process of several phases: (1) diagnosis, (2) planning, (3) data collection, (4) analysis, and (5) feedback.

The diagnostic phase (1) includes definition of present status of the object of research/development (e.g. a process or an organization). In the planning phase (2), the diagnosis is compared to a goal state or goal effect that the development project is trying to achieve. The purpose is to identify activities or/and interventions that might be necessary to undertake in order to achieve these goals. After the data collection phase (3) comes the analysis phase (4), where these identified activities and interventions are tested as hypotheses. Will the suggested actions lead to the stated goals or/and effects? The feedback phase (5) has the purpose of interpreting the results together with the clients and to "tune" future activities/interventions so that stated goals and effects can be better achieved. The model is shown in Figure 7:1.

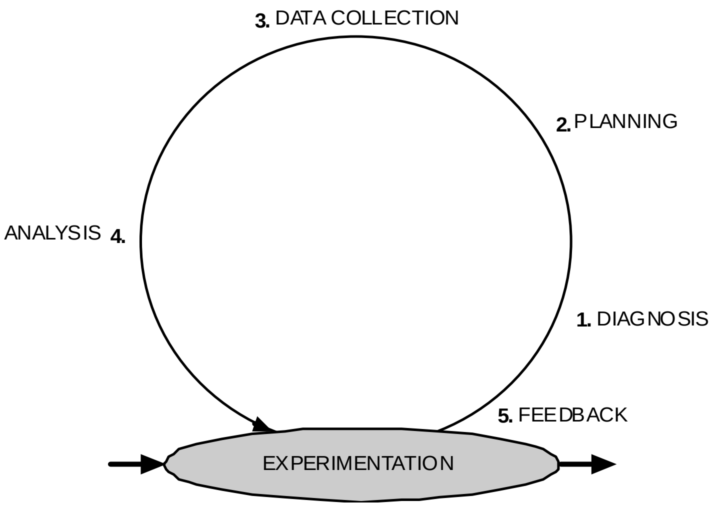


Figure 7:1. A generic control model of experimentation.

However, one single control loop is not sufficient when the evaluation process stretches over a period of time. An iterative approach is necessary: formulating hypotheses; testing them in various test environments; turning back to the initial model for evaluation and revision; and generating new hypotheses to be tested in additional tests and experiments.

In “Code of best practice for experimentation” (Alberts & Hayes, 2002) the need of “experimental campaigns” — a series of interconnected experiments aiming at exploring and developing “Mission Capability Packages” — is stressed. An important implication is that feedback of results and experiences from any experiment is relevant both in the short and the long perspective. It should be possible to use results as early as possible for decisions on further work in the development project and also to use data and experiences from an experiment after a considerable amount of time. Presumably, this will enhance the possibility to create a “body of knowledge” which is traceable to results from earlier experimentation. This “body of knowledge” should also be used in planning of future experiments and activities.

7.3 Our feedback model

We argue that it is possible to view the feedback process from any single experiment as a series of different sequential phases. We suggest the following phases:

Phase 1: During experimentation. The normal procedure is often to focus on data collection during an experiment. Analysis is conducted some time after the experiment. During an experiment, when researchers, clients, and other relevant competence are present at the test location, there is reason to use this opportunity for analysis based on the multi-disciplinary competences. The purpose of feedback in this phase might be to give a basis for decisions to make changes between runs.

Phase 2: After Action Review (AAR). Directly after an exercise it is common procedure to gather all participants to discuss the conducted exercise in an after-action review. The purpose of feedback in this phase is to present and discuss preliminary results from the experiment with the participants in order to validate and interpret findings and also, a very important issue in the AAR: to collect additional data such as the participants’ opinions.

Phase 3: Post Mission Analysis. It is not always possible to conduct an in depth analysis of compiled data in the AAR phase. It is not possible to include and discuss all results in an AAR as available time, resources for preparation, and the actual AAR session are limited. Additional analysis of data in cooperation with a limited number of clients and domain experts might be necessary to conduct after the experimentation and the AAR phases. The purpose is to perform as much data exploration as needed for enhanced testing of defined hypotheses.

Phase 4: Experiment Conclusion. At some point in the feedback process it is necessary to make a closure of the conducted experiments and start focusing on the next action. The normal feedback in the closure phase is a written a report in order to present a final and official documentation of results from an experiment.

Phase 5: Lessons Learned. After final closure of an experiment, the compiled databases can still be of value. Questions might arise later that were not recognized at the time when the experiment was conducted and other related projects might also be interested in exploring data from earlier experiment. The purpose of this phase is therefore to construct a searchable generic database of experiences from previous phases to be used as a base for future interventions and actions.

Phase 6: Planning. In the initial planning phase of succeeding experiments, results from earlier experimentation should be considered. The planning should be undertaken by an integrated team of researchers, clients, and relevant domain experts. The purpose of this phase is to define a basic research design. The essential outcome of the planning phase is a definition of what to study and how to do it. Thus, this phase is connected to the issues discussed in chapter 4. Earlier results from experiments should be included in the modeling process.

Phase 7: Pre-Action Presentation. Even with a well developed research design the analysis needs an in depth preparation. This is especially true if there is a demand of feedback during the experiment. Of course, one such source of preparation is to present data from earlier experiments. The purpose of this phase is to prepare participants and/or the analysis team before the next experiment. One way is to study interventions and actions from previous experiments.

Each separate phase can be described as a control model shown in figure 7.1. Several such sequential control models might be interconnected so that output from one becomes input in another, as depicted in figure 7:2.

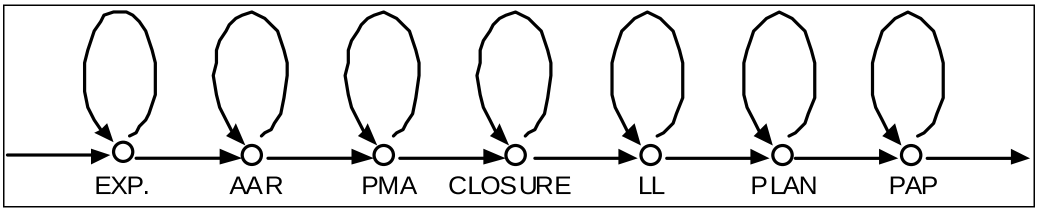


Figure 7:2 Suggested feedback phases in military research and development.

7.4 Illustrative cases

In this section we will discuss the suggested phases more thoroughly. To illustrate the discussion, some examples of feedback of results in some of these phases are presented. Each phase calls for a different approach which is exemplified with cases. This approach can not be fully applied in this chapter since there are not yet concrete results for all the phases.

One might argue that each of the phases in an experiment cycle should contain well structured feedback processes. However, due to the specific circumstances of every single experiment, the nature of feedback of results might differ. It is our experience that a fully conducted feedback cycle in every phase can seldom be attained. Thus, the selected cases are examples of how feedback of results has been performed in a single phase and not for the whole cycle.

7.4.1 Feedback of results during an experiment

During an on-going experiment the involvement of the clients is crucial. The analysis of results should not be left solely to the researchers. As researchers, clients, and other relevant competence are present at the test location, there is reason to use this opportunity for analysis based on the multi-disciplinary competences. Often, an experiment is conducted as a series of runs during an exercise. Thus, time is crucial if the results are to be interpreted during an on-going data-collection series during an exercise. This is especially accentuated in an action research approach (see section 3.2), where the results from the initial test runs should be input for further runs. However, the need of feedback during a single case study (see section 3.2) is also present as it might save time and resources invested in the post mission analysis.

7.4.1.1 Case 7A: Demo 05 V

Experimentation is an important feature of the Swedish Armed Forces' transformation from invasion focus to a flexible net-work based operational defense. Demonstrations and tests of new technology and methods will be used to resolve uncertainties, test technical (and financial) feasibility, and illustrate alternative courses of action. The results will be analyzed and assessed through study activities before deciding on continued development activities. One important sub-program in this process is LedSystM which is responsible for development of command and control (C2) methods. As a part of the demonstrator program, command and control methods and principles will be tested in a series of experiments.

In the exercise Demo 05 V, held in April 2005, the focus of the experimentation was to further develop a procedure for planning under time pressure. The model for the procedure, Planning Under Time Pressure (PUT), is based on a satisficing rationale approach, that is, the purpose of the planning process is to find a good-enough option early in the process and then refine and corroborate it through war gaming. The model is originally developed at the Swedish Defence College and was first introduced in the Swedish armed forces as a tactical planning model in 2003. A full description of the PUT-model is provided by Thunholm (2003, 2005).

More precisely, the aim was to develop procedures for the cooperation between staffs and the distribution of commander's intent in the context of the PUT model. The experiment was conducted during a series of war-gaming sessions. The test design was based on an action-research approach. The intention was to conduct the analysis according to the following procedure:

1. During, and instantly after, each war-gaming session a set of variables were measured with a number of different instruments, i.e. observers, questionnaires and registration, and classification of radio and data communication. The variables - task knowledge, cooperation, product quality, distribution of commander's intent, and communication pattern - were used as indicators of organizational performance.

2. Between each war gaming session the result from the measurements were analyzed together with personnel from LedSystM. Participants of the exercise also conducted workshops on the theme "What can be changed in the planning process in order to improve the PUT process?".

3. Based on results from the steps above decisions were taken together with personnel from LedSystM on what to change in the PUT process during the next war-gaming session. Instruments of measurements could also be changed.

4. During the next war-gaming session the organization was measured using basically the same set of variables. The basic question to be answered was if the changes introduced made any difference. If so, what was the nature of the effect? Refined and supplementary questions and hypotheses might be added to the analysis.

