Managing CSIRT Capacity as a Renewable Resource Management Challenge: An Experimental Study

Agata Sawicka
Faculty of Science and Engineering
Agder University College
NO-4876 Grimstad, Norway
Tel: + 47 37 25 33 58 / Fax: + 47 37 25 30 01
Email: agata.sawicka@hia.no

Jose J. Gonzalez
Faculty of Science and Engineering
Agder University College
NO-4876 Grimstad, Norway
Tel: + 47 37 25 32 40 / Fax: + 47 37 25 30 01
Email: jose.j.gonzalez@hia.no

Ying Qian
Faculty of Science and Engineering
Agder University College
NO-4876 Grimstad, Norway
Tel: + 47 37 25 34 25 / Fax: + 47 37 25 30 01
Email: ying.qian@hia.no

Abstract: CSIRTs are security incident handling organizations serving a parent organization or a "constituency" of independent organizations. CSIRTs struggle coping with the increasing number and sophistication of incidents; staff is overloaded with work; managers 'over-utilize' their teams. The CSIRT 'mismanagement' problem can be framed as a case of natural resource management. Studies by Moxnes suggest that misperception of dynamics may contribute to natural resources mismanagement. We replicate experiments by Moxnes (2004), reframing the one-stock reindeer rangeland management task as a challenge in sustainable CSIRT management. Our results suggest: 1) The misperception of dynamics persists when the problem context changes; 2) people employ a simplistic anchoring-and-adjustment decision rule to deal with the problem; 3) our data do not support the version of the rule proposed by Moxnes. We hypothesize that the observed misperception might at least in part depend on the way in which the task was presented.

The work presented in this paper was carried out as a part of the postdoctoral research project of Agata Sawicka founded by Grant 160789/V30 from the Research Council of Norway. Contribution by Jose J. Gonzalez (supported within the same grant) concerned primarily participation in the experimental task conceptualization and paper review. Ying Qian (supported by Grant 164384/V30 -AMBASEC from the Research Council of Norway) helped in running the experiment, data analysis and paper review.
The first author is indebted to Professor Erling Moxnes for granting the permission to adapt his experimental simulator to the needs of this study, and to dr Robert J. Bois for granting the permission to use parts of the self-assessment questionnaire developed by him in this study. The first author is especially grateful to Johannes Wiik for sharing his insights regarding management challenges faced by CSIRTs and providing relevant references, and to Deborah Campbell for her valuable feedback regarding the experimental task and the task instructions.

Introduction

Rapid development of computer technology over the past decade has changed the way modern organizations function. Their operations are now supported by ever more complex and attractive applications. The same IT revolution, however, has facilitated development of increasingly effective tools for exploitation of software vulnerabilities (Lipson 2000, Schneier 2000): In the 1980's malicious agents needed both will and knowledge. Now, with the automated intrusion tools available, in many cases, will and minimal e-literacy suffice. To combat this increasing and serious threat, the Carnegie Mellon University CERT® Coordination Center encourages organizations to establish CSIRTs – Computer Security Incident Response Teams (CERT/CC 1998, West-Brown, Stikvoort et al. 2003).

CSIRTs are specialized service units that assist their parent/constituency organizations in handling computer incidents and staying at guard. CSIRTs need to master the changing security threat landscape and deploy automated tools to cope with an ever increasing volume of computer intrusions. Hence, it is essential that some of the CSIRT activities are directed towards improvement of their know-how and capability. Still, funding of the response teams does not depend directly on their capacity, but is a function of the services offered.

Achieving an appropriate balance between the capacity development and service activities seems to be one of the main challenges faced by the CSIRT managers: Focusing too much on the know-how and capability development will impede a CSIRT's ability to provide services, threatening its funding and, hence, survival. Increasing service level is likely to yield greater funding; however, if done without a sufficient capacity backing, it would impede a CSIRT's ability to maintain and develop its know-how and capability, leading to collapse in a long-run.

It seems that a parallel could be drawn between the challenges faced by the CSIRT managers and managers of natural renewable resources: In both cases, striking a balance between exploitation and protection of the utilized resource is necessary to achieve a sustainable enterprise. In case of CSIRTs, the renewable resource is the CISRT capacity to provide services.

