15th ICCRTS

"The Evolution of C2"

A Microworld Study of Task Force Commanders Executing a Maritime Escort Mission

Topic 5: Experimentation and Analysis

Author and Point of Contact:
Christofer Waldenström
STUDENT
Swedish National Defense College
Department of Military Studies
Division for Command and Control Studies
PO Box 27805
SE-115 93 Stockholm
Sweden
christofer.waldenstrom@fhs.se

A Microworld Study of Task Force Commanders Executing a Maritime Escort Mission

Christofer Waldenström
Swedish National Defense College
Department of Military Studies
Division for Command and Control Studies
christofer.waldenstrom@fhs.se

This paper presents an exploratory microworld study with the aim to identify individual differences between participants, and relate those difference to how well the participant solves the task. Six officers, rank from lieutenant commander to flotilla admiral, were studied when they commanded a maritime escort mission. The experiment was conducted using a microworld where the participant had to control all own units while the computer controlled enemy and neutral units. Data collection consisted of think-aloud protocols, screen captures of the microworld's tactical screen, questionnaires, and battle outcomes. Performance was determined using a measure of mission success and a general model of the participants' decision making process was constructed. This model was used to identify individual differences and relate those to task performance. The results suggest that there is no correlation between how often the participants perform a certain decision making activity, and how well they perform in the microworld. On the other hand, the results suggest a strong correlation between how well the participants perform in the microworld and how many different decision making activities they visit during one coherent reasoning chain. The result seems to suggest that it is more important to consider many aspects of a problem at the same time, and that no decision making activity is more important that another.

Command and control (C2) systems can be viewed as artifacts, and as such they are best understood in the terms of the logic they were constructed: the logic of design (Brehmer, 2007). This approach requires that the system is analyzed in terms of its purpose, its functions, and its form. The system's purpose describes *why* the system exists, and in the case of C2 it is to produce the military effects by providing direction and coordination of the military effort. The function describe *what* must be accomplished by the system in order to fulfill its purpose, and here Brehmer (2006, 2005) proposes *data collection*, *sensemaking*, and *planning* as the necessary functions. The final step in the design is to settle on the system's form i.e. describe *how* the functions should be implemented. The form is the embodiment of the C2-system and is comprised of organizations[1], procedures, and support systems. From the view of design logic, form is normative and describes the designer's idea of an optimal C2-system. However, the actual workings of a C2-system (the C2-process) often deviate from the designer's intention. This deviation may lead to worse or better performance than what could be predicted from form alone. Thus, improvements may come from bringing process closer to form, e.g. by training, or by bringing form closer to process, e.g. by changing procedures (see for example Schmitt & Klein, 1999; Thunholm, 2003). However, more radical improvements will likely come if we find better ways to fulfill the requisite functions (Brehmer, 2007). In either case, if we we want to improve C2 we need to study the C2-process. Only then can we identify if process matches form and how well form fulfills the functions. Thus, this paper presents an exploratory microworld study of experienced navy officers executing a maritime escort mission.

Military manuals point to command as the first element of C2. Command "includes the authority and responsibility for effectively using available resources and for planning the employment of, organizing, directing, coordinating and controlling military forces for the accomplishment of assigned missions" (US Navy, 1995, p. 7). The major process of command has traditionally been the military planning process, which has rendered detailed descriptions in the doctrines (US Navy, 2007; Swedish Armed Forces, 1987). The second element of C2 is control, and here the commander monitors and influences the action to accomplish what must be done. Feedback is vital to control since it gives "a way to monitor events, adapt to changing circumstances, adjust allocation of resources, and harmonize the efforts of

This work was funded by Swedish Armed Forces Doctoral Program

[1] It should be noted that an organization contains roles, and that each role requires a certain competence. Thus, how well the personnel match these competence requirements of course have impact on how well the C2-system can perform.

the force" (US Navy, 1995, p. 9). Consequently, control has been related to the process of execution, but in contrast to planning, execution has not been described to the same level of detail as planning has in the manuals.

C2 can also be conceived as a controller of a military operation, as illustrated by the many of cybernetic models proposed to describe it (see Levis & Athans, 1988, or for later examples see the DOODA-loop by Brehmer, 2006, or the CECA-model by Bryant, 2003). Being a controller, it is important to notice that the preconditions for control are the same for the C2-function as it is for any other controller. It must have a goal to steer the system towards. It must be able to both observe and affect the state of the system. It must possess a model of the system that tells the controller which actions leads to what states, so that an appropriate course of action can be selected. Thus, when studying the C2-process we must study the human when he is set to control a dynamic system.

The human as a controller has been studied within the field of dynamic decision making (or complex problemsolving, as it is also called, see Frensch & Funke, 1995). What differentiates a dynamic decision task from a traditional decision making task is that a series of decisions have to be made, where the decisions are not independent of each other. Further, the state of the decision task changes, both as a consequence of the decision makers actions but also autonomously. Finally, time is important. (Brehmer, 2000).

To investigate how people solve dynamic decision tasks researchers have studied them while they interact with a computer simulation of some system, a so called *microworld*. Besides being dynamic, the microworld presents the controller with two further difficulties. First, in the microworld many variables are interconnected which make it impossible to do only one thing. All actions have side effects and these must be taken into account when trying to achieve the main effect. Second, the microworld is opaque. This means that the controller cannot directly ascertain the state of the microworld, but has to take actions in order to do that. The interest of researchers have been to explore the extent to which people can learn how to control the complex task presented by a microworld, i.e. how people develop a model of the complex system and to what extent they are able to do so.

In order to succeed in the dynamic task the participant must handle the difficulties posed by the microworld (for a more extensive review the demands a microworld puts on a participant see for example Brehmer & Dörner, 1993; Dörner & Wearing, 1995). First, the participants must engage in goal analysis. In many of the microworlds the goals are on a global and abstract level and these goals must be expressed in set of more explicit subgoals. If there is conflict between the goals the subject must find ways to resolve those. Furthermore, if there exists several goals, and the importance of those goals changes over time, the participant must be able to prioritize goals and change these priorities as the situation demands. Second, the participant must learn about the microworld. Even if some parts of the microworld are evident to the participants some parts of the system are unknown to them. So, in order to act, participants must gather information about the microworld and form hypotheses about the hidden structure. Third, the participants must make predictions concerning the microworld's future development. This is necessary because the inertia of the system - primarily the dead-time and time-constant - creates a delay between the time a participant orders an action and the time that action starts to take effect. Fourth, the participant must develop courses of action and implement them. Fifth, they must monitor the effects of the selected course of action and the monitor the development of the microworld. Not a trivial task, since the participant must determine what effects are the results of the selected course of action, and what effects are the result of the the microworld's own development. Sixth, the participants must monitor and evaluate their own strategies employed to solve the activities presented here. Finally, they must organize all these activities in to "some coherent whole" (Brehmer & Dörner, 1993, p. 181). For example, they must decide when they should start planning instead of collecting information.