The process described above was implemented with the help of military personnel allocated to the work from LedSystM and is described in more detail below:

a) The initial task was to explain the results from the analysis of data from the first war-gaming session. Results showed a strong negative correlation between the commander of the joint staff and his staff members on the ranking task of the importance of different events that occurred during the session. This result might be interpreted as a failure of the commander to communicate his intent within his own staff. The personnel from LedSystM used the data base from the digital questionnaires in order to find explanations of the result. It was also possible to trace data from staffs, observers, and personnel for further data collection. The most plausible explanation found was that the result was due to the fact that distribution of information to the subordinate staff had been prioritized. Thus, the commander's intent had not been formally communicated in the joint staff. Another explanation might also be that officers from different services had different competence and traditions and therefore interpreted the same situation differently.

b) The next step in the procedure was to come up with solutions to overcome the identified problems. In this case, the major problem was judged to be the inconsistent perception of the situation in the joint staff and the absence of communication of intent within the staff. The opinion was that the most significant tool was that each staff should establish a timetable in order to push the planning process forward.

c) Finally, two changes in the experimental design were defined. First, changes of the measurement of distribution of commander's intent were implemented. The aim was to overcome the problem with the difference in perspective on the mission between services. Instead of just ranking events of primary interest for the joint operational level, the ranking also included lists of events of importance for each of the services. Second, a minor experiment was also conducted between the first and second session in order to compare different principles to formulate a commander's intent.

In this minor experiment twelve officers were divided into two groups. Each individual in the groups acted as a liaison officer receiving a commander's intent communicated by the commander of the joint staff. One group received commander's intent expressed in terms of the ”required effect” while the other group received the same commander's intent expressed in ”required capacity”. Communication between the officers within or between the groups as well as questions to the commander where not allowed. Measurement of distribution of commander's intent consisted of ranking of importance of a selection of defined events. Each participants were asked to rank the order of a number of injects regarding their threat to mission success. The participants ranking of importance were compared within each staff as well as with the commander (whose ranking were considered to be the correct ranking in accordance with the communicated intent). Result of the data analysis indicated that the group subjected to commander's intent expressed in ”required effect” showed a higher consistency in ranking of importance relative to the “correct” commander's ranking.

Results from the intervening experiment were not put forward as a finding to be generalized. Instead, the results were used as a basis for an AAR discussion a few minutes after the “intervening” experiment. The purpose of the discussion was to try to explain the result and to come up with actions to improve transfer of commander's intent. During the discussion a hypothesis, intended to be tested in the next war gaming session, was formulated: “Commander's intent should be mediated in terms of “*required effects*” but the following dialogue should be based on the ”*required capacity*”. A template to support a dialogue based on this hypothesis was constructed and informally tested in the following session. The possibilities to observe this dialogue were limited due to limited space for observers.

7.4.2 Feedback of results during an After Action Review - AAR

Directly after an exercise it is common procedure to gather all participants in order to have a post exercise review. Such an After Action Review (AAR) is a common approach for this kind of feedback session. AAR was first implemented in the mid 1970's (Rankin et al., 1995). The AAR approach has been developed ever since to provide and deliver feedback after a collective training exercise in order to enhance training value. This is made possible by collection and presentation of data that characterize the performance of the units in relation to preset goals and mission outcomes.

Traditional use of AAR. As defined in Training Circular (TC) 25-20, A Leader's Guide to After-Action Reviews (U.S. Army Combined Arms Center [CAC], 1993), an AAR aims to answer the general question "*How did the unit do?*". This question is broken down into three more concrete sub-questions: "*What happened during the collective training exercise?*", "*Why did it happen?*" and "*How can units improve their performance?*". The goal of an AAR session is to enhance the training effect.

The answer to the main question *"How did the unit do?"* may not be immediately obvious to the participants or to those who planned and controlled the exercise. Morrison & Meliza (1999) explain how this understanding is easily lost or obscured by "the fog of war" or just the simple fact that it is impossible for each and everyone to have an overview of the amount of distributed, but related, events and actions going on during a collective training exercise. An important part of the AAR is thus the ability to raise the level of awareness about the actual course of events. FOI has been working with development of methods and tools to support reconstruction and exploration of the chain of events for an exercise. This work is described in more detail in Morin, Jenvald & Thorstensson (2003).

AAR in experimentation. The traditional use of AAR methods and technology, described above, is to structure training and evaluate exercises in order to achieve or increase training value. Based on the traditional content of AAR, we argue that the approach is applicable, with some modifications, in the context of experimentation. The general idea is that the AAR approach can be used to support feedback of results from tests and experiments in the context of a development process. The AAR is an opportunity to present data and preliminary results in a comprehensive way in order to discuss and analyze the material with the participants in the time frame of a traditional AAR and, most importantly, to *collect additional data from the participants*. The main effort is not to reflect upon individual and group performance but to reflect on the questions of the experiment.

However, it is important to realize that the AAR approach tailored for training is not directly applicable to report findings from tests and experiments. Instead, we argue that in experimentation the AAR should be used to discuss preliminary results in order to validate and interpret them and to collect complementary data. Hence, in experimentation the AAR is a data source to support analysis, whereas during training the AAR is an important part of the training process.

With this changed focus of AAR we need to modify the three questions for self examination. The main question of *"How did the unit do?"* is no longer necessarily in focus and is therefore, in a hypothesis-driven experiment, modified to *"Which hypotheses were rejected?"* The results of the planned data collection and analysis are presented to the relevant participants of the experiment, in the form of supported or rejected hypotheses. The traditional questions are thus slightly modified as shown in table 7.1:

Table 7.1: *Modified AAR questions for experiments compared to AAR for training*

Traditional (training) AAR questions	*Modified (experiment) AAR questions*
"What happened during the collective training exercise?" AAR participants attempt to specify the facts (i.e., important actions and outcomes) of the exercise.	"What happened during the experiment?" *AAR participants attempt to specify the facts of the course of events in the experiment.*
"Why did it happen?" Given the facts of the exercise, the participants attempt to explain the causes of particularly important actions and outcomes.	"Why did it happen?" *Given the facts of the experiment, the participants attempt to explain the causes of the outcomes of the test of the hypotheses.*
"How can units improve their performance?" *Given that the previous two questions are answered, the participants determine actions to solve identified problems. For example, changes to standard operating procedures (SOPs) or increased training on basic drills.*	"What is the implication for the development project?" *Given that the previous two questions are answered, the participants identify problems (and possible solutions) relevant to the development project.*

Some challenges for using AAR in experimentation. One important issue in the development of new C2 systems is that highly abstract processes such as decision making, situation awareness, etc. must be analyzed and developed. Consequently, data on these matters must be collected, analyzed, and presented in an AAR. This data collection and analysis brings out inherent problems.

Since the time span between a test and the following AAR should be as short as possible, the amount of time available for data analysis is limited. Consequently, data from decision-making processes are normally limited to quantitative data for statistical analyses. Our experience is that the presentation of results from such analyses poses a problem as evaluation often becomes a lecture on statistics rather than evaluation of a decision process (Wikberg & Lundin, 2002a). We evaluated the "normal" technique to present statistics in the format of graphs in the context of an AAR in an 'Experimental Simulation Exercise' (ESE). The evaluation tested how the involved experienced officers understood the results presented (Wikberg & Lundin, 2002a). During the ESE, two additional inquiries were distributed to the participants. The first, distributed directly after the regular questionnaire concerning the problem studied in the ESE, was designed to survey to what degree the participants understood which comparison the graphs intended to represent. The second, distributed directly after the presentation of the results from the regular inquiry, was designed to elucidate which results the participants believed they would attain. The results indicate that the foundation for the AAR was incorrect since the participants could not explain the information in the presented graphs. In the multiple-choice questionnaire they performed slightly poorer than a random outcome. This could partly be explained by the fact that there was a difference between the expected and the attained result, which is a cognitive dissonance. Consequently, statistical graphs should be avoided in an AAR. Results used in an AAR must be presented in a familiar context and in an understandable manner to the participants. As the time available for presenting and discussing the results normally is limited, it is important to find procedures to present statistics in a straightforward way.

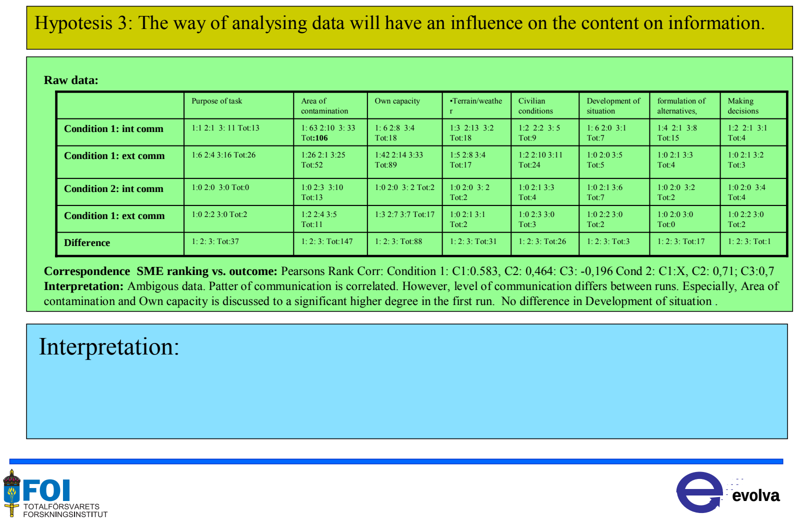
We have tried out ideas on how to tackle this problem. One solution is to present results in terms of rejected and supported hypotheses. Each hypothesis is stated together with a verbal description of the result of data analysis. The results are presented on power-point slides to the participants using a projector. On the slide it is also possible to document comments and additional information from the participants. As a complement to this the MIND system (see section 3.5) can be used to replay the mission history and explore collected data in order to give the audience a connection from the hypothesis and the statistical data to the actual chain of events during the exercise.

7.4.2.1 Case 7B: AAR in Case Study at SkyddC

An experiment was conducted at the Swedish NBC Defence Centre during an exercise with the national NBC protection task force. In the experiment, witch used an experimental strategy in a case-study approach, the task force had to manage a situation with chemical warfare agents discharged over a major city.

Two settings were compared. One was a situation were the existing NBC C2 system was used. In the other situation a simulation of a future C2 system outlined in the NBC Demo project was used. The difference between the settings was the degree of access to “intelligent” sensors. The futuristic system had the ability to automatically analyze and distribute data and thereby bypassing the human element in the analysis process. Based on the basic question “*Will this improve the decisions by the operators in the C2 system?”* a number of hypotheses were formulated. The experiment focused on three specific situations during the exercise. The first situation was input of data from a deficient sensor. The second situation was introduction of contradictory information from different sources and the final situation was to be forced to deliver an instant answer to an official authority on a request. Data was collected using observers, questionnaires, and registration and classification of radio and data communication.