Among the best known system dynamics studies on management of renewable resources are those conducted by Moxnes (see e.g. Moxnes 2000). In this paper we report on an experimental study in which we replicate the Moxnes' experiment with one-stock reindeer rangeland management task (Moxnes 2004, treatment T1), reframing the task as a CSIRT capacity management challenge.

This paper contributes to the system dynamics field with an experimental validation of Moxnes' findings regarding misperceptions of dynamics. By embedding the task in a different problem domain we test the general validity of the original results. Replication is an essential part of any research process (Sidman 1960, Cooper, Heron et al. 1987). Still, while the body of experimental studies within the system dynamics field increases, an effort to replicate earlier results (with notable exceptions by Howie, Sy et al. 2000, Bois 2002, Jensen 2005) seems to be rather limited. The results of this study will also provide input for our future investigations concerning identification of more effective ways of communicating dynamic aspects of complex problems.[1]

The paper contributes to the computer security field with a simple case study that might serve as a classroom example illustrating challenges involved in management of computer security incident handling. The computer security curricula – traditionally dominated by strictly technical aspects – are in need of cases that could be used to teach about the human and dynamic aspects. Two recent publications by Melara, Sarriegi et al. (2003) and Martinez-Moyano et al. (2005) indicate that system dynamics provides an attractive platform for devising such generic cases. The CSIRT capacity challenge developed for the purpose of this study provides yet another prototypical system dynamics-based case that could be included in the enhanced computer security curricula.

The paper proceeds as follows: First, the concept of CSIRT capacity and how it may be seen as a renewable resource is briefly explained. Next, we discuss the experimental task. This is followed by presentation and discussion of our experimental results. The closing section summarizes our findings and outlines how the results will fuel our future research.

CSIRT capacity as a renewable resource

The CSIRT capacity may be thought of as ability to provide computer incident related services: the greater the capacity, the greater the challenges that can be handled by the CISRT. It may be expressed as a total number of average person-hours that a CSIRT is capable of delivering during one hour. The contribution of the CSIRT members will vary depending on their proficiency to carry out the tasks (experience means faster performance).

The CSIRT capacity expressed in terms of average person-hours is illustrated in Figure 1. Both CSIRTs depicted in Figure 1 consist of 5 staff members. Despite of the same level of staffing, each team has a different overall capacity. This is due to the difference in the levels of expertise of the staff on each team.

[1] This research will be conducted as a part of the project titled: *Disseminating Insights from Complex Models to a Broader Audience: Case of system dynamics models*, which is a postdoctoral fellowship for Agata Sawicka funded by the Research Council of Norway (grant 160789/V30).