In microworld research three different approaches can be distinguished: the individual differences approach, the case-study approach, and the experimental approach (Brehmer & Dörner, 1993). In the individual differences approach participants interact with the same microworld and the researcher is interested in what differentiates 'good controllers' from 'bad controllers' (see for example Dörner, 1980). In the case study approach the purpose is to generate hypotheses of how people deal with complexity and uncertainty. The experimenter examines the behavior of the participants and tries to identify what is typical and what is atypical. In the experimental approach properties of the microworld are systematically manipulated to determine how these manipulations influence participants performance (see for example Brehmer & Nählinder, 2007).

The experiments with microworlds have revealed a collection of errors that people commit. These errors are of great interest since they give insight to the limitations people have when coping with complex systems (the list is compiled from Brehmer & Dörner, 1993, and Dörner & Wearing, 1995):

1. People have problem understanding development tendencies if they receive data about these tendencies as isolated events in time. Exponential growth is seen as linear, oscillations are interpreted as chaos.
2. People have difficulties in taking into account the distant effects and side effects of their actions
3. People tend to focus on the 'here and now'. What will happen in the future or what has led the participant to the current situation is of less interest.
4. People have problems in coping with the delayed feedback concerning the results of their actions.
5. In crisis situations the participants adopt a "fire and forget" attitude and do not check the effect of their actions.
6. When forming hypotheses people tend to reduce them to only one or a few variables which they believe are of central importance.

7. People have problems in combining many courses of actions into one. This is important when many subgoals must be met simultaneously. Instead people carry out each course of action and do not take the effects they have on each other into account.
8. Under some conditions people have problems in choosing the appropriate level of detail of their plans. Planning with a low level of detail ends quickly because conditions and consequences for action are considered only superficially. Planning with a high level of detail never comes to an end.

So, what can we expect to observe from the participants as they interact with the microworld in this study? Well, first of all they will engage in goal analysis. Both for the overall escort mission, but also for subgoals that emerges during execution. As the resources are limited, conflicts will probably arise: should I use these resources for anti surface warfare or for anti submarine warfare? As people have problems in constructing a course of action that satisfy several goals at the same time it can be expected that such goal conflicts will probably pose problems to the participants. Evidence for this would be that the participants devote time to engage in problem solving when identifying such conflict.

The microworld used in this experiment is designed to include the most important aspects of naval warfare, so when it comes to learning about the microworld it can be expected that the participant will probe into how well it corresponds to what they know from experience. What is opaque to the participant, however, is of course the enemy, and as understanding the enemy is considered the most difficult part of C2 (Waldenström, 2008) much effort will probably be invested in this activity.

Furthermore, the participant will probably only receive intermittent information about the enemy's position, so he or she must build hypotheses about where the enemy is and what the enemy is going to do. Since people have problem understanding development tendencies based on piecemeal information, how the participants build and maintain such hypotheses about the enemy is of great interest. Problems here will have major influence on action. Further, peoples tendency to simplify hypotheses also bear on this problem.

Signs of problems with creating and maintaining hypotheses could be that the participants have problems in organizing the different enemy reports into plausible enemy courses of action, or that the participant switches back and forth in his or her belief of what a detected enemy is going to do. Both these cases will make it difficult for the participant to plan and initiate action to counter the enemy. Other signs could be that the participant creates an impossible hypothesis given the data at hand. For example, given that the enemy is spotted at two different locations with some time between them, the participant creates the hypothesis that it is the same enemy unit, despite the fact that it is impossible for that unit to move such distance given the time and the unit's maximum speed.

In order to coordinate own actions, and handle the inertia of the system, the participants must try to predict the development of the operation, so actions can be initiated in due time. This requires the participant to maintain at least two predictive models: (1) a model of what own units will do in the future and how long that will take, and (2) a model of what the enemy will do in the future and how long that will take. The participant will constantly receive information about the own unit's positions, so signs of problem in maintaining the first model would be that the participant makes estimation errors when it comes to planning movement of own units i.e. they arrive earlier or later than expected. The second model is assumed to be more difficult for the participant to maintain since it requires at least two things: First, the participant must use the intermittent enemy observations and create an understanding of where the enemy is and what the enemy is going to do, as described above. Second, given this understanding the participant must estimate the future enemy movement. Problems in maintaining these models can lead to situations where the participants find himself or herself being 'to late'. This might force the participant to make late minute changes to counter an unexpected threat, or high risk taking when it comes to movement.

As execution is in large feed-back driven it can be expected that the participants will spend a lot of time monitoring the situation. The purpose of monitoring is to "adapt to changing circumstances, adjust the allocation of forces, and harmonize the effects of the force"(US Navy, 1995, p. 9). The participant will have to identify emerging problems and devise proper action to counter those problems i.e. identify threats and devise action to control that threat (an overview of the principal elements in this activity is described in Waldenström, Ekenberg, & Danielsson, 2009). Problems in identifying threats can lead to situations where the enemy unexpectedly engages the participant.

Method

Participants

The participants were acting and retired officers from the Swedish navy, one lieutenant commander, three commanders, one captain and one flotilla admiral, who were either paid or given two movie tickets for participating. Their mean age was 53 (min=40, max=65) and their mean years of service were 31 (min=21, max=40).

Design

This study was designed as an individual differences study with the purpose of investigating if differences in decision making processes led to differences in performance. All participants solved the same scenario and performance was determined using a measure of mission success. A general model of the participants' decision making process was constructed. This model was used to identify individual differences and relate those to task performance.

Microworld

The microworld used in the experiment was the Simple Surface warfare Model (SSM) in which the participant assumes the role of a navy commander located at his or her