The study was conducted in one day. The day after the exercise, participants were assembled for a half-day AAR. The purpose of the AAR was additional data collection and validation of preliminary results. The AAR had its basis in the preliminary result of the analysis of collected data. The preliminary analysis of collected data had been restricted to elementary statistics (rank correlation and t-test). Results were compiled on power-point slides in order to structure the discussion. Each slide presented the result of one of the hypotheses. An example of one of the slides is shown in figure 7:3.



Hypotesis 3: The way of analysing data will have an influence on the content on information.

Raw data:

	Purpose of task	Area of contamination	Own capacity	•Terrain/weather	Civilian conditions	Development of situation	formulation of alternatives,	Making decisions
Condition 1: int comm	1:1 2:1 3: 11 Tot:13	1: 63 2:10 3: 33 Tot:**106**	1: 6 2:8 3:4 Tot:18	1:3 2:13 3:2 Tot:18	1:2 2:2 3: 5 Tot:9	1: 6 2:0 3:1 Tot:7	1:4 2:1 3:8 Tot:15	1:2 2:1 3:1 Tot:4
Condition 1: ext comm	1:6 2:4 3:16 Tot:26	1:26 2:1 3:25 Tot:52	1:42 2:14 3:33 Tot:89	1:5 2:8 3:4 Tot:17	1:2 2:10 3:11 Tot:24	1:0 2:0 3:5 Tot:5	1:0 2:1 3:3 Tot:4	1:0 2:1 3:2 Tot:3
Condition 2: int comm	1:0 2:0 3:0 Tot:0	1:0 2:3 3:10 Tot:13	1:0 2:0 3: 2 Tot:2	1:0 2:0 3: 2 Tot:2	1:0 2:1 3:3 Tot:4	1:0 2:1 3:6 Tot:7	1:0 2:0 3:2 Tot:2	1:0 2:0 3:4 Tot:4
Condition 1: ext comm	1:0 2:2 3:0 Tot:2	1:2 2:4 3:5 Tot:11	1:3 2:7 3:7 Tot:17	1:0 2:1 3:1 Tot:2	1:0 2:3 3:0 Tot:3	1:0 2:2 3:0 Tot:2	1:0 2:0 3:0 Tot:0	1:0 2:2 3:0 Tot:2
Difference	1: 2: 3: Tot:37	1: 2: 3: Tot:147	1: 2: 3: Tot:88	1: 2: 3: Tot:31	1: 2: 3: Tot:26	1: 2: 3: Tot:3	1: 2: 3: Tot:17	1: 2: 3: Tot:1

Correspondence SME ranking vs. outcome: Pearsons Rank Corr: Condition 1: C1:0.583, C2: 0,464: C3: -0,196 Cond 2: C1:X, C2: 0,71; C3:0,7
Interpretation: Ambigous data. Patter of communication is correlated. However, level of communication differs between runs. Especially, Area of contamination and Own capacity is discussed to a significant higher degree in the first run. No difference in Development of situation .

Interpretation:

Figure 7:3 Example of an AAR presentation of a hypothesis test. In the upper box the hypothesis is stated. The green box contains a verbal description of the result of the data analysis. The blue box is used to document comments and additional information from the participants.

In the initial part of the AAR the participants were reminded of the purpose of the experiment. The rationale for this was to establish a common point of reference for the following discussion.

In the next step a discussion about possible confounding factors took place. The participants were informed that there had been three specific situations studied in the experiment. They were asked to comment on the situations during the scenario regarding their realism. Other confounding factors were also discussed, such as the effect of training between runs and their personal experience compared to their role in the task force. Data collected on these issues were presented to the participants. The purpose of this discussion was to establish some basis for estimating validity and reliability of the analysis.

Then, each of the hypotheses tested during the experiment was discussed and each of the defined hypotheses was presented to the participants as shown in figure 7.3. The purpose of the discussion was to further analyze and, if possible, explain the received result. The starting point was, of course, if they agreed with the preliminary result. Comments and complementary data were successively documented on each slide.

Finally, the AAR was completed with a discussion on the meta-level about consequences on the task force's organization and procedures in case the suggested technological system would be implemented. Another issue in this discussion was the organization of the exercise and possible future work.

Based on the collected data and the discussion in the AAR the conclusions from the experiment were:

- Data indicate a more time efficient C 2 process in the high tech condition.
- However, there is a lower degree of control and verification of information in the high tech condition.
- The implication of the experiment is that new technology will not automatically decrease the need of training.
- Another implication is that today's organization of the task force might be inappropriate if the NBC Demo system is introduced.

7.4.3 Feedback of results during Post Mission Analysis (PMA)

With extensive data collection in experimentation, it is necessary to allocate time after each instance of experimentation to compile data; in the short time frame in order to support AAR, but also to support subsequent activities utilizing a substantial data set. The phase following the AAR is the Post Mission Analysis (PMA). In the PMA methodological experts and domain experts work together with collected data to elaborate the analysis and focus on specific questions from earlier phases as well as issues identified in the ongoing analysis. PMA is normally performed about a week after the specific mission (or experiment) when data has been compiled to a certain degree. The time utilized for PMA is normally limited, but can vary depending on the amount of analysis questions and available resources. PMA procedures are quite well developed for training but are still to be defined for experimentation. The details of the case in the following section are not fully developed.

7.4.3.1 Case 7C: Demo 05 V

After conducting the series of experiments involved in the Demo 05 V exercise, we compiled collected data using the MIND system. The compilation of data consisted of a re-playable dynamic mission history. An example is shown in Figure 7.4.The mission history replay was then the foundation for analyzing the friction in communication between the different command levels within the C2 structure. The aim of the PMA after Demo 05 V was to demonstrate the possibilities in the available system.

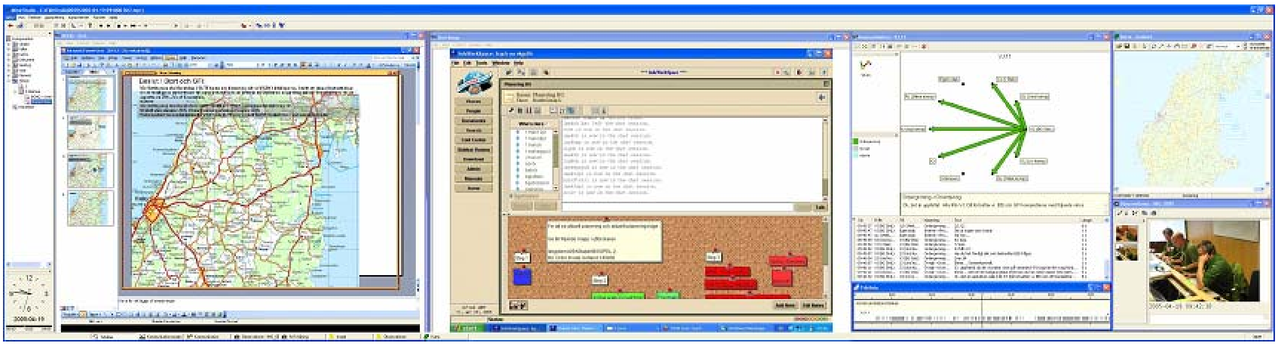


Figure 7:4 An example from the MIND system showing a PMA visualization of communication from the Demo 05 V exercise. The time-synchronized model in this instance contains two different operators' screens (two leftmost large windows), a communication event (green arrows), a map, and a situation photograph from the battle group head quarters (BGHQ).

7.4.4 Feedback of results in the final experiment conclusion

At some point in the feedback process, it is necessary to make a closure of the conducted experiment and start focusing on the next action. Normally, this means that results and experiences are documented in a written report. The level of quality of this written report might vary. The normal procedure for checking the quality of a scientific study is the peer-review method. The report is scrutinized by some external individuals who are independent of both the researchers and the clients involved in the study. Of course, this calls for resources to be put aside for this purpose: A decision where the client makes the ultimate judgment.

One important goal is to engage the client in the production of the "final" documentation of an experiment. A written report from a large scale experiment might be quite extensive and thus hard to penetrate for a client. Alternative procedures to produce a report with conclusion are desirable. We have tried to tackle this problem by presenting the elements of a traditional report in a graphical model-based layout (Wikberg & Lundin, 2002b). A case illustrating such a process is described in case 7.D.

7.4.4.1 Case 7D: Model-based documentation in a study of the C2 system of a division

Focusing on the cooperation between different levels in the chain of command in an army division (Wikberg & Lundin, 2002b) a study was conducted in the spring of 2001 during the Army's staff and communication exercise. The purpose was to study roles responsible for supplying the system with information in accordance with a successively identified need. The problem in focus of the study was structured in a modelling session, where personnel from the Army and the Swedish Defence Research Institute participated. In the modelling, the most relevant factors of the problem were identified. Methods for measuring these factors in the given context were also defined.

The model was used as an observation protocol by military staff instructors. Following each staff the staff instructors documented identified events, incidents procedures, and so forth that corresponded to the defined measures on site. The model and its definitions form the basis for the data collection. Every day an evaluation meeting was held, where each staff instructor presented his documentation. Figure 7.5 shows a situation during such an evaluation meeting.

Figure 7:5. A model-based evaluation meeting

All data where related to the factors of the model by using a print-out of the model on a paper sheet of approximately one by two meters (see figure 7:5). Documentation and interpretation were conducted successively during frequent evaluation meetings. Each documentation was included in the model as raw data. A new version of the model was printed for the next meeting. After a couple of days of data collection, the documented data were used to draw some preliminary conclusions on a heuristic basis on each of the factors identified in the modelling session. These conclusions on factors where presented in a more general discussion. The resulting graphical model based documentation is shown in figure 7:6. The spreadsheet contains the same information as an ordinary test report (including an introduction, sections on method, results, and conclusions).

The major advantage of the approach was that the primary recipients of the results were engaged in the evaluation sessions. This meant that they had first hand access and knowledge of the results and the context of the study in advance of the documented report. A major part of the documentation is actually the clients’ own wordings. This approach makes it possible to present an analyzed and documented result when the test is finished, as opposed to spending days or weeks to analyze the data before being able to present any results.