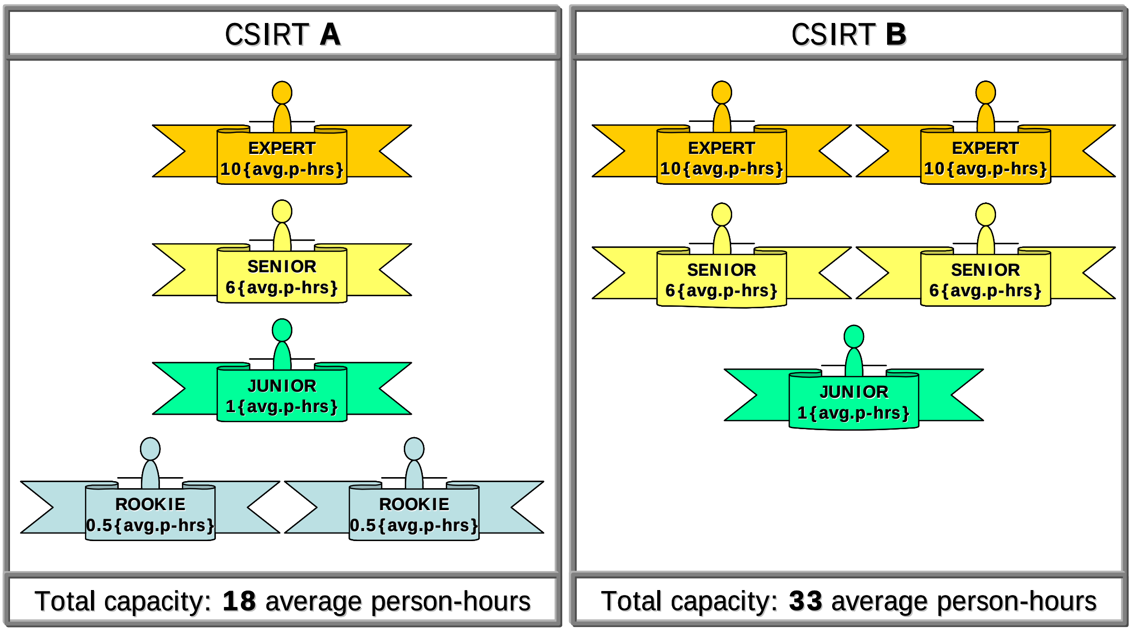


Figure 1 Conceptual illustration of the CSIRT capacity estimation.

The CSIRT capacity may be increased through hiring of new staff as well as through a range of developmental activities, such as training, exploration of new technologies, participation in various research projects, development of software tools to automate services, etc. Although these activities increase the overall CSIRT capacity and hence its efficiency, they reduce the CSIRT's ability to engage in service-related activities. The inherent conflict of capacity development and throughput results in a trade-off situation quite typical for security or quality management. In such tradeoff situations, managers frequently overlook non-productive needs, focusing their attention primarily on the tangible and pressing throughput goals (see e.g., Reason 1997, Sterman 1997). Given the continuous and rapid changes in the computer security threat landscape, such bias in the context of CSIRTs seems especially dangerous: Firstly, over time CSIRT capacity becomes inevitably obsolete. Prolonged periods without sufficient development will lead to depreciation in the CSIRT capacity. Secondly, it has been noted that continuous performance of service-related activities only is likely to lead to staff burnout, impeding seriously staff's ability to handle effectively incoming inquiries (Wack 1991, Smith 1994). To counteract the service-related capacity utilization, involvement in less-stressful and more creative developmental activities is seen as essential; specific recommendations vary, indicating that the CSIRT staff spent anywhere from 20% up to as much as 45% of their time engaged in various developmental activities (see West-Brown, Stikvoort et al. 2003, van Wyk and Forno 2001, respectively). These activities not only help to reenergize the staff, regenerating the service-related capacity utilization, but also prevent the staff's skills and knowledge from becoming obsolete. They may also contribute directly to development of new capacity (e.g., by the staff acquiring new skills and knowledge, or by provision of new automated tools yielding more effective incident handling, etc.).

Accordingly, CSIRTs ought to operate with balanced level of services and sustain developmental activities. Still, the results of a recent survey of CSIRTs (see Killcrece, Kossakowski et al. 2003) shows that many teams struggle with achieving the highest sustainable service level. Reports of excessive workloads, frequent staff burnout, and high turnout rate are common. Many CSIRTs seem to be stretched to their limits and are unable to devote sufficient resources towards capacity maintenance and development. Faced with ever increasing number of incidents and inquiries, teams devote less and less time towards the capacity enhancement activities. Consequently, the CSIRT's ability to provide services efficiently gradually declines, and excessive workload increases yet again.

The situation seems to resemble management challenges already discussed in the system dynamics literature in other contexts: First, the situation might be seen as an instance of the 'working hard' versus 'working smart' dilemma. Management repeatedly falls for the 'working hard' strategy, attending reactively to the pressing throughput demands rather than proactively investing in the capacity enhancement. Basic dynamics of the 'working hard' and 'working smart' strategies are discussed by Repenning and Sterman (2001). In one of the parallel presentations intended for this conference, other members of our research group[2] discuss how these dynamics operate in the context of CISRT management (Wiik and Gonzalez 2005). On the other hand, the situation, in which excessive workloads lead to CSIRT capacity overutilization, seems to resemble the case of overexploitation, commonly observed in the context of natural renewable resource management (Kneese and Sweeney 20021985). It is this perspective that is taken on in this paper.