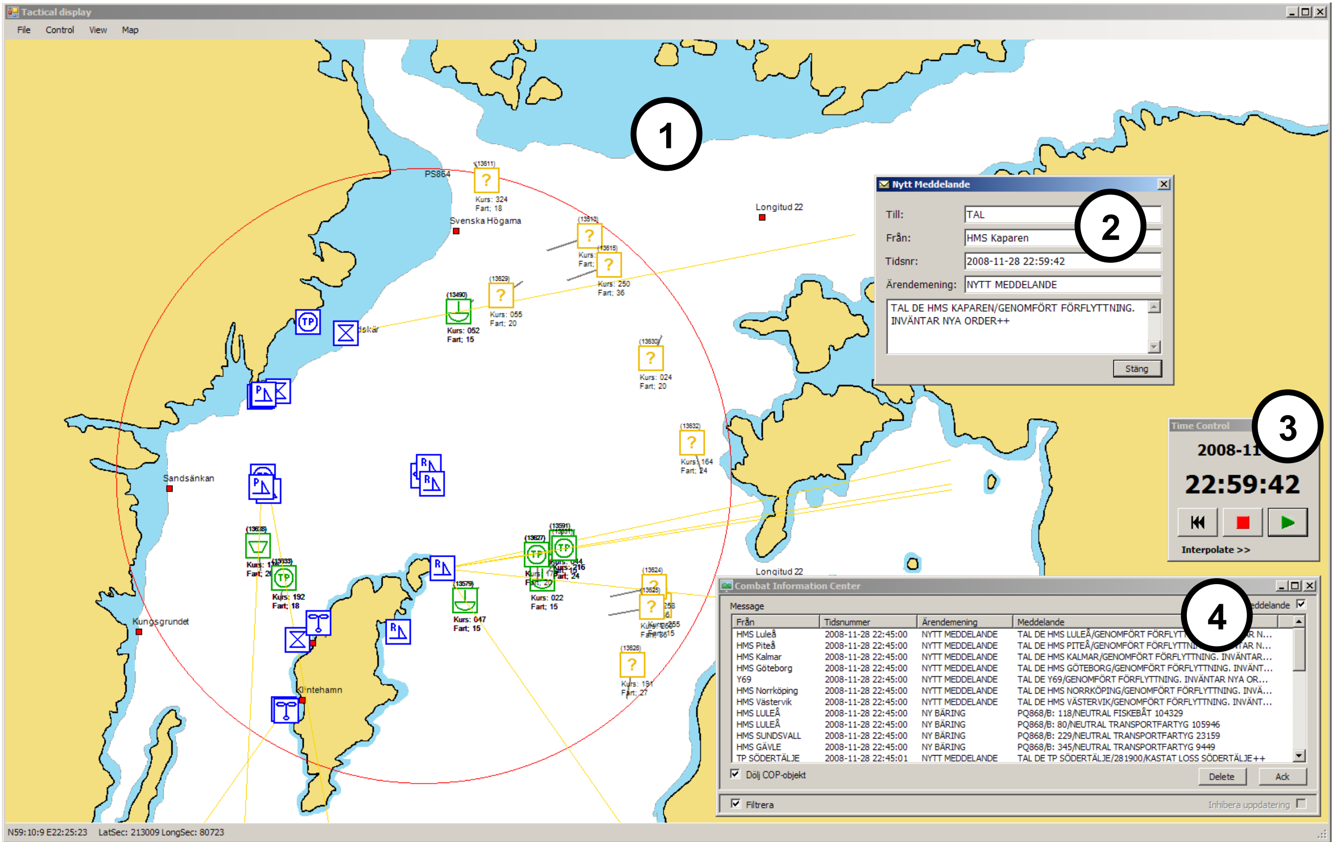

Figure 1. The main screen of the microworld. (1) Map of the mission area. (2) Messages sent by own units. (3) Time control. (4) Message list

command post. The overall task that the participant has to solve depends on the scenario, but what the participant does in the microworld is to give orders to own units, either to go somewhere, survey its surroundings with a sensor, or fire a weapon. The results of these actions are reported back to the commander, who then must make new decisions based on this new information. This continues until the overall missions is either solved or failed.

The main screen of the microworld is shown in figure 1. The central part of the main screen is the *tactical map* (1). On this map all own units, together with a common operational picture (COP) is displayed. The COP is vital to the commander because it is a representation of everything the own force detects with its sensors at any given moment[2]. The COP is created by merging data provided by the sensors on all own units, and its functioning is best illustrated by an example. In figure 2(a) a symbol representing an own unit is displayed: OWN Fast Patrol Boat. The line extending from the symbol represents that the unit has detected another unit with a sensor that provides information about the direction to the detected unit and its class, in this case an 'Enemy Attack Corvette'. In figure 2(b) another unit in the force has detected the same enemy unit with a sensor similar as the one of the first unit. The COP-logic now automatically creates a symbol representing the detected unit and places it at the intersection of the two lines. Another way to obtain the class and position of an enemy unit is displayed in figure 2(c) and figure 2(d). In figure 2(c) an own unit detects a unit with a sensor that provides information about position, heading and speed (e.g. a radar). If this information is combined with the information obtained from a sensor providing information about direction and class, the COP-logic produces a symbol of the enemy unit as illustrated in figure 2(d). At any moment, the COP is the best view of the operations area that can be created by the sensors in the own force, and if the commander wants to identify any unidentified targets or survey an area not yet covered, he or she must dispatch units with appropriate sensors to do this.

The *time control* (3) in figure 1 controls the time in the microworld. The time can be started and stopped as the participant wants and it can be made to run at any speed up to a factor of 30, i.e. thirty simulated seconds every real second.

All reports from the units are displayed in a dedicated *message window* as illustrated in figure 1 (2). Whenever a message is displayed the time is stopped to give the participant time to read the message. When the participant has read the message it is closed by clicking on it and the time starts running again. In order to not "flood" the participant with messages regarding the COP, only the first contact with

[2] Actually everything the own force has detected within a certain time interval