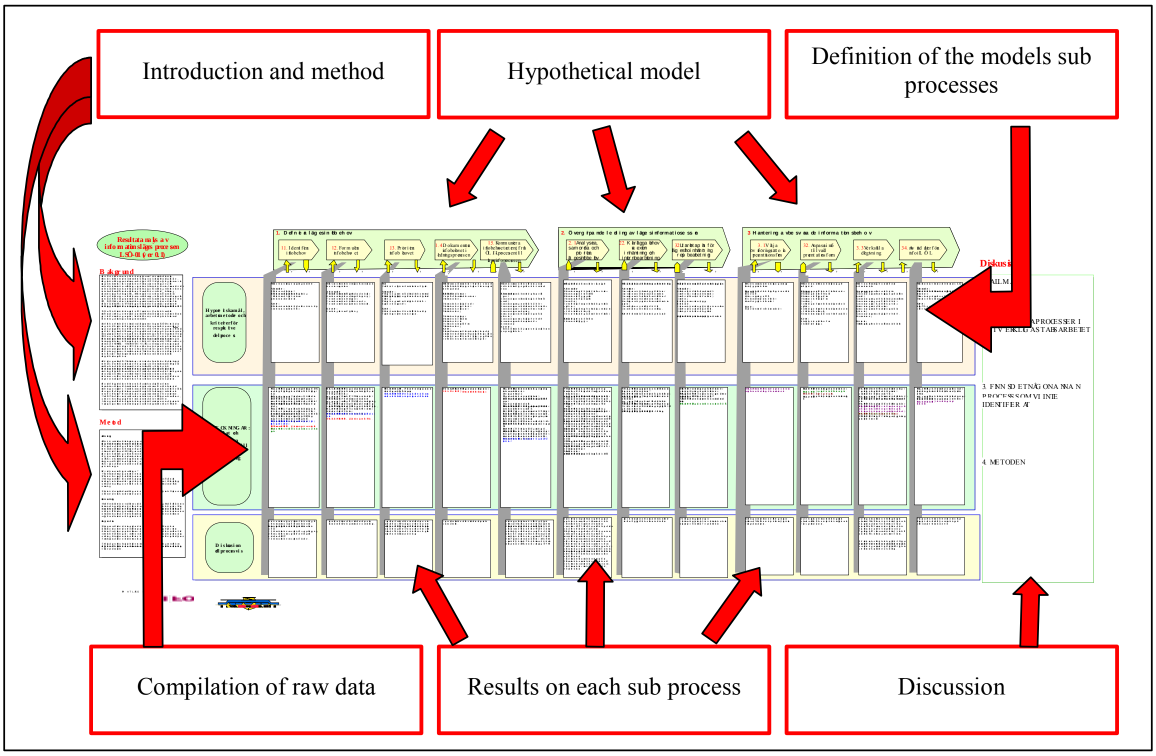


Figure 7:6. The graphical layout of the documentation of a model-based evaluation. In this particular case we used a paper sheet of approximately 1x2.5 meters, wherefore it is not possible to read the text in the figure shown here.

As the lay-out of the spread sheet contained all the necessary elements of an experimental report (se figure 7:6) it is a quite straightforward process to "cut and paste" the documentation into a regular format, i.e. an A4 text document.

In the case described above, a rather qualitative and pragmatic approach to analysis of data was used. There is of course no reason why the same principle can be used for other approaches to data analysis.

7.4.5 Feedback of results as lessons learned

Accumulated experience from a series of experiments can be documented in different ways. Using a re-playable mission history is one way of using data as a means for conveying lessons learned (LL). Traditional LL databases mostly consist of written reports on what people have identified as lessons from specific events. We argue that a re-playable mission history could complement written reports effectively. Future analysis of experimental series will also benefit from having data from previous experiments available and graspable.

However, the format and content of the documentation from an earlier experiment might not be appropriate for other purposes. A re-playable mission history has to be put together from a specific purpose. Using the same large data material for another purpose calls for different ways of analyzing and representing data. Developing procedures for reusing data from earlier experiments is an important challenge for future work.

7.4.6 Feedback of result during a planning phase

In the planning phase of a series of successive experiments, results from earlier experimentation should be considered. The purpose of this phase is to define a basic research design which, at least to some part, is based on earlier experiments. The essence of the research design is a definition of what to study and how to do it. Thus, results from earlier experiments is an important input in using modeling as a tool to define evaluation measures, an issue discussed in chapter 4. However, a more specific procedure on how to use recorded data from earlier experiments is still to be defined.

7.4.7 Feedback of results during a Pre-Action Presentation

The purpose of this phase is priming of participants, analysts, and observers before an experiment. To achieve this, a variety of procedures are possible to use. One source of such preparation is presentation of data and results from earlier experiments.

Military units can prepare a mission by using data and course of events from previously performed similar missions in a Pre-Action Presentation (PAP). In the PAP, strengths and shortcomings from previous missions can be visualized, discussed, and analyzed with the aim to improve coming performance. We argue that this is applicable for experiments as well. The audience in a PAP for experimentation can be the operators, who can assimilate previously gained experience to improve performance in the experiment, but an equally important group of recipients of the PAP is the allocated observers. Observers of an experimentation exercise must have detailed knowledge of the studied organization. Observers who lack experience of either method or domain must be treated as novice observers and be prepared accordingly (Jenvald et al., 2002). Methods for using data and results from previous experiments in a PAP need to be developed.

8 Conclusions and future work

8.1 Rationale of our work

The rationale for our work have been that experimentation is an important feature of the Swedish Armed Forces' transformation from being invasion focused to a flexible network-based defense with battle groups capable to participate in international operations. Evolva has been carried out between 2003 and 2005 with an aim to develop knowledge on tools and procedures to support this and forthcoming transformation processes with experimentation. The project has had a focus on development of military command and control but we argue that experimentation is applicable to other areas, such as effect evaluation, as well. The project has been inter-disciplinary with a mix of competences from behavioral science, engineering, and military operation analysis.

Considering the high costs of developing military systems, the experimentation process should aim at a high degree of scientific rigor and professionalism. Our work has focused on developing a ”practice” based on a scientific approach to conduct experimentation. The basis of our work has been the hypothetical-deductive approach to science as described by Hempel (1952, 1965, & 1966). Such an approach implies that the organizations' ”best guesses” about what they think would be proper solutions on identified problems should be experimentally tested. Put in the context of a development process, a series of related tests and experiments, referred to as "experimental campaigns" (Albert & Hayes, 2005), is normally necessary. Thus, development is seen as an iterative design process of formulating hypotheses, contrasting them in various environments, turning back to the initial expectations for evaluation and revision, and generating new hypotheses to be contrasted in additional simulations.

Some of the problems of experimentation considered in the project have been that the research problem is often relatively vague and not theoretically well developed. Another issue is limited time and resources making it necessary to find efficient ways to work. Finally, exercises in an experimentation campaign are often used for several different, sometimes conflicting, purposes.

We have assumed that the development process is in general initiated with some kind of problem or idea, often very vaguely formulated. In order to make the problem or idea explicit we suggest a *modeling* approach using multi-disciplinary domain competences to define hypotheses and test settings. Experimentation always includes some kind of *data collection* as data is the information gathered which is assumed to represent the characteristic of the studied phenomena. In order to be able to do this as efficient as possible, we have explored network enhanced possibilities for data collection. Another important issue has been to explore and develop alternatives to ”highly formalized laboratory experiments” and "field exercises" *as test environments*. Finally, in order to enhance *feedback of results* from experimentation we have studied alternatives to the traditional research report to transfer results to clients. In practice, we argue that the suggested process is equivalent to the term "concept development and experimentation" (CDE).

In the following sections we discuss the projects achieved goals, aspects of the scientific procedure covered by the project, and possible future work.

8.2 Achieved goals

One of the goals of the project was to generate a well documented body of research by publishing results from the studies on methods conducted in the project. During the project, and including this report, 15 different scientific publications have been presented. Most international publications are included in proceedings (see section 2.2.1). The proportion of peer-review journal publications could have been higher. It has been a matter of balancing resources between running experiments and publishing scientific documentation. The time needed for acceptation of publication in peer-review journals is normally higher compared to proceedings and FOI-reports. The consequence of increasing the number of publications in journals would probably have decreased the number of experiments. However, it must be emphasized that a peer-review procedure is always carried out prior to acceptance in proceedings and scientific FOI-reports. This was considered to be a "good enough" quality assurance of work undertaken by the project.

Another important contribution to the body of research is of course list of references in this report.

The other goal was to create an education package on research methods for officers working in development projects. The course was given the first time between the fall of 2004 and spring of 2005. The students performed well (7 of 8 students were graduated with excellence grades). A compilation of the studies performed by the students is published in Wikberg, Strangert, & Danielsson (2005a). The students feed-back on the course have been very positive. The evaluation of the first course is published in Wikberg, Strangert, & Danielsson (2005b). The students of the first course have also been requested as resources to different experimentation activities, for example the SAF demonstrator program and the NATO Viking -05 exercise.

A second run of the course is planned to start in the spring of 2006.

8.3 Modeling of experimental design

We argue that our approach to use modeling as a mean for conducting the problem analysis seems to work well. End users and other system experts can be engaged in the work of defining the research design without the procedure being to time-consuming.

However, the tools used to support the modeling process need further development. Products from the modeling of research design, such as instruments of measurements, should be operationalized and tested in practice as soon as possible. For example, instead of making a prototype of a questionnaire in MS word, the prototype should be constructed directly in digital questionnaire software (see section 6.4.1). This will enable an immediate testing of the questionnaire design. The same principle should also be applicable for other kinds of instruments of measurements such as system registrations or observer protocols. A related issue is to enhance the possibility to conduct the modeling with a minimum of face-to-face meetings in order to optimize the use of resources.

Another experience of the project is that participants of the modeling should be assigned roles in accordance with their competence in the context of a scenario. A study (Holmström, 2003) have shown that in modeling participants assigned with such a role, the participants knowledge also contribute to the final result in a significantly higher degree. Thus, an interaction structure based on roles would make modeling sessions more efficient, and would be the preferred strategy when modeling complex systems.