Overexploitation describes a situation in which a renewable resource, being below its optimal sustainable level, is exploited at rate that exceeds its self-regeneration. Continuing this type of exploitation inevitably leads to destruction of the resource. Overexploitation traditionally has been seen as a result of an unrestricted access to given renewable resource – the so-called 'tragedy of the commons' (Gordon 1954, Hardin 1968, Hardin and Baden 1977). However, experimental studies by Moxnes (1998, 1998, 2000, 2004) suggest that overexploitation might occur even when people are granted exclusive property rights. For example, when managing a simulated reindeer rangeland or fishery, people often misperceive the dynamics of the system. In particular, they fail to understand how an inverse U-shape of the renewable resource net growth rate affects the system.

To the best of our knowledge there are no formal estimates of the net growth rate of CSIRT capacity. However, it seems plausible to assume that the growth rate could be described as an inverse U-shape function of the CSIRT capacity: When the CSIRT capacity is low, the net growth rate would be only very small due to insufficient resources to support the developmental activities. When the capacity is high and approaches its maximum, the development of new capacity would again be very small (majority of effort would be directed towards update/upgrade of the existing capacity). Somewhere in between these extremes, the net capacity growth rate would reach its maximum. Note that the inverse U-shape net growth is also consistent with the law of diminishing returns, frequently referred to when explaining why the resource increments are not directly proportional to made investments.

Given that the CSIRT capacity net growth could be described by an inverse U-shape function of the CSIRT capacity and that CSIRT managers have exclusive 'property rights' over the capacity of their teams, one could suspect that the problems they experience might at least in part be due to misperceptions of the capacity development dynamics. The following section discusses an experimental environment devised to explore whether overexploitation due to misperceptions of dynamics observed by Moxnes in the context of

---

[2] Security & Quality in Organizations, http://ikt.hia.no/sqo

management of reindeer rangelands (Moxnes 2004) and fisheries (Moxnes 1998) would also occur in the context of CSIRT capacity management.

Experimental task analysis

In this section, we first describe how the task of reindeer rangeland management developed by Moxnes (2004) has been reframed as a CSIRT management challenge. Next, we discuss the expected results.

Dynamics of the CSIRT management

The task used in our experiment is based on the one-stock version of the reindeer-rangeland management task developed by Moxnes (2004). In the original task, the subjects are asked to restore the highest sustainable reindeer herd size based on information about the developmental dynamics of lichen – the plant essential to the reindeer survival during the winter season. In our case, we ask the subjects to arrive at the highest sustainable service level for the CSIRT they manage, given information about the dynamics of the CSIRT capacity development.

The system underlying the challenge is of the Lotka-Voltera type. In the original task by Moxnes (2004) the reindeer prey on lichen; in our version of the task the CSIRT services 'prey' on the CSIRT capacity. The stock-and-flow structure of the task is presented in Figure 2. The subjects have full control over the predator population dynamics – the number of CSIRT services is defined by the subjects and there are no variations in the service level in between the decision periods. On the other hand, the dynamics of the prey population (i.e., the CSIRT capacity) are fully controlled by the simulation. The nonlinear CSIRT capacity net growth rate, identical to this used in Moxnes (2004), is depicted in Figure 3.

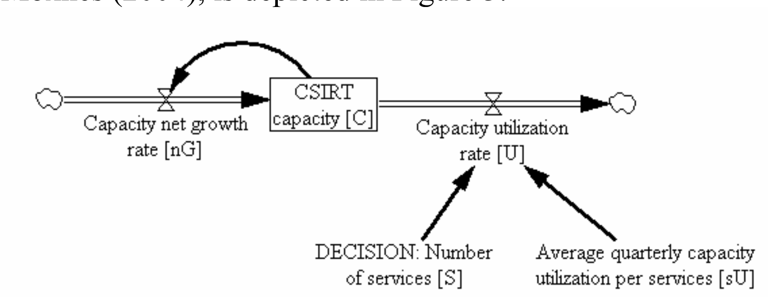


Figure 2 The stock-and-flow structure of the CSIRT capacity management task.[3]

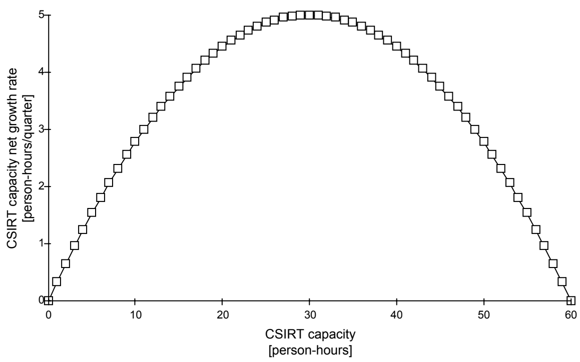


Figure 3 The nonlinear, CSIRT capacity-dependent, CSIRT capacity net growth rate.