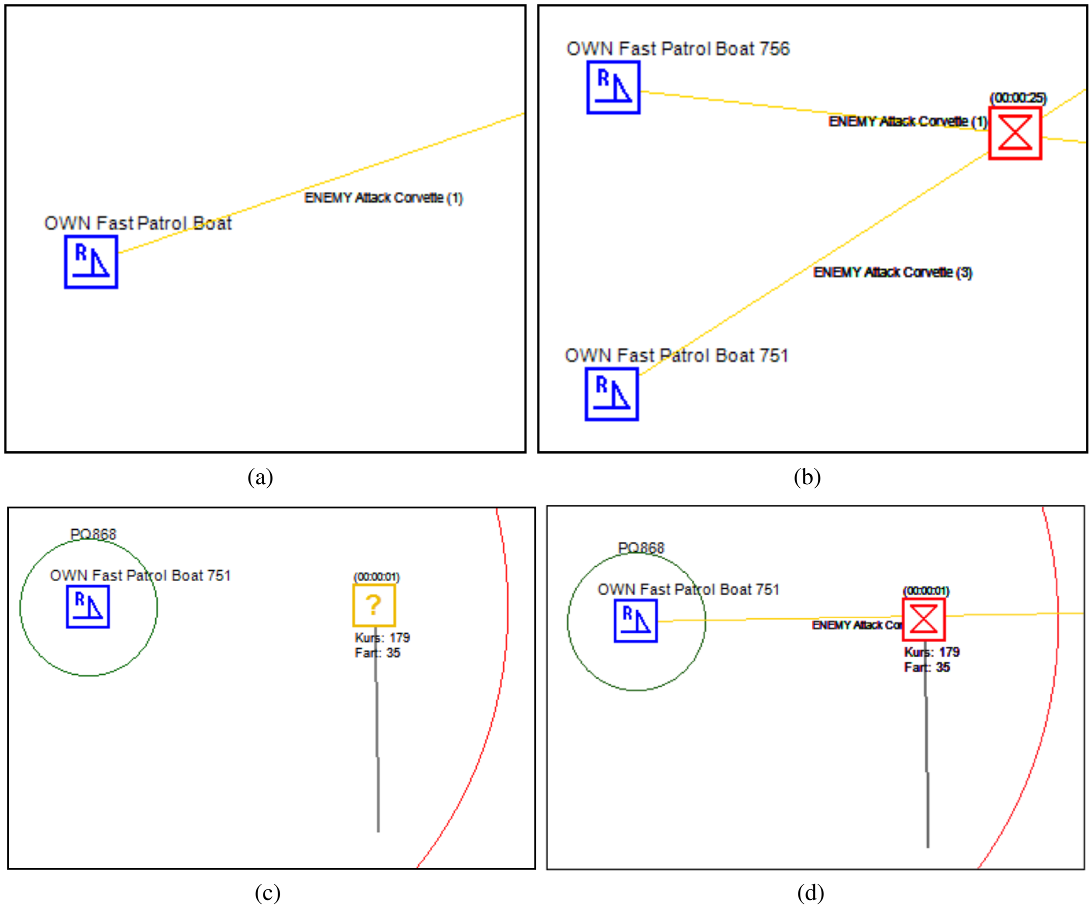

Figure 2. Creating the common operational picture

a new unit is displayed as a message. All further updates of that contact is done automatically by the COP-logic until no unit in the force has contact with that unit anymore. At this point a message is displayed to inform the participant that the force has lost contact with that target. All messages that have been displayed are stored in a message list. This gives the participant the opportunity to go back and look at old messages if needed.

To give orders to the own units, one clicks on the unit's symbol. This brings up a menu that contains the actions that can be performed by the unit. From this menu the participant can issue movement orders, sensors can be turned on or off, and weapons can be fired. It is also possible to display the ranges of all sensors as well as the ranges of all weapons. The computer controls all enemy and neutral units. Neutral units do nothing but move along their specified routs. The purpose of having them in the microworld is to make it harder for the participant to identify contacts, e.g. when the participant uses a radar to survey an area, is the detected units neutral or hostile?

Enemy units behave in the same way as neutral units. They proceed according to their routes unless they detect a unit from the participants' force, at which point they will fire an appropriate weapon if the unit is within range. The enemy base decisions to fire weapons on the information in the enemy COP, which is created in exactly the same way as the participants COP. However, a few simple rules are used to identify the contacts in the COP that should be engaged.

Scenario

The scenario took place in the Baltic Sea and the time was set to November 28th 15:00 hours. Reliable intelligence stated that the enemy was planning to invade the island of Gotland as part of an operation primarily aimed at another country in the region. To prevent this, an infantry brigade had to be transported from the Swedish mainland to Gotland. The enemy could invade Gotland no earlier than December 1st, but did have some capacity to disturb the transport. On Gotland two ports, Visby and Klintehamn, could be used for debarkation. To make sure that the brigade could establish defensive positions, the brigade had to debark no later than November 30th at 18:00.

Task. The participant had to take command of a naval task force, and plan and execute the transport from the Swedish mainland to the island of Gotland. The naval task force consisted of the following units: Four anti submarine warfare (ASW) and anti surface warfare (ASuW) corvettes, four ASuW missile crafts, and four ASW/ASuW fast patrol boats. The infantry brigade was loaded onto three transport ships, of which two could depart at 19:00 tonight, and one at 22:00. The port of Visby allowed for two transport ships to unload at the same time, and the port of Klintehamn allowed for one. It took five hours to unload one transport ship. The participant also controlled one radar surveillance helicopter and two ASW helicopters.

Detailed intelligence stated that six enemy attack corvettes and one enemy radar surveillance helicopter were currently active in the area of operations. Of these, four attack corvettes were in the Gulf of Finland and two attack corvettes in the Bay of Riga. These corvettes would, however, not move west of longitude east 22° earlier than 23:00 the first evening according to reliable sources. It was also suspected that an enemy submarine was patrolling the waters north west of Gotland. The rules of engagements stated that the enemy could be engaged without warning.

Own plan. The former task force commander had prepared a plan for the mission. This plan exploited the intelligence report saying that the enemy could not move west of longitude east 22° no earlier than 23:00 the first evening. This gave an opportunity to "rush" with two transport ships and make them reach Gotland before the enemy could move within range to engage them. The last transport ship would then move south at slow speed and take position in the mainland archipelago west of Gotland. When own forces had established control in the waters west of Gotland the last transport ship would be escorted from its position to either Visby or Klintehamn.

Enemy plan. The idea in the enemy plan was to initially conduct raids on Swedish forces with the objective to disturb any reinforcement of Gotland, while later try to establish control in the waters east of Gotland as preparation for the landing forces. The plan for raiding was to be very active north of Gotland using the four attack corvettes from the Gulf of Finland, while moving silently with the corvettes from Bay of Riga to a position south of Gotland. From this position they would engage any Swedish forces moving west of Gotland. Eight hours before landing forces on Gotland, all remaining attack corvettes plus two additional attack corvetts would be used to establish control west of Gotland.

Injects. Four injects were prepared. The first inject, given to the participant November 28th at 22:15 scenario time[3], stated that enemy forces had crossed longitude east 22°, one hour earlier than anticipated. If the participant had used the original plan this meant that the enemy now threatened at least one transport ship. The second inject, given to the participant at 23:43 November 28th, revealed that mines had exploded just outside Visby harbour. This meant that the participants were forced to unload all transport ships in Klintehamn, which would lead to an extended debarkation process. The third inject, given to the participant November 29th at 02:18, stated that the enemy had begun loading troops in several ports in Bay of Finland. This indicates that the enemy is ready to invade Gotland earlier that anticipated. As a result the brigade must be unloaded no later than 22:00 November 29th. The last inject was given to the participant 08:14 November 29th and stated the enemy attack corvettes are moving in the south of the Baltic Sea and might threaten the participant's south flank.

Procedure

The participants were informed about the purpose of the experiment and were briefed about how it would be conducted. To train the participants in 'thinking aloud' they were instructed to solve the 'Necklace problem'[4] while verbalizing their thought processes. They were given 15 minutes or until the problem was solved. The next step was to train the participants in using the microworld and a simple search-and-destroy task was used as a training scenario. This training session continued until the participant had solved the task.

The participant were told that he or she, on short notice, had to replace the former task force commander who had disappeared under mysterious circumstnaces. The participant were given four different documents: A *scenario description* containing the task to be solved, a roster of own forces, and intelligence about the enemy; a *combat-estimate* made by the former task force commander; a geographical representation of the plan; and a spread-sheet illustrating the temporal aspects of the plan. The participants were also provided with a chart of the operations area together with rulers, compasses, pencils and paper and were given 45 minutes to read the documents. Clarification was given if it was deemed not to interfere with the experiment. The participants were also instructed to make any changes in the plan that he or she felt necessary.