We also argue that the notion of "discovery experiments" by Alberts & Hayes (2002) is somewhat unfortunate. They make a distinction between 'discovery experiments', 'hypothesis testing experiments' and 'demonstrations'. Discovery experiments refer to the initial exploration of an unknown area. Hypothesis testing experiments are the classic type used by scholars to advance knowledge by seeking to falsify specific hypotheses or to discover their limiting conditions. Demonstrations are the illustration of a known fact or effect comparable to a teacher demonstrating a known law of physics. However, our experience is that the term 'discovery experiments' is often used as an excuse for "sloppy" experimentation based on an argument of the type of "*We don't know enough so we just do something and see what happens*". In practice that often means to skip the problem analysis. We argue that this is a major mistake in experimentation. It is also a misinterpretation of Alberts & Hayes intentions (2002) of the nature of the discovery experiment. They specifically point out the need of a model to guide experimentation:

> *"Failure to develop an explicit model of the problem and processes under study also makes it very difficult to run a successful experiment or campaign of experimentation. The approach that, "we'll try it and see what happens," almost invariably means that the experimentation team will have an incomplete or erroneous data collection plan. As a result, they will have a very difficult time generating meaningful empirical findings. Even discovery experiments should be supported by a simple model of what the team believes is important and the dynamics they expect to observe"* (op cit pp. 13-8).

As mentioned before, a basis for our approach to the problem analysis has been to define the clients' "best guesses" as hypotheses. Alberts & Hayes (2002) suggest the use of propositions instead of hypotheses in discovery experiments "*because the level of knowledge is not yet mature enough to state the problem as formal hypotheses*" (op cit pp. 6-2). However, this distinction between propositions and hypotheses is vague. A practical development project should be able, at least to some extent, to define the "best guesses" of what to expect as a result of their work. It is hard to imagine a developmental project that is unable to formulate such hypotheses even early in the project. If not, it might be an indication that the project has a lack of focus. To label these "best guesses" propositions or hypotheses are in practice merely a semantic matter. In the worst case, it is used as another excuse to skip the problem analysis.

8.4 Test environments

We have explored a variety of test environments mainly focused on the middle ground of the "space of test environments" illustrated in the model suggested by Gist, Hopper, & Daniels (1998) presented in chapter 5 (figure 5:1).

The results from our studies indicate that the use of commercial computer software has proven to be promising as a low-cost substitute to real military exercises. The difference in level of investment between a commercial product and tailor-made software or real exercises are substantial. Another major advantage is the insignificant need of system maintenance as new products continuously enter the market. The low cost also makes it possible for every type of unit to choose among a set of simulated environments in different software, thus tailoring the simulation to the purpose and nature of the test or exercise. As our experiments on PC games indicate, the environment is good enough as a simulator for hypothesis testing. Thus, combining simulated battle in commercial software with the use of real C2 systems is recommended as an effective and realistic enough resource setting for research on C2. As a final note, we would like to stress that we have not explored the use of commercial computer software as a tool for training. Thus, we recommend cautiousness before generalize our finding to use PC software for training.

Another conclusion rises from the fact that exercises in an experimentation campaign are often used for several different, sometimes conflicting, purposes. Large scale exercises often have to be adjusted in relation to the planned execution. Changes in scenario and procedures might of course corrupt the experimental design unless they are a part of the design. As the action research tradition is based on planned change, our experience is that this tradition is often the most applicable approach in large-scale experimentation when using an experimental strategy. Note that the recommendation is only valid in case it is possible to implement changes during a series of experimental runs within an exercise, such as during a demonstrator exercise. If a case-study approach, where an experimental strategy is to be used, the exercise must have a prioritized focus on experimentation with a minimum of changes of conditions. To adjust to this problem, we recommend a combination of data collection in larger settings and smaller case studies as described in case 7A. If an observational strategy is used both case studies and action research is applicable.

8.5 Data collection

We have explored network-enhanced technology to enable the compilation, processing, and analysis of data during ongoing exercises. As a result, vast amount of data collected during an experimental exercise must be filed, retrieved, and analyzed. The task of finding the right data in an overwhelming data set might of course be time consuming. We have used the data matrix, described in section 6.3, as a prototype to link the measurement model with collected data. We have also made some initial work together with the MARULK project on a tool to make the connection between hypotheses, data, and conclusions traceable (see figure 8:1). Problems concerning data analysis have not been fully covered by this project but we recognize the need of development of this area. The primary problem does not concern the theoretical considerations of data analysis such as statistical methods. The problem is the rather pragmatic issue to avoid being overwhelmed by the vast amount of data.

Another major restriction is bandwidth on the local network which limits the possibility to have "on-line" access to streaming data. The data collection strategy must be "selective" in order to optimize the use of available resources. To use predefined measures based on the measurement model is the principal strategy to address this problem. Another methodological solution is to use designated observers to locate relevant data. The observers use a network connected observation tool which in time and space pin-points relevant events. This tool is not yet fully functional but initial studies have been undertaken.

Finally, we would like to stress that even if data collection is digitized, the need of control for reliability and validity of data is even more present. For example, unforeseen flaws in data collection software might severely corrupt data. Precision and rigor in experimental control is as relevant as ever.

8.6 Feedback of results

In chapter 7 we have suggested a process of feedback of results from experimentation divided into a number of phases. Each phase calls for different feedback products and services. We argue that the suggested feedback process will enhance the dialogue between analysts and clients. It will be easier to define the nature of the feedback product or service depending on the phase in the experimentation cycle. This makes it easier to avoid work that does not correspond to the need of the client. A summary of the suggested feedback phases are presented in table 8:1.

Only parts of the suggested model have been tested during the course of the project. Especially, the “during experimentation” and “After Action Review” phases has been explored. Our recommendation is that representatives of the client are integrated in data collection and analysis on a full time basis during an experiment. The “After Action Review” should be presented to as many participants as possible. We do not recommend a presentation to a limited selection of participants. Another recommendation is to avoid presenting the “After Action Review” as a “briefing of facts”. The core rationale of AAR is the dialogue with participants in order to enhance interpretation of results and to collect complimentary data.

Table 8:1 *Suggested feedback phases in military experimentation*

Phases	**Purpose**	**Examples of feedback products or services**
During experimentation	Comparison between conditions. Feedback in order to make changes between runs.	1. Spreadsheets 2. Database 3. Predefined datasets 4. Refined models
After Action Review	Interpretation of results from experimentation in dialogue with participants in order to make proper conclusions. Collection of complimentary data.	1. Documentation of opinions on interpretation 2. New hypotheses 3. Preliminary results
Post Mission Analysis	Analysis of compiled data from AAR and experimentation in cooperation with domain experts.	1. Enhanced testing of hypotheses 2. New hypotheses 3. Testing of hypotheses outside of the initial plan
Experiment Conclusion	Construction of a report in order to give a final feedback of results from the experimentation.	Written documentation or oral presentation (Peer-Review paper or in clients organizational communication channel)
Lesson Learned	Compilation of experiences from previous phases to be used as a base for future interventions and actions.	Searchable and dynamic database
Plan	Planning of interventions and actions based on experiences from previous phases (and experiments).	Research design (new or revised)
Pre-Action Presentation	Implementation of interventions and actions trough a presentation to participants before next experimentation phase.	Plan for analysis

One characteristic that should be present in the feedback material is an evident link between the definitions of the problem to the interpretation of data. Interpretations of data made from a specific point of view should be obvious. If an earlier experiment has made a specific recommendation, the documentation should clearly state on which data and assumptions this recommendation is based upon. Initial studies to define a tool to enhance such documentation have been undertaken by the project MARULK, which is the implementation of the MIND framework in Swedish Armed Forces C4I Development centre, LedUtvC. These initial studies have focused on a tool to administer data related to hypotheses. The prototype of the tool is shown in figure 8.1.

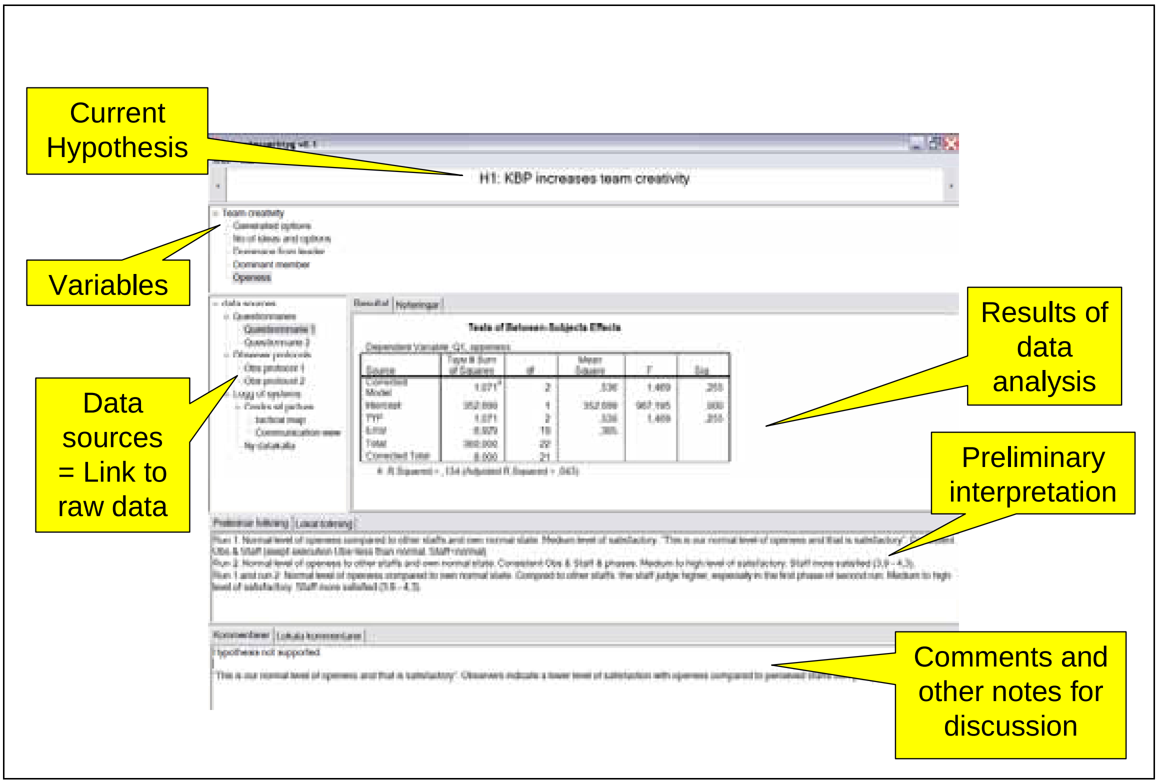


Figure 8:1. Tool to administer hypotheses.