[3] Fully documented Vensim model is provided in the supplementary materials.

To solve the task, one needs to realize that the maximum sustainable number of services is achieved when the service-related capacity utilization rate (U) equals the maximum CSIRT capacity net growth rate (nG_{max}). Using the net growth curve presented in Figure 3 we can easily see that nG_{max} equals 5 [person-hours/quarter] in our case. Knowing the quarterly capacity utilization per service (sU=0.04 [(person-hours /service)/quarter]),[4] the maximum sustainable service level ($S_{max\ sustain}$) may be calculated from the following:

$S = S_{max\ sustain} \leftrightarrow U = S*sU = nG_{max}$

$S_{max\ sustain} = nG_{max}/sU = 5$[person-hours/quarter]/0.04[(person-hours/service)/quarter] = 125[services]

The optimal decision algorithm for achieving the sustainable service level service level is presented in Table 1: Initially, the CSIRT capacity is overutilized and is Δ = 5.6 [person-hours] short of its optimal level (C_{opt} = 30[person-hours]). The quickest way to achieve the optimum is to cancel all the services for the first decision period. The CSIRT capacity increases to 29.2 [person-hours] in the second decision period. This is still short of the optimum level: To achieve the required level, the CSIRT capacity needs to be further increased by 0.8 [person-hours]. Given no services, the capacity would increase by 5[person-hours]. This is too much – 105 services should be provided to prevent capacity overdevelopment. In the third decision period the CSIRT capacity reaches its optimum level of 30 [person-hours] and the CSIRT services can be set to their maximum sustainable number $S_{max\ sustain}$ = 125 [services].

Table 1 Steps necessary to return to the highest sustainable level of services.

Decision period:	**I**	**II**	**III**
C : Current capacity level	24.4	29.2	30
Δ : deviation of the current capacity from the optimum capacity level (C_{opt} =30[person-hours]) **Δ = | C – C_{opt} |**	5.6	0.8	0
Desired service utilization rate should equal nG_{max}. Current desired service utilization rate U: **IF** Δ> nG_{max} **THEN U=0[person-hours/quarter]** **ELSE U=**nG_{max} **– Δ**	0	4.2	5
Current desired number of services **S = U/sU**	0	105	125

As in the study by Moxnes (2004), our subjects get only textual description of the CSIRT capacity growth dynamics.[5] Given this, they need not only to realize that the maximum sustainable CSIRT service level is achieved when the service-related capacity utilization equals the maximum CSIRT capacity growth rate, but they also must estimate the maximum growth rate. The only information given about the capacity net growth function is that it is an inverse U-shape function of CSIRT capacity. To identify its maximum, one should inspect the historical time series. Table 2 outlines the required calculations. As we can see to estimate the net growth, one needs to calculate at least some of the capacity increments (ΔC) and the service related capacity utilization (U). Given the capacity utilization in the given period and how the capacity level increment,

[4] The subjects are provided this information in the instructions, see the *Managing CSIRTs* handbook (p. 4) included in the supplementary materials.

[5] The description as well as the entire task instructions are analogous to the task instructions used in the Moxnes' study; compare *Appendix 1* in Moxnes 2004 and the *Instructions* section in the *Managing CSIRTs* handbook (see pp. 4-5 in *ManagingCSIRTs.pdf* included in the supplementary materials).

one can estimate the capacity net growth for that period, nG=ΔC+U. Keeping in mind that the capacity net growth function has an inverse U-shape, one should be looking for the maximum net growth rate when inspecting the historical times series – this might be identified as equal to 5[person-hours/quarter] for the 14th quarter.

Table 2 Estimating the capacity net growth curve.