When the participant was ready the initial changes to the plan was entered into the microworld and execution began. During the experiment the participant told the experimenter what he or she wanted to do in the microworld, and the experimenter entered the necessary instructions into the system. The main screen of the microworld was displayed on a 52" HD TV and the participants sat by the table or stood in front of the screen as they saw fit. The experiment was ended when the task was solved or when the participant aborted the mission. The duration of the experiment was about six hours, two hours of training and preparation, and four hours of execution.

Measurements

The variables measured was (1) the participants *decision making process*, (2) *task performance*, (3) *quality of sce-*

[3] All time references here refer to scenario time in the microworld.

[4] One version of the problem is: "You are given four separate pieces of chain that are each three links in length. It costs 2$ open a link and 3$ to close a link. All links are closed at the beginning of the problem. Your goal is to join all 12 links of chain into a single circle at a cost of no more than 15$"

nario, (4) *quality of execution*, (5) the participants *general command experience*, and (6) the participants *command experience in the current scenario*.

Decision making process was operationalized as the cognitive activities used when interacting with the microworld, and was measured using think-aloud protocol together with screen shots from the microworld. Think-aloud protocols have been used extensively within problem solving research (see for example Newell & Simon, 1972) and it is suggested that it does not interfere with the problem-solving process (Ericsson & Simon, 1980). To determine individual differences, two different measures of the decision making process was developed: (1) a quantitative measure that described the distribution of the participant's decision making activities, i.e. the percentage the participant engaged in each activity, and (2) a qualitative measure that described how elaborate the participant's decision making process was. The qualitative measure was operationalized as the mean number of decision making activities covered by the participant in one coherent reasoning chain. The more statements that are included in reasoning chains, and the longer the chains, the higher this measure will be.

Task performance was operationalized as mission success and was measured as the weighted sum of number of transport ships safely escorted to the goal harbor, number of own losses, number of neutral ships sunk, and number of enemy units sunk. The measure reflected the following rule: Getting all transport ships to the goal harbor was ranked as most important, so getting three transport ships to harbor but losing all your own forces was a better result than getting two out of three to harbor without losing any own units. To distinguish between those who got the same number of transport ships to the goal harbor the number of own losses, number of neutral ships sunk, and number of enemy units sunk was used. Here an own unit and a neutral unit were ranked equal; losing four own units was equally bad as losing two own units and sinking two neutral units. Losing one own unit or sinking one neutral unit was ranked as bad as it was good to sink two enemy units. So losing one own unit could be 'compensated' by sinking two enemy units. This yielded the following formula for calculating mission success using variables, performance (p), number of enemy units sunk (u_{enemy}), number of own units sunk (u_{own}), and number of neutral units sunk ($u_{neutral}$) :

$$p = m \times \max\left(u_{enemy} - 2 \times (u_{own} + u_{neutral})\right) + \\ + u_{enemy} - 2 \times (u_{own} + u_{neutral}) \qquad (1)$$

Quality of scenario was operationalized as the participants subjective opinion of the scenario and was based on the sum of two quality criteria measured on a six level Likert-scale, (a) 'Do you think there was enough information in the scenario to allow you to play your role?' (1=not at all, 6=completely), and (b) 'How did you perceive the level of uncertainty in the scenario?' (1=very low, 6=very high). The purpose of measuring this variable was to determine if the scenario allowed the participant to take the task seriously.

To establish the *quality of scenario* it was assumed that sufficient amount of scenario information, and a medium level of uncertainty should be ranked the highest, and that too little scenario information plus either too low uncertainty or too high uncertainty would be ranked the lowest. To reflect this assumption the following rule was used, where $mean_{scale}$ denotes the mean value of the uncertainty scale (in this case $mean_{scale} = 3.5$ since the scale had six levels, yielding a scale with levels 2-12):

$$\text{quality} = \text{enough information} + \\ + 2 \times mean_{scale} - |\text{uncertainty} - mean_{scale}| \qquad (2)$$

Quality of execution was operationalized as the participants subjective opinion of playing the commander and was based on the sum of two quality criteria measured on a six level Likert-scale, (a) 'Was there enough time to allow you to command in a good way?' (1=very little time, 6=very much time), and (b) 'Did the microworld reproduce the corresponding situation at sea sufficiently enough to allow you to immerse in the scenario and play your role?' (1=not at all, 6=completely). The resulting scale had levels 2-12. The purpose of measuring the first variable (a) was that high levels of time pressure and stress tends to reduce the cognitive capacity that a participant can devote to solve a complicated task (Zakay, 1993).

General command experience was operationalized as the time the participant had commanded on corresponding levels during his or her career, both during exercises and missions, measured in weeks.

Command experience in the current scenario was operationalized as *general command experience* but only measured the number of weeks in similar scenario as in the experiment. Both experienced variables was measured because it can be expected that experience is correlated with task performance.

Data collection and analysis

To capture the participants decision making process the participant was fitted with a head-mounted camera that recorded both what the participant was saying and what he or she was looking at. As a back-up system an overview camera was placed behind the experimenter and was setup to record all sound in the room together with the main screen of the microworld. To determine the events triggering the participants decision making process a screen-capture program took screen-shots of the microworld every other second producing a 'movie' of what was displayed of the screen. To establish task performance all missiles fired and ships sunk were logged by the microworld. The quality of the scenario, the quality of execution, the general command experience, and the command experience in current scenario were collected by a questionnaire.

The voice recordings of the participant and the experimenter were transcribed verbatim. The head-mounted camera, the overview camera, and the movie of the microworld were loaded into a reconstruction software called F-REX (see Andersson, Pileman, & Hallberg, 2008). F-REX allows the researcher to play several data sources simultaneously and was used to identify the microworld events and the scenario time related to the participants verbalized thought-process. The events and scenario time was entered in the transcription. Each transcription was then reduced by the author in three steps. The first step consisted of reducing the text to more compressed statements as illustrated by the following example:

Transcription:

> <15:21 contact with enemy radar surveillance helicopter>
>
> Participant: Radar surveillance helicopter, ok, it's probably over land somewhere, over the Baltic states in this scenario...oh that far up north...and we had no opportunity to use own fighter aircrafts?
>
> Experimenter: Sorry
>
> Participant: And we have no ships that can strike it. Well then you have to assume that this picture not then, for the opponent starts to clear up maybe that there are movements in our archipelago towards south.

Statement:

> <15:21 contact with enemy radar surveillance helicopter>
> Up in the north. Then you have to assume the the picture does not start to clear up, maybe that there are movements in our archipelago towards south.