The tool is based on a procedure which makes a distinction between data reduction, data analysis, and interpretation of results. Data reduction is the process of obtaining raw scores or characteristics from the raw data through consolidation, organization, and computing the descriptive statistics or qualitative characteristics of raw data. Data analysis refers to the computational procedure of hypothesis testing and estimation – does data support the hypothesis? Interpretation refers to the process of using the results from the data analysis to judge whether the results answer the research question and how this answer contributes to the knowledge of the problem domain.

The top window presents the hypotheses or ”best guesses” defined in the measurement model of the relevant study. The window beneath presents the variables related to the hypothesis in question. In this case the main variable is creativity which consists of a set of sub-variables. The actual data sources used to measure these variables are visualized in the window to the far left. In this case, a number of different questionnaires, observer protocols, and log files were used to measure the sub-variables of creativity. The purpose of the list of data sources is of course to create links between the data sources and the raw data in the database. In the window to the right of the data sources, preliminary results from the analysis of data are documented. In this case, data was mainly analyzed using the statistical tool SPSS. Finally, the two frames at the bottom are used to document the discussion during the interpretation phase together with complimentary information.

The idea is that the hypothesis tool will be used to organize and compile important data collected during experimentation. The compilation constitutes a ”mission history” with a traceable link between hypotheses, data, and conclusions. It also becomes the entrance to the considerably wider set of data that has been generated during the experiment.

Note that data and results used in figure 8:1 is not taken from an experiment where this tool was used. The set of data and results are just used for illustrative purpose and was chosen because the material was documented in English. The case is described in connection to figure 4:6.

8.7 Future work

We have chosen to focus our work on a few selected issues of experimentation. We have also chosen a strategy to cooperate with other relevant projects in the C2 domain. These strategic approaches have had consequences on the outcome of the project.

Data-analysis techniques and procedures are theoretically quite well developed in science and therefore excluded from the issues focused in this project. Note that we refer to data analysis as the step between data collection and interpretation, e.g. to statistically process collected data. Even if there are (or possibly needs of) development in inferential statistics, theories of nonlinear systems, and computational power of systems it was not judged to be a task for the project. Over time, as data-collection techniques have evolved, it became evident that this is an important area that needs to be developed in the future. As mentioned before, the primary problem is to avoid being overwhelmed by the vast amount of data.

Another area that needs to be looked into is the development of "standard" measures on C2. This issue has been covered by other projects. Another reason to exclude this from the project was to accentuate the need of a unique problem analysis before any experimentation. We argue that a strategy of using the same standardized set of measures in all experimentation on C2 is a dead end. However, a "library" of measures anchored in relevant theory and carefully operationalized before every unique run of experimentation might be a valuable tool to support the problem analysis.

The cooperative strategy meant that we have been able to accomplish a relatively large number of studies. As a consequence, the studied issue has varied between different experimentation activities. Aiming at developing standard methods for experimentation we found out that this was an advisable approach. The problems studied in the project include access to real-time information, automated sensor-information processing, delegation of command and control responsibility, and comparison of different ways to present a situational picture. However, the different experiments have, as compared to each other, had a nature of "one shot" trials with the only common theme of being different methodological problems. Consequently, we haven't been able to create a coherent experimental campaign on a focused C2 issue. Future work should have such an ambition.

In general, we recommend that the experiences from this project are integrated in other projects were experimentation is a mean for progress. This is of course especially relevant for projects aiming at developing C2. As experimentation methods are general in nature, we argue that the procedures are applicable to other areas as well. Further development of method experimentation should be an integrated part of the objectives of these projects. We recommend further work in two different domains:

1) Development of a more extended course on experimentation.
2) Continued methodological and technological development for efficient experimentation in distributed exercises.

8.7.1 An extended course on experimentation.

The initiated university course gave an introduction to a research method applied to development activities in the Swedish Defence Force. Our opinion is that the participants have been able put the theoretical rationale of the scientific method into practice. Compared to ordinary university students they even have performed better.

Due to the limited extent of the course it was not possible to cover all aspects of experimentation. For example, it has not been possible to cover statistical methods. In addition, the course was purely focusing on methods. No subject fields, such as social psychology or perception, were covered. However, we believe that an extended course, organized in a similar way, would be of value in order to enhance the Defence Forces competence on experimentation.

The principles for such a supplementary course have been outlined. The same concept as in the already initiated course should be used with the students having their own real study assignment to which they can apply methodological concepts and principles. Teaching should mainly be based on individual supervision. However, the focus will shift from one experiment to a series of experiments or an experimental campaign. The extended training on real experimentation will make the students more able to manage and conduct experimentation rooted in established scientific practice.

An extended course should also have an increased academic ambition. Studies of relevant subject fields should be mixed with studies on research methods. A completed course should make the students entitled to begin postgraduate studies. In practice, this will mean an additional career opportunity for officers interested in development and experimentation.

8.7.2 Methods and technology for experimentation in distributed settings

Development of methods for handling the large set of data produced during exercises and experiments will continue. Our intent has been to exploit network technologies to make experimentation more efficient. In general, we have tried to integrate this methodological and technological development in the MIND system. Development of procedures to administrate massive sets of data in distributed experimentation can and should be integrated in relevant research and development projects. As described in section 8.6, work have been initiated for extended possibilities of data analysis in the MIND system. Other examples of work already or partially in progress are:

Distributed modeling of experimental design. An ambition for future work is to enhance the possibilities to model experimental design with a minimum of face-to-face meetings.

Tool to administer hypotheses. As mentioned, initial studies to define a tool to enhance documentation of results from experimentation with a traceable link between hypotheses, data, and conclusions have been undertaken. The idea is that the hypothesis tool will be used to organize and compile important data collected during experimentation. The prototype of the tool is described in section 8.6 and shown in figure 8:1.

Streamed data. The ability to stream collected data over a network is a future functionality of central importance. There are still phases in the data-collection process that are carried out manually. Instead of filing data in local databases and later copy these manually to a central database, it is more suitable to continuously stream the data. Besides decreased amount of manual administration, it gives the opportunity to observe and work with data in real time.

Dedicated network for data collection. The use of the ordinary tactical network for data collection might interfere with the exercise due to limited bandwidth. For selected development environments, such as FM LedUtvC, a planned solution is to create a parallel network dedicated for data collection. An alternative is to expand the existing network to Gbps capacity.

Configurable data sources. Large scale experimentation involves more components such as units, staffs, individuals, and computers, than what has traditionally being handled. To manually equip each component with measurement instruments becomes problematic. Initial tests of network solutions for managing some measurement instruments, registration of voice communication, and logging of computer interfaces have been conducted. These solutions will be developed further to enable the administration of a wider set of data sources.

Object related events other than positions. Many object related events concerning individuals and units other than positions might be of interest in experimentation. Examples of such data are combat value, status, fuel, and ammunition. However, the number of object-related events possible to register has so far been limited. Our ambition is to expand this repertoire to be able to handle more types of object-related events.

System interaction. In a modern C2 system, events occurring in a number of technical systems might be of interest to register, e.g. registration of system failures. We will explore the possibilities and restrictions of using this kind of data in experimentation.

C2-process events. MIND has traditionally focused on the reconstruction of objective and concrete courses of events. As a complement, there is a need to be able to visualize more abstract processes. As of today, process-related events can only be presented in text. In a near future, a component of MIND will be developed to presents dynamic-process events. In the component model of, for an example, a planning process, a model could be presented where data on empirical events are linked to certain sub-processes of the model.

Extended tools to administrate data on communication. Channels of communication has changed and increased in recent years. Examples of ways for communication implemented in new C2 systems are text-chat, voice-chat, order-tools, e-mail, etc. In a network setting, everyone can connect to everybody else for communication, and this is the great challenge in compiling and processing communication data; i.e. to connect sender and receiver of each communication event. MIND has originally been developed to handle radio communication. Consequently, a more dynamic representation of communication is needed. All communication is now administrated in the same way even though there are many ways of communicating. The development of more powerful tools for processing communication data, for example to merge and divide groups of individuals, has also begun.

Window for non time-stamped data. MIND has traditionally supported non time-stamped data rather inadequately. Data from questionnaires, interviews, AAR, and other non time-stamped observations are important to include in documented mission histories. New components are being developed to enable the presentation and linking of this type of data to other data collected during an experimental exercise. One solution under development is to create links between different data using the Metadata tool (Albinsson et al., 2004). By clicking on a note in the Metadata tool, for example a note concerning situational awareness, references to different data relevant to problems concerning this issue are presented.

9 References

Albinsson, P.-A., Morin, M., & Thorstensson, M. (2004). Managing metadata in collaborative command and control analysis. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society,* pp. 600-604.

Albinsson, P.-A., Wong, W., Pilemalm, S., & Morin, M. (2005). Using a Reconstruction-Exploration Approach to Empower Cognitive Analysis methods: Possibilities and challenges. *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, September 26-30, Orlando, USA.

Alberts, D., & Hayes, R. (2005). *Campaigns of Experimentation: Pathways to Innovation and Transformation*. The Command and Control Research Program, USA: CCRP publication series.

Alberts, D., & Hayes, R. (2002). *Codes of best practice for experimentation.* The Command and Control Research Program, USA: CCRP publication series.

Andersson, J., Lindoff, J., & Wikberg, P. (2003). *Uppgiftsanalys av ledningsprocess – en beskrivning med operatörsfokus.* FOI MEMO 03-1545/L.

Annett, J., Duncan, D., Stammers, R. B., & Gray, M. J. (1971). *Task Analysis*. London: Her Majesty's Stationary Office.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman and Hall.

Albinsson, P.-A. (2004). *Interacting with Command and Control Systems: Tools for operators and designers.* Linköping Studies in Science and Technology, Thesis No 1132, Linköping, Sweden: Linköping University.

Albinsson, P.-A., Morin, M., & Thorstensson, M. (2004). Managing metadata in collaborative command and control analysis. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 600-604 ,September 20-24, New Orleans, USA.

Blalock, H. (1982). *Concepualization and Measurement in the Social Sciences*. London: Sage.

Brehmer, B. (2002). *Självsynkronisering: En ny Form för Kraftsamling i det Nätverksbaserade Försvaret.* Rapport Nr 3 från projektet NetWork Warfare. Swedish National Defence College, Sweden.