Decision period [t]	CSIRT capacity [C]	No. of services [S]	Service utilization rate [U = S * sU]	Change to occur in CSIRT capacity [$\Delta C = C_{t+1} - C_t$]	CSIRT capacity net growth rate [nG = U + ΔC]
1	50	115	4.6	-1.82	2.78
2	48.17778	120	4.8	-1.64	3.16
3	46.54205	125	5	-1.52	3.48
4	45.02183	130	5.2	-1.45	3.75
5	43.56819	135	5.4	-1.42	3.98
6	42.14543	140	5.6	-1.42	4.18
7	40.72592	145	5.8	-1.44	4.36
8	39.28678	150	6	-1.48	4.52
9	37.80765	155	6.2	-1.54	4.66
10	36.26899	160	6.4	-1.62	4.78
11	34.65065	165	6.6	-1.72	4.88
12	32.93049	170	6.8	-1.85	4.95
13	31.08278	175	7	-2.01	4.99
14	29.07627	180	7.2	-2.20	5.00
15	26.87153	185	7.4	-2.45	4.95
16	24.41715				

Calculations presented in Table 2 show that to find the optimal solution the subjects not only need to understand how the nonlinear capacity development influences the capacity level together with the service-related capacity utilization rate, but they also need to perform a rather extensive initial analysis of the historical time series. Most subjects performed poorly in the studies conducted by Moxnes (2004). Given that in our experiment we do not provide the subjects with any additional information or aids, we would expect much of the same performance.

Exploring subjects command of the system

As outlined in the previous section, our experimental task mirrors the one-stock reindeer task developed by Moxnes (2004): the minor numerical differences[6] do not introduce any qualitative difference in the underlying mathematical structure of the problem. The structural and numerical equivalence between the dynamical tasks does not necessarily guarantee the same performance. For example, the results reported by Moxnes and Saysel (2004) indicate that performance is likely to be better in a more familiar task context. Still, given that our subjects, as the subjects participating in the original study by Moxnes (2004), are not likely to have any particular knowledge of the task context, we expect to observe the same type of performance.

Moxnes found that during the first trial only a few subjects seem to realize what is needed to return the system to its sustainable equilibrium. He argues that the poor performance is due to the subjects' *"inability to formulate an appropriate model for the*

[6] There are three numerical differences between the task in Moxnes' study (2004) and our study: (1) the decision-making is extended from 15 to 16 periods, (2) the initial service level and (3) the quarterly capacity utilization rate per service are reduced tenfold.

decision problem" (2004, 12). Indeed, it is a common finding within dynamic decision making research that people have difficulties to form an appropriate mental representation of the task (see e.g., Dörner 1975; 1989 (1996), Sterman 1987, 1989, Brehmer and Allard 1991).

However, there is more to the problem as acquiring an accurate mental model of the task. In case of the original reindeer task to achieve the sustainable herd size, the subjects had to understand the lichen's nonlinear net growth and how it affected the lichen level together with the grazing rate. In previous studies, where a more complex version of the task was used, presenting the net growth curve improved the subjects' performance. Still, Moxnes notes that *"even researchers with considerable experience in formal analysis happened to misperceive the figure"* (2004, p. 157). This observation is in accord with findings reported by Jensen (2005). She found that even when presented with the net growth curve, only 3 out of 28 subjects articulated correctly how this information should be taken into account when setting the reindeer quota levels; still, one of these subjects *"consistently, through all three trials, cut quotas so slightly below the equilibrium level that the rebuilding of lichen was far too slow"* (Jensen 2005, p. 128). This would indicate that even when people seem to understand the underlying dynamics, they still might have problems with translating this understanding into effective action. This is consistent with the second tenet of the misperception of feedback hypothesis which says that human ability to infer correctly the dynamics of dynamic systems is poor (see Sterman 1994, Sterman 2000).

To gain a better insight into how the subjects perceived and tackled the problem, we enhanced the experimental protocol used by Moxnes (2004) and asked our subjects to log into provided workbooks any type of analysis or calculations they perform as they work through the three trials. This supplementary data should help us to distinguish more precisely between the subjects who saw the important dynamic aspects of the problem but were not able to act upon this understanding, and the subjects who failed to develop an accurate mental model.

We expect that most of our subjects would not be able to develop an accurate mental representation of the task. As in the original one-stock reindeer task by Moxnes, the subjects in our experiment receive only a textual description of the prey population's net growth curve along with the historical time series, presented as graphs and tables. As indicated in the previous section, the provided information is sufficient for estimating the optimal solution. However, the required analysis (see Table 1 and Table 2) is – in our opinion – more than what most subjects would be prepared to do.