The second step of the reduction consisted of arranging the statements according to scenario time on large sheets of paper. This gave an overview of each participants decision making process. The author then analyzed the statements and reduced sequences of statements to even more compact activities. In the final step of reduction the activities were grouped and related to each other to form a model of the participants decision making process.

To obtain the quantitative measure of the participant's decision making process the transcripts were reanalyzed. Statements in the transcripts that described a decision making activity was transferred to the corresponding section in a separate log. The log contained in total 22 categories. The number of statements belonging to each activity was then counted and the distribution was obtained by dividing the number of statements in each activity by the total number of statements for that participant.

The log was also used to obtain the qualitative measure of the decision making process. It was recorded how many activities that was covered in a coherent reasoning chain. A statement covering only one activity was treated as a chain with length one. Then the number of reasoning chains was determined for each participant. The mean length was determined by dividing the total number of statements for a participant with the number of chains for that participant.

To assess the validity of the categorization, 100 statements (of 1212) was randomly selected so that their distribution reflected the overall distribution of statements. A second rater was instructed of how the categories should be interpreted, and were also given three randomly selected example statements of each category. The second rater then categorized the 100 statements yielding a inter-rater measure of 74% (74 of 100 statements were categorized the same).

Results

The participants mean *command experience* was 140 weeks (min=71, max=212), and their mean *experience in current scenario* was 47 weeks (min=3, max=137). The mean *quality of scenario* was 11.7 (min=11, max=12), and mean *quality of execution* was 9.7 (min=7, max=11). The mean task performance was 5.8 (min=-12, max=16, sd=10.2).

When solving the task the participants engage in mainly three different processes: *Find Enemy Units*, *Counter Threat and Exploit Opportunities*, *Evaluate Effects of Actions and Events*.

Find enemy units

The participants spend most of the simulation time monitoring the COP with the purpose to determine which of the contacts tracked by the own force are enemy units (see figure 3). As the operation begins, own surveillance produce a set of contacts. Which areas that are surveyed and what sensors the units use are determined by the current course of action. The microworld immediately classifies each contact as either enemy, neutral or unknown. As neutral units do not pose a threat they do not trigger any action from the participants part. Enemy units, on the other hand, trigger action. But since they have been classified they do not pose a problem to the participant when it comes to identifying them, so from this process' point of view they are uninteresting.

The remaining problem facing the participant is to establish the identity of the unknown units and to accomplish this the participants use the set of unknown contacts, together with a model of the civil shipping in the area - the *neutrals model* - and a model of the enemy - the *threat model*.

The neutrals model. The neutrals model contains the expected behavior of civil shipping in the area, such as merchant speed ranges, sea-lanes in the area, major ports, etc. As said above, neutral units are left unattended by the participant apart from he or she using them to update the neutrals model, to establish what some participants called "a normal picture". As new neutrals are detected the model is updated - "a bit more civil shipping than expected", "quite high speeds for merchants" - until the participant is satisfied. The process of 'tuning' the neutrals model occurs during the early stages of the experiment.

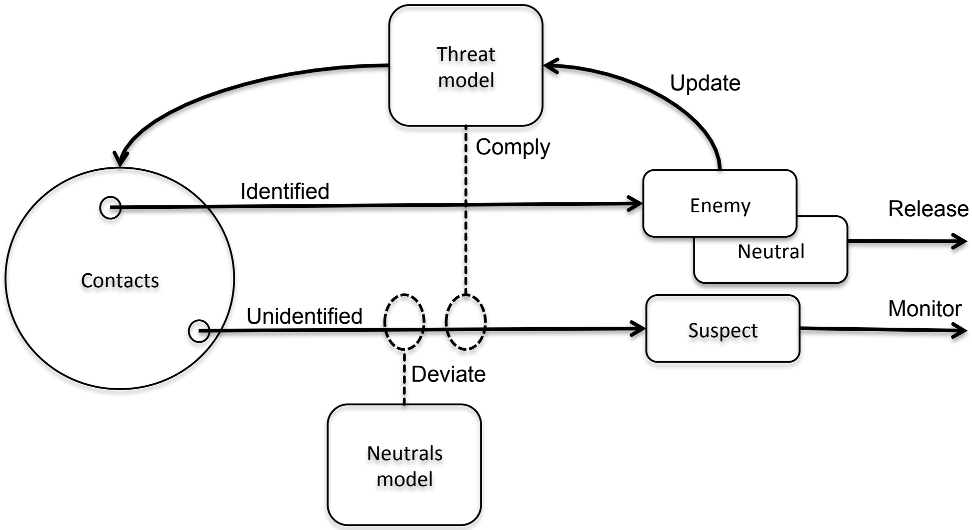

Figure 3. Finding the enemy

The threat model. The threat model is based on the enemy intelligence and creates expectations of which enemy units are in the area, where they are, and when they can be expected to operate in certain areas. The model also includes the participants' knowledge about the capabilities of the enemy units, together with their tactics. There is a reciprocal relation between the detected enemy units and the threat model. When detecting an enemy unit its location is compared to the predicted location by the threat model - is it plausible to find the enemy at that location given what I know about the enemy? If there is no conflict between the expected and detected situation, the threat model is updated given the new information about the enemy. However, if there is a conflict between the threat model and the perceived data, the participants starts problem solving in order to fit the conflicting data into the threat model in a coherent way.

There are some differences between the poor performers and the good performers when it comes to maintaining the threat model. Upon receiving new information about the enemy, either by own sensors or by intelligence reports, they all try to determine what the enemy is up to given this new information. For example, when the enemy leaves the Bay of Riga and starts move south, all participants detect this. However, the bad performers only do a superficial analysis of how this *influences own operations*, and thus have problems in drawing conclusions of what to do. This hinders them from taking action in due time and they run the risk of being overmanned later on. Having investigated how the participants create the neutrals and the threat models, we can now go back and look at how they use these models to identify the unknown units.

Unknown units are of special interest to the participant, since every unknown unit is potentially an enemy. To determine which of the unknown units that may be enemies, the participant uses both the neutrals model and the threat model. Unknown contacts with a behavior that deviates from the behavior predicted by the neutrals model are treated as suspects and is monitored more closely. Unknown contacts can also be 'upgraded' to enemies if they display a behavior that is very different from what is expected by civil shipping, and at the same time show a behavior that is very consistent with the behavior predicted by the threat model, e.g. using speeds than no civil shipping would use. Upgrading may also occur if a suspect keeps the deviant behavior for longer periods of time, or if it moves to a position where it can threaten the participants own operations.

When it comes to building and maintaining the neutrals model there is no evident differences between the good and bad performers. However, some difference is observable when it comes to identifying and upgrading suspects, especially regarding the best performer. He or she is very vigilant when it comes to identify and track suspects. Small deviations from the 'normal picture' is observed and analyzed, which leads to early identification and good tracking of enemies. This is in contrast to the poor performers who are less vigilant and needs larger deviations to identify a suspect.