Brehmer, B., & Svenmarck, P. (1994). Distributed decision making in dynamic environments: Time scales and architectures of decision making. In J.-P. Caverni, M. Bar-Hillel, F. H. Barron & H. Jungermann (Red.), Contributions to decision making. pp. 155-1749. Amsterdam: Elsevier Science.

Campbell, D. T. (1969). Reforms as Experiments. *American Psychologist, 24*, 409-429.

Campbell, D. T. (1974). *Methods for Experimenting Society.* Evaluation Research Program Paper series, No 5. Evanton, Ill: Northwestern University.

Cheah, M., Thunholm, P., Chew, L. P., Wikberg, P., Andersson, J., & Danielsson, T. (2005). C^2 Team Collaboration Experiment – A Joint Research by Sweden and Singapore on Teams in a CPoF environment. *Proceedings to 10th International Command and Control Research and Technology Symposium: The Future of Command and Control,* June 13-16, McLean, VA. Command and Control Research Program (CCRP), Washington, D.C.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage can Model of Organizational Choice. Administrative Science Quartely 17, March: 1-25.

Cyert, R. M., & March, J. G. (1963). *Behavioral Theory of the Firm.* Englewoos Cliffs: NJ, Prentice-Hall.

Dawson, R.E. (1962). Simulation in the social sciences. In H. Guetzkow (Ed.), *Simulation in social sciences*. Engelwood cliffs NJ: Prentice-Hall.

Drury, C. G., Paramore, B., Van Cott, H. P., Grey, S. M., & Corlett, E. N. (1987). Task Analysis. In G. Salvendy (ed.), *Handbook of Human Factors*. New York: John Wiley & Sons.

Flood, R. L., & Carson, E. R. (1990). *Dealing with Complexity*. New York: Plenum Press.

French, W. L., & Bell, C. H. (1999). *Organization Development: Behavioural Science Interventions for Organizational Improvement.* Upper Saddle River, NJ: Prentice Hall.

Gasparski, W. W. (1991). Systems Approach as a Style: A Hermaneneutics of Systems. In M. C. Jackson, G. J. Mansell, R. L. Flood, R. B. Blackham & S. V. E. Probert. *Systems Thinking in Europe*. New York: Plenum Press.

Gavin, J. F. (1984). Survey Feedback: The Perspectives of Science and Practice. *Group and Organizational Studies*, 9, no. 1, pp. 29–70.

Gist, M. E, Hopper, H., & Daniels, D (1998). Behavioural Simulation: Application and Potential in Management Research. *Organizational Research methods*, Vol 1. No 3 July 1998 pp. 251-295.

Graziano, A. M, & Raulin, M. L. (2004). *Research methods. A process of inquiry.* 3rd ed. New York: Longman.

Harré, R. (2002). *Cognitive Science: A philosophical introduction*. London: Sage.

Harvey, R. J. (1991). Job Analysis. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of Industrial and Organisational Psychology 2 ed., vol. 2.* (pp. 71-163). Palo Alto, CA: Consulting Psychologists Press, Inc.

Hempel, C. G. (1966). *Philosophy of Natural Science*. Englewood Cliffs NJ: Prentice Hall, Inc.

Hempel, C. G. (1965). *Aspects of Scientific Explanations*. New York: The Free Press-MacMillan.

Hempel, C. G. (1952). Fundamentals of concept formation in empirical science. *International encyclopedia of unified science*, Vol. II (pp. 20-62). Chicago: University of Chicago Press.

Hollnagel, E., & Woods, D. D., (1983). Cognitive Systems Engineering: New Wine in New Bottles. *International Journal of Man-Machine Studies*, 18, pp. 583-600.

Holmström, H. (2003). Modeling of complex systems: Role demands and interactions. *Proceedings of 11th European Congress on work and organizational Psychology (EAWOP)*, 14-17 Maj 2003, Lissabon.

Jenvald, J., Crissey, M. J., Morin, M., & Thorstensson, M. (2002). Training Novice Observers to Monitor Simulation Exercises. *In Proceedings of The 13th International Training and Education Conference (ITEC 2002)*, p. 68–78, 9–11 April, Lille, France.

Katz, D., & Kahn, R. L. (1978). *The Social Psychology of Organizations.* New York: John Wiley.

Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, CT: Ablex.

Kolbe, R. H. (1991). Content analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research*, 18, 243-250.

Krantz, D., Luce, R.D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement. Vol. 1*. San Diego: Academic Press. pp. 1-33.

Llewellyn, K. N. (1948). Case Methods. In E. Seligman & A. Johnson (Eds.), *Encyclopedia of the Social Sciences*. New York: Macmillan.

Lewin, K. (1952). Group Decision and Social Change. In G. E. Swanson, T. N. Newcomb, & E. L. Hartley (Eds.). *Reading in Social Psychology*, rev. ed. New York: Holt.

Miller, D., & Starr, M. (1967). *The Structure of Human Decisions.* Englewoos Cliffs: NJ, Prentice-Hall.

Morin, M. (2002). *Multimedia representations of distributed tactical operations*. Linköping Studies in Science and Technology: PhD Dissertation No 771, Linköping, Sweden.

Morin, M., Jenvald, J., & Thorstensson, M. (2003). Utvecklingsmetoder för samhällsförsvaret. FOI user report: FOI-R--1064—SE.

Morrison, J. E., & Meliza, L. L. (1999). *Foundations of the After Action Review Process*. Special Report 42. U.S. Army Research Institute: Alexandria, Virginia.

Mårtensson, B., & Nilstun, T. (1988). *Praktisk Vetenskapsteori.* Studentlitteratur: Lund.

Popper, K. R. (1959). The Logic of Scientific Discovery. New York, Basic Books.

Rankin, W. J., Gentner, F. C., & Crissey, M. J. (1995). After Action Review and Debriefing Methods: Technique and Technology. In *Proceedings of the 17th Interservice / Industry Training Systems and Education Conference*, pp. 252–261, Albuquerque, New Mexico, USA.

Rasmussen, J. (1993). Analysis of Tasks, Activities and Work in the Field and in Laboratories. *Le Travail Humain*, 56, pp. 133-155.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: Wiley.

Reichenbach, H. (1938). Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge. *The University of Chicago Press*, Chicago, Illinois, USA.

Sandberg, L., & Sandberg, R. (2002). *Elementär statistik.* Stockholm: Studentlitteratur.

Santalainen, T. J., & Hunt, J. G. (1988). Change Differences from an Action Research, Results-Oriented OD Program in high- and Low-Performing Banks. *Group and Organizational Studies*, 13, no. 4, pp. 314–440.

Schultz, R. L., & Sullivan, E. M. (1972). Development in simulation in social and administrative science. In H. Guetzkow, P. Kotler, & Schultz, R. L. (Eds.) *Simulation in social and administrative science.* (pp. 3-47). Engelwood cliffs NJ: Prentice-Hall.

Shani, A. B., & Eberhardt, B. J. (1987). Parallel Organization in a Health Care Institution. *Group and Organizational Studies*, 12, no. 2, pp. 147–173.

Singleton, W. T. (ed.) (1979). The Study of Real Skills. *Compliance and Excellence* Vol. 2. Lancaster: MTP.

Silverman, D. (2001). *Interpreting qualitative data. Methods for analyzing talk, text and interaction.* 2nd ed. London: Sage.

Simon, H. A. (1957). *Administrative Behavior*. New York: The Free Press.

Skinner, B. (1953). Science and human behavior. New York: MacMillan.

Skinner, B. (1968). The technology of teaching. New York: Appleton-Crofts.

SPSS. (2005). *Statistical Package for the Social Sciences.* SPSS Inc. http://www.spss.com.

Staw, B. (1983). The Experimenting Organization: Problems and Prospects. In B. Staw (Ed.) *Pshychological Foundations of Organizational Behavior 2nd ed.* Glenview Ill: Scott, Foresman and Company.

Stein, H. (1952). Case Method and the Analysis and the Analysis of Public Administration. In H. Stein (Ed.), *Public Administration and Policy Development*. New York: Harcourt Brace Jovaniovich.

Stevens, S. S. (1951). Marthematics, Measuremnt, and Psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York & London: Basic Books.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, 103,* pp. 677-680.

Strangert, B. (2005). *Kompendium i försöksmetodik för utvecklingsansvariga i Försvarsmakten – förhandsutgåva*. FOI-MEMO 1429.

Swedish Defence Forces (2002). *Målsättning för Systemdemonstratorerna 2005 och 2006.* HKV Beteckning 09 100:67492.

Swedish Defence Forces (2005). Interoperabilitet avseende ledningssystem – C KRI Beslut i Stort (BIS) – Tillägg 1. HKV Beteckning 09100:64557.

Thorstensson, M. (1997). Structured reports for manual observations in team training. M. Sc. Thesis LiTH-IDA-Ex-97/64, Linköpings universitet, Linköping.

Thunholm, P. (2003). *Military Decision Making and Planning: Towards a New Prescriptive Model.* Doctoral thesis, Stockholm University. Sweden, Edsbruk: Akademitryck.

Thunholm, P. (2005). Planning Under Time-pressure: An attempt Toward a Prescriptive Model of Military Tactical Decision Making. In H. Montgomery, R. Lipshitz, & B. Brehmer (Eds.), *How Experts Make Decisions*. New Jersey: Lawrence Erlbaum.

Towl, A. R. (1969). *To Study Administrations by Cases.* Boston: Harvard University Business School.

U.S. Army Combined Arms Center. (1993). *A leader's guide to after-action reviews* (Training Circular 25-20). Fort Leavenworth, KS: Author. Retrieved November 18, 1999, from the World Wide Web: http://call.army.mil/products/trngqtr/tq1-98/table.htm.

Weber, R. P. (1990). *Basic Content Analysis*. 2nd ed. Beverly-Hills, CA: Sage.

Weisbord, M. R. (1987a). Toward Third-Wave Managing and Consulting. *Organizational Dynamics*, pp. 5–24.

Wikberg, P., Strangert, B., & Danielsson, T. (2005a). *Utvärdering av kurs i försöksmetodik för utvecklingsansvariga inom Försvarsmakten – 10 poäng.* FOI MEMO-1427.