Moxnes finds that only 2 out of 34 subjects seem to recognize from the start that the herd size ought to be reduced. However, their reductions are not sufficient, indicating that their mental models of the dynamics is still too simplistic. Moxnes hypothesizes that initially most subjects are likely to rely on a simple static model saying *"the more animals, the less lichen, and vice versa"* (2004, p. 151), adjusting the herd size only gradually according to a simple anchoring-and-adjustment model. The hypothesized anchoring-and-adjustment decision model (presented in more detail in the *Discussion* section of our paper) represents some sort of feedback-based model. However, it has little relevance to the true nature of the task: The model is formed largely through a trial-and-error approach to the task, rather than reflective analysis of the problem. The tendency to follow trial-and-error strategies was detected in other experimental studies deploying the reindeer management task (Moxnes 1998) as well as in the context of other dynamic challenges (Jensen and Brehmer 2003, Dörner 1975; 1989 (1996), Crossman and Cooke 1974).

An interesting question is whether the subjects relying on the strategies developed through trial-and-error approach are content with their control of the system. It is conceivable that the subjects followed strategies that seemed to work just to get through with the experimental task. At the end of the day, the subjects are supposed to 'manage' not 'understand' the system. To explore to what extent the subjects tried to understand the system's workings a short post-test questionnaire was administered. The subjects were asked to comment on: (1) their own performance, (2) their perceived understanding of the task, and (3) their perceived ability to control the system.

To test whether there is a correlation between the subjects' effort and their performance, we also asked the subjects to assess how much effort they put into tackling the problem.[7] Earlier observations by Bois (2002) and Jensen and Brehmer (2003) suggest that performance in dynamic decision making tasks seems to be positively correlated with the individual effort.[8]

The final post in the questionnaire invited the subjects to assess decision aids they had at their disposal during the experiment and to suggest other aids that in their opinion could have helped them to achieve better results. The subjects were asked to consider the question again in their student project groups, following the experiment debriefing.[9] Assessment of the provided decision support as well as suggestions for improvements will provide an important input for our future studies, discussed in the closing section of this paper.

Experimental study

Subjects and experimental procedure

The experiment was conducted with 38 students taking a one semester system dynamics course at Agder University College during the spring 2005 term. At the time of the experiment the subjects had participated in 8 lectures in system dynamics, covering approximately Chapters 1-8 of Sterman's *Business Dynamics* (Sterman 2000). The experiment was conducted as part of the obligatory course assignment: All enrolled students had to participate in the study. However, their actual performance on the experimental task did not to influence their course grades in any way. The only grading related to the experimental study concerned the follow-up reports, prepared in the regular student project groups.

The experimental session was scheduled to last up to 3 hours and involved the students reading the instructions,[10] performing the experimental task and filling out the

[7] The self-assessment was developed based on the questionnaire proposed by Bois (2002, Appendix H) with the author's permission.
[8] Bois (2002), exploring people's ability to manage a business simulator (a version of the STRATEGEM game [Sterman 1987]), found that subjects who in the post-test questionnaire reported to have invested more effort tended to perform better. Jensen and Brehmer (2003) also observed that subjects, who seem to exhibit more of a *'fighting spirit'* tended to perform better (Ibid., p. 122).
[9] During the debriefing session we discussed with the students the optimal policy and the system dynamics model underlying the task.
[10] See the Managing CSIRTs handbook: *ManagingCSIRTs.pdf* included in the supplementary materials.

questionnaire.[11] At the start of the experiment, the subjects were ensured that all collected data would remain confidential and that their performance during the experiment would not have any impact on their grade in the system dynamics course. They were also promised that the person who performed best in each trial would receive a symbolic prize. This incentive is analogous to the one used by Moxnes (2004, see p. 144).[12]