Counter threats and exploit opportunities

The purpose of *counter threats and exploit opportunities* is to determine if the participant needs to take any action given the new situation. There are two events that trigger this process: either a change in the enemy contacts tracked by the own force, or a change in the enemy force predicted by the threat model (see figure 4). An example of triggering by tracked enemy contacts could be that the force detects an enemy radar surveillance helicopter. An example of triggering by the threat model is that enough time have elapsed since the last contact with an enemy unit, to allow that unit to move to a position where it could strike at an own unit.

When detecting an enemy unit the participants immediately tries to determine if it poses any immediate threat to own operations. Looking at the unit's capability to detect own units and its capability to strike at own units does this. The conclusion is used to determine the consequences for

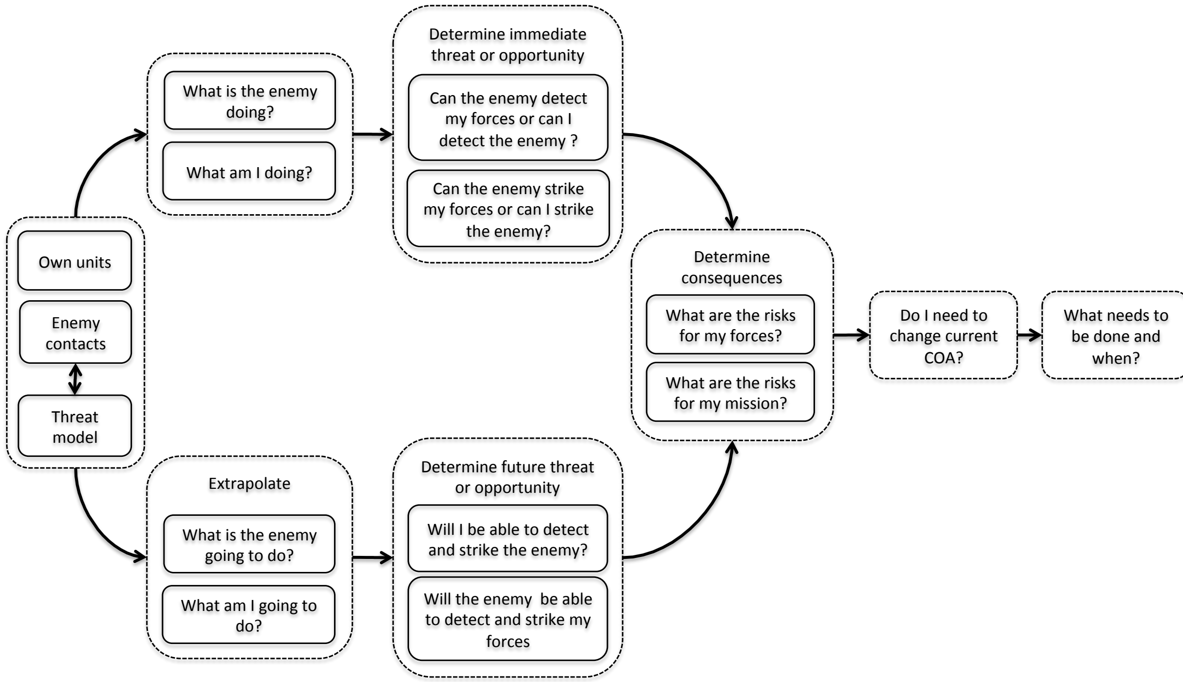

Figure 4. Countering threats and exploiting opportunities

the own force and the mission. Once the consequences have been established, the participant has to decide whether or not he or she has to make any changes to the current operations. Following examples illustrates this process:

> Ok, then we have identified two opponents out there...my southern missile ships will engage them when they come within range. I assume however that my ships have positioned themselves so that the enemy cannot engage them. [They have not] Ok, then move them into a protected position.

The threat model is also used to determine if any future threat or opportunity is about to develop. The participant compare the threat model's predicted behavior of the enemy to the own course of action and determines if he or she will be able to track and strike the enemy in the predicted situation, and if the participants have enough forces to handle the situation.

> How should I handle the attack corvettes when they finally approach from the south, because they will come from the south, it is their direction of movement. And at least keep, I will be divided here in, my forces, if we think about, if a submarine will be detected [...] you want to keep the ASW-helicopters and costal corvettes together but it will be impossible I will have to, will be forced to split [them] if I should be able to keep a guard to the south.

Evaluate effects of actions and events

When the participants have taken action or when they receive information about changed mission parameters or receive new intelligence, he or she tries to determine if this new situation forces the participant to change the current course of action. Three, such events have been observed: *Evaluate effects of battle*, *Changed mission parameters*, and *New intelligence*.

One evident process is the evaluation process following a battle (see figure 5, top diagram). In this process the participant seeks to determine how many enemy ships that have been sunk in the battle, and how many own ships were lost. Given the result they update the threat model and then try to determine if the surviving own forces can handle the remaining threat. If the participant decides on changing, new actions are devised and initiated.

What differentiates the good performers from the bad in this process is that they are more thorough when analyzing new situation. Good performers do not only look to the battle - I killed two of them, they killed one of mine - they also reflect upon the overall force balance - they have four in the north and two in the south, I have two in the north and two in the south. Further, the good performers tend to count their strength in missiles, not in ships, in contrast to bad performers e.g "I have 48 missiles in the north, he has 32".

Change of mission parameters is initiated when the participant receives information that has effect on the mission (see figure 5, middle diagram). Injects two and three - the increased mine threat outside Visby which led to that only one harbor could be used for debarkation, and the order stating that the brigade must have taken defensive positions earlier

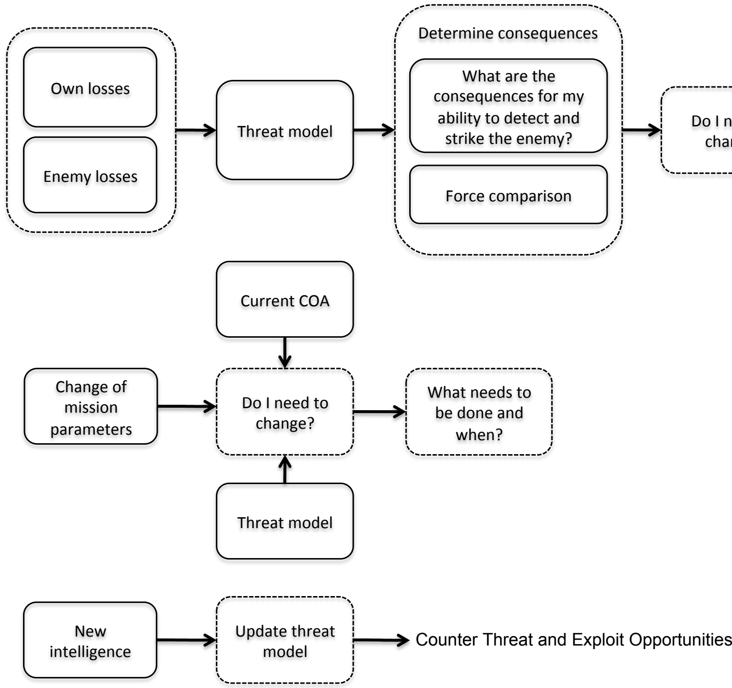

Figure 5. Evaluate effects of actions and events

than planned for - are the main examples from this experiment. These changes to the mission are so large that the participants are forced to engage in thorough analysis. In this analysis the own course of action is analyzed with the current threat model to determine whether or not changes must be made. However, no evident differences between good and bad performers were observed.