Wikberg, P., Strangert, B., & Danielsson, T. (2005b). *Kurs i försöksmetodik för utvecklings-ansvariga inom Försvarsmakten 2004-2005. Kursens uppläggning och genomförda försök.* FOI MEMO-1428.

Wikberg, P., Danielsson, T., & Holmström, H. (2005). A Study of the Effect of a Decision Support System with Automated Data Analysis. *Proceedings of 12th Congress of European Association of Work amd Organisational Psychology*, May 12-15,2005, Istanbul, Turkey.

Wikberg, P., Adolfsson, M, Strangert, B., & Holmström, H. (2005). *Beslutsprocesser för ledning. Beredningsbeslut, förenklade beslut, samverkansbeslut samt insatsledning.* Underlagsrapport, FOI-R--1644—SE.

Wikberg, P., Strangert, B., & Danielsson, T. (2005a). *Kurs i försöksmetodik för utvecklingsansvariga inom Försvarsmakten 2004-2005.* Kursens uppläggning och genomförda försök. FOI MEMO-1428.

Wikberg, P., Albinsson, P-A., Andersson, D., Danielsson, T., Holmström, H., Johansson, M., Thorstensson, M., & Wulff, M-E. (2005). *Feedback of results from experimentation in development projects.* FOI MEMO-1365.

Wikberg, P., Lundin, J., Albinsson, P.-A., Danielsson, T., Holmström, H., Johansson, M., & Östensson, M. (2004). *Dokumentation från workshop om After Action Review (AAR) av ledningsprocesser 040322-040326.* FOI MEMO-915.

Wikberg, P., Strangert, B., & Danielsson, T. (2004). *Halvtidsutvärdering av kurs i försöksmetodik för utvecklingsansvariga inom Försvarsmakten – 10 poäng.* FOI MEMO-1105.

Wikberg, P., Andersson, J., Berggren, J., Hedström, J., Lindoff, J., Rencrantz, C., Thorstensson, M., & Holmström, H. (2004). *Simulerade insatsmiljöer i kommersiella PC-spel som försöksplattformar. Skillnader mellan genomförande av jägarinsats i en virtuell miljö med ett genomförande i riktig terräng med lös ammunition och simfire-utrustning.* Swedish Defence Research Agency, FOI-R--1416--SE.

Wikberg, P., Hasewinkel, H., Lindoff, J., Stjernberger, J., & Eriksson, L. (2004). Game Based Simulation: Using Commercial Game Software in a Military Command and Control Exercise. *Proceedings of 28th International Congress of Psychology*, pp. 1287, August 8-13, 2004, Beijing, China.

Wikberg, P., Albinsson, P.-A., Thorstensson, M., Stjernberger, J., & Holmstrom, H. (2004). Methodological Aspects of Development and Evaluation of Military Command and Control Systems. *Proceedings of 28th International Congress of Psychology*, pp. 263, August 8-13, 2004, Beijing, China.

Wikberg, P. (2004). *Försöksmiljöer för utveckling av nätverksbaserad ledning.* FOI MEMO-840.

Wikberg, P., & Holmström, H. (2003). Experimental Simulation of Command and Control Processes. *Proceedings of Human Factors of Decision making in Complex Systems.* 8-11 September 2003, Dunblane.

Wikberg, P., & Hasewinkel, H. (2003). Spelhåla i jägarstab visar chefens rätta jag. *Framsyn*, 2003:6, pp. 22-23.

Wikberg, P., & Thorstensson, M. (2003). Rapid Evaluation of Complex and Distributed Systems. In *Proceedings of the 11th European Congress on Work and Organisational Psychology*. May 14-17, 2003 Lisbon, Portugal.

Wikberg, P., & Lundin, J. (2003). Experimental Simulation of Complex Social Systems. In *Proceedings of the 11th European Congress on Work and Organisational Psychology*. May 14-17, 2003 Lisbon, Portugal.

Wikberg, P., Albinsson, P.-A., Thorstensson, M., Stjernberger, J., Holmstrom, H., & Östensson, M. (2003). *Methodological Aspects of Development and Evaluation of Military Control.* FOI scientific report: FOI-R--1034—SE.

Wikberg, P., Hasewinkel, H., Lindoff, J., Stjernberger, J., Eriksson, L., & Persson, A. (2003). *Kommersiella PC-spel som försöks- och övningsmiljö: Prövning av jägarbataljons ledningsmetod vid tillgång till realtidsinformation.* FOI-R--0989—SE.

Wikberg, P., & Lundin, J. (2002a). Perception and Interpretation of Statistical Results in After Action Reviews. *In Proceedings of the 2002 Hawaii International Conference on Statistics and Related Fields*. June 5–9, 2002, Honolulu, Hawaii.

Wikberg, P, & Lundin, J. (2002) *Grafiska rapporter.* Underlagsrapport, FOI-R--0487—SE.

Wikberg, P. (1997). *Using Expert Descriptions in Systems Analysis*. Umeå: Department of Psychology, Umeå University, Sweden.

Windsor, D., & Greanias, G. (1983). The Public Policy and Management Program for Case/Course Development. *Public Administration Review*, 26, pp. 370–378.

Yin, R. (1984). *Case Study Research. Design and Methods. Applied Social Science Research Methods Vol. 5.* Beverly Hills: Sage Publications.

Yin, R. (1993). *Applications of Case Study Researchs. Applied Social Science Research Methods Vol. 34.* Beverly Hills: Sage Publications.

Appendix 1. The project's publications

1 International publications

Albinsson, P.-A., Wong, W., Pilemalm, S., & Morin, M. (2005). Using a Reconstruction-Exploration Approach to Empower Cognitive Analysis methods: Possibilities and challenges. *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, September 26-30, Orlando, USA.

Albinsson, P.-A. (2004). *Interacting with Command and Control Systems: Tools for operators and designers.* Linköping Studies in Science and Technology, Thesis No 1132, Linköping, Sweden: Linköping University.

Albinsson, P.-A., Morin, M., & Thorstensson, M. (2004). Managing metadata in collaborative command and control analysis. In *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 600-604, September 20-24, New Orleans, USA.

Cheah, M., Thunholm, P., Chew, L. P., Wikberg, P., Andersson, J., & Danielsson, T. (2005). Command and Control Team Collaboration Experiment – A Joint Research by Sweden and Singapore on Teams in a CPoF environment. *Proceedings to 10th International Command and Control Research and Technology Symposium: The Future of Command and Control,* June 13-16, McLean, VA. Command and Control Research Program (CCRP), Washington, D.C.

Holmström, H. (2003). Modeling of complex systems: Role demands and interactions. *Proceedings of 11th European Congress on work and organizational Psychology (EAWOP)*, 14-17 May 2003, Lisbon, Portugal.

Wikberg, P., & Holmström, H. (2003). Experimental Simulation of Command and Control Processes. *Proceedings of Human Factors of Decision making in Complex Systems.* 8-11 September 2003, Dunblane, Scotland.

Wikberg, P., & Lundin, J. (2003). Experimental Simulation of Complex Social Systems. *Proceedings of 11th European Congress on Work and Organizational Psychology.* May 14-17, 2003 Lisbon, Portugal.

Wikberg, P., & Thorstensson, M. (2003). Rapid Evaluation of Complex and Distributed Systems. *Proceedings of 11th European Congress on Work and Organizational Psychology*. May 14-17, 2003 Lisbon, Portugal.

Wikberg, P., Hasewinkel, H., Lindoff, J., Stjernberger, J., & Eriksson, L. (2004). Game Based Simulation: Using Commercial Game Software in a Military Command and Control Exercise. *Proceedings of 28th International Congress of Psychology*, pp. 1287, August 8-13, 2004, Beijing, China.

Wikberg, P., Albinsson, P.-A., Thorstensson, M., Stjernberger, J., & Holmström, H. (2004). Methodological Aspects of Development and Evaluation of Military Command and Control Systems. *Proceedings of 28th International Congress of Psychology*, pp. 263, August 8-13, 2004, Beijing, China.

Wikberg, P., Danielsson, T., & Holmström, H. (2005). A Study of the Effect of a Decision Support System with Automated Data Analysis. *Proceedings of 12th Congress of European Association of Work amd Organisational Psychology*, May 12-15, 2005, Istanbul, Turkey.

2 Scientific reports published at FOI

Wikberg, P., Hasewinkel, H., Lindoff, J., Stjernberger, J., Eriksson, L., & Persson, A. (2003). *Kommersiella PC-spel som försöks- och övningsmiljö: Prövning av jägarbataljons ledningsmetod vid tillgång till realtidsinformation.* FOI-R--0989--SE.

Wikberg, P., Albinsson, P.-A., Thorstensson, M., Stjernberger, J., Holmström, H., & Östensson, M. (2003). *Methodological Aspects of Development and Evaluation of Military Control.* FOI: FOI-R--1034—SE.

Wikberg, P., Andersson, J., Berggren, J., Hedström, J., Lindoff, J., Rencrantz, C., Thorstensson, M., & Holmström, H. (2004). *Simulerade insatsmiljöer i kommersiella PC-spel som försöksplattformar. Skillnader mellan genomförande av jägarinsats i en virtuell miljö med ett genomförande i riktig terräng med lös ammunition och simfire-utrustning.* FOI-R--1416--SE.

3 Memos published at FOI

Wikberg, P., Albinsson, P-A., Andersson, D., Danielsson, T., Holmström, H., Johansson, M., Thorstensson, M., & Wulff, M-E. (2005). *Feedback of results from experimentation in development projects.* FOI MEMO-1365

Wikberg, P. (2004). *Försöksmiljöer för utveckling av nätverksbaserad ledning.* FOI MEMO-840.

Wikberg, P., Lundin, J., Albinsson, P-A., Danielsson, T., Holmström., H. Johansson, M., & Östensson, M. (2004). *Dokumentation från workshop om After Action Review (AAR) av ledningsprocesser 040322-040326.* FOI MEMO-915

Wikberg, P., Strangert, B., & Danielsson, T. (2004). *Halvtidsutvärdering av kurs i försöksmetodik för utvecklingsansvariga inom Försvarsmakten – 10 poäng.* FOI MEMO-1105.

4 Other articles

Wikberg, P., & Hasewinkel, H. (2003). Spelhåla i jägarstab visar chefens rätta jag. *Framsyn*, 2003:6, pp. 22-23.