New intelligence is initiated when the participant receives new information about the enemy (see figure 5, bottom diagram). In this experiment injects one and four were of this type. What the participant does when receiving this information is to update the threat model and then enter *counter threat and exploit opportunities* to determine if changes to own operations must be made.

The relation between the decision making process and task performance

The decision making process was analyzed using both the quantitative and the qualitative measure. When using the quantitative measure, all activities found in *counter threats and exploit opportunities* was used, together with *find enemy units*,which was treated as one category. Finally four activities regarding orders were added: task orders, sensor orders, movement orders, and firing orders. This gave a total of 22 activities. The correlation (Pearson's r) between *task performance* and the the proportion of statements in each category was calculated, yielding no significant results.[5] The qualitative measure of the decision making process was used to calculate if there were any correlation between *task performance* the mean length of the reasoning chains. This yielded a significant correlation (Pearson's r) of $r = 0.87$, $p = .025$.

Discussion

The purpose of this study was to investigate differences in the decision making processes of experienced navy officers while they executed a maritime escort mission, and try to relate those differences to differences in task performance. To achieve this, a model of the decision makers were constructed from verbal protocols and a performance measure was constructed using outcomes in the microworld. To examine the decision making process two measures, one quantitative and one qualitative, were constructed, and the correlation between task performance and these two measures were calculated.

The results suggest that there is no correlation between how often the participants perform a certain decision making activity, and how well they perform in the microworld. On the other hand, the results suggest a strong correlation between how well the participants perform in the microworld and how many different decision making activities they visit during one coherent reasoning chain. This may be interpreted as no decision making activity is more important than another, but rather that they who perform well in the microworld considers more aspects of the task at the same time when trying to solve it. This is also in line with the observations of the participants.

The participants in this study was quite experienced, both form the navy in general, but also in the scenario used in

[5] There were three correlations that were significant on the .05-level: 'Can I detect the Enemy?', $r = -0.89, p = .015$, 'What are the risks for my mission?, $r = 0.83, p = .042$, and 'Unit task orders', $r = 0.82, p = .044$. But none when using an adjusted α-level

the experiment. The results suggest that the participants believed that the scenario was of high quality, and also that the microworld and the pace of the game was good. This gives some reason to believe that they engaged in the task, and that they solved the task in similar fashion as they would have done in a real setting. The study, however, only include six participants and to get more reliable results the study must be expanded. Further, the inter-rater reliability was moderate which indicate that the decision making activities used needs to be better defined in order to get a more definite result.

Nevertheless, the most interesting finding in this study is that it suggests that it is more important to consider many aspects of a problem at the same time, rather than that certain decision making activities are more important that others.

References

Andersson, D., Pileman, S., & Hallberg, N. (2008, May). Evaluation of crisis management operations using reconstruction and exploration. In *Proceedings of the 5th international ISCRAM conference.*

Brehmer, B. (2000). Dynamic decision making in command and control. In C. McCann & R. Pigeau (Eds.), (chap. 16). US, NY: Kluwer Academic/Plenum Publishers.

Brehmer, B. (2005). The dynamic OODA loop: Amalgating Boyd's OODA loop and the cybernetic approach to command and control. In *Proceedings of the 10th international command and control research and technology symposium.* US Department of Defence.

Brehmer, B. (2006). One loop to rule them all. In *Proceedings of 11th international command and control research and technology symposium.* US Department of Defence.

Brehmer, B. (2007). Understanding the functions of C2 is the key to success. *The international C2 journal*, *1*(1), 211-232.

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

Brehmer, B., & Nählinder, S. (2007). Achieving what cannot be done: Coping with the time constants in a dynamic decision task by doing something else. *Scandinavian Journal of Psychology*, *48*, 359-365.

Bryant, D. J. (2003). *Critique, explore, compare, and adapt (CECA): A new model for command decision making.* (Technical report No. TR 2003-105). Defence R&D Canada.

Dörner, D. (1980). On the difficulties people have when dealing with complexity. *Simulation and Games*, *11*(1), 87-106.

Dörner, D., & Wearing, A. (1995). Complex problem solving: Towards a (computersimulated) theory. In P. Frensch & J. Funke (Eds.), (p. 65-99). Hillsdale: NJ: Lawrence Erlbaum.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological review*, *87*(3), 215-251.

Frensch, P., & Funke, J. (1995). *Complex problem solving: the european perspective*. Lawrence Erlbaum Associates.

Levis, A., & Athans, M. (1988). Science of command and control: Coping with uncertainty. In S. Johnson & A. Levis (Eds.), (p. 4-9). Fairfax, VA: AFCEA International Press.

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

Schmitt, J., & Klein, G. (1999). A recognitional planning model. In *Proceedings of the 4th ICCRTS.* US Department of Defence.

Swedish Armed Forces. (1987). *Taktikreglemente för flottan (tactical regulations for the navy).* Stockholm, Sweden: Chefen för Marinen (Head of Navy).

Thunholm, P. (2003). *Military decision making and planning: towards a new prescriptive model*. Unpublished doctoral dissertation, Stockholm University, Stockholm, Sweden.

US Navy. (1995). *U.S. navy doctrine document 6: Command and control*. Washington, D.C.: Department of the Navy.

US Navy. (2007). *Navy planning NWP 5-01.* Newport, RI: Department of the Navy.

Waldenström, C. (2008). What is difficult in naval sensemaking? In *Proceedings of the 13th international command and control research and technology symposium.* US Department of Defence.

Waldenström, C., Ekenberg, L., & Danielsson, M. (2009, July 14-18). Threat and control in military decision making. In *6th international symposium on imprecise probabililty: Theories and applications (ISIPTA 09).*

Zakay, D. (1993). The impact of time perception processes on decision making under time pressure. In O. Svensson & J. Maule (Eds.), (Time Pressure and Stress in Human Judgement and Decision Making ed., p. 170-180). US, New York: Plenum Press.