Our experimental procedure was essentially the same as in the original study by Moxnes (2004). The subjects received only a written task instruction.[13] To maintain equivalence between experiments, the basic task instructions in our case mirror the original instructions, the only difference being that in our case the subjects are managers of CSIRT not a reindeer rangeland and in order to be successful should understand the net growth dynamics of the CSIRT capacity rather than lichen (compare the *Instruction* section in from the *Managing CSIRTs* handbook with *Appendix 1* in Moxnes [2004]). Our task instructions are additionally supplemented with a short introductory section where we explain what the CSIRTs are and what is the relationship between the CSIRT capacity and the service level (see pp. 1-3 in the *Managing CSIRTs* handbook). We felt that such introductory discussion was necessary in our case as most subjects are not likely to have any knowledge of CSIRTs; whereas in case of the reindeer-lichen task, the introduction was not needed, as most subjects were likely to have an intuitive understanding of the lichen-reindeer interaction. In the introductory section we also stipulated what was meant by the sustainable state of the system. This was done to prevent the misinterpretations of the term 'sustainable' detected on few occasions by Moxnes (2004, p. 147).

As in the experiment by Moxnes (2004), the subjects performed three trials. The simulator used in our study is an adapted version of the MS Excel based simulator developed by Moxnes (2004). The adaptation concerned primarily modification of the variable labels in the decision-making interface. Additionally, a few numerical properties of the simulator had to be adjusted to fit the particular context of our task version.[14] The customized simulator decision-making interface is presented in Figure 4.

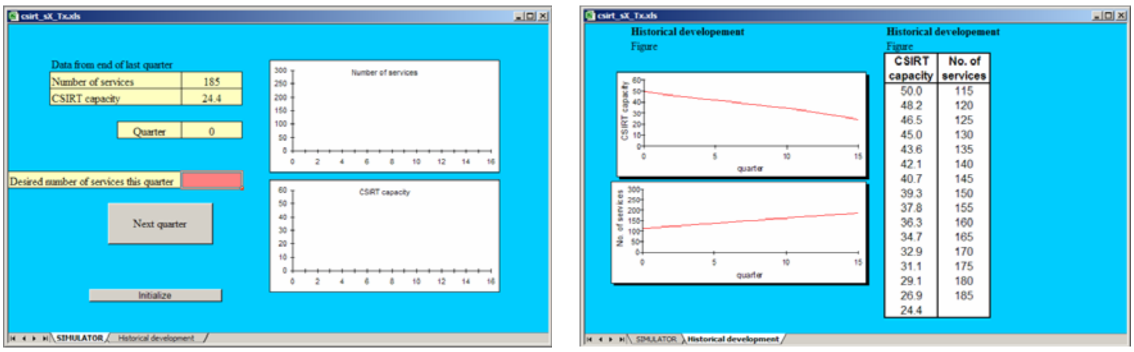


Figure 4 The simulator interface for managing CSIRT capacity. The interface is analogous to the one used in Moxnes (2004).[15]

In our case each trial consisted of 16 rather than 15 decision periods. The number of decision periods is increased to fit better the timescale assumed for our experiment – in

[11] See *Questionnaire.pdf* included in the supplementary materials.
[12] For the discussion of prize-based incentives in experimental studies see e.g. Bolle 1990.
[13] See the Managing CSIRTs handbook: *ManagingCSIRTs.pdf* included in the supplementary materials.
[14] See footnote , p. 8
[15] The simulator was adapted with permission from Professor Erling Moxnes.

our task the simulated decision periods correspond to quarter rather than years. Asking the subjects to manage the CSIRT for 4 years, yields 16 decision periods. As in the original study by Moxnes (2004), the subjects were asked to record all their decisions into the decision log. Additionally, we asked the subjects to use workbooks for performing any type of analyses or calculations they deemed necessary in the course of the experiment. Once all three trials were completed, the subjects were asked to complete the questionnaire.[16]

The debriefing session was conducted 2 days later. First, the best performers for each trial were rewarded with small prizes. Next, we discussed in detail the optimal solution of the task. Once all the questions from the students were answered, they were asked to work in their regular student groups[17] and to prepare a short report evaluating decision-support provided during the task.[18]

Results

During the experimental study we intended to collect both numerical decision logs for each of the subjects and a descriptive record of the subjects' performance. The descriptive record was supposed to be elicited through the workbook and questionnaire. Most of the subjects did not use actively the workbook. Hence, we limit our current analysis to the numerical logs of the subjects' decisions and data collected through the questionnaires. The section is divided into two parts. First, we present an overview of our results at a group level. Next, we present a more detailed view of the subjects' performance for each of the three trials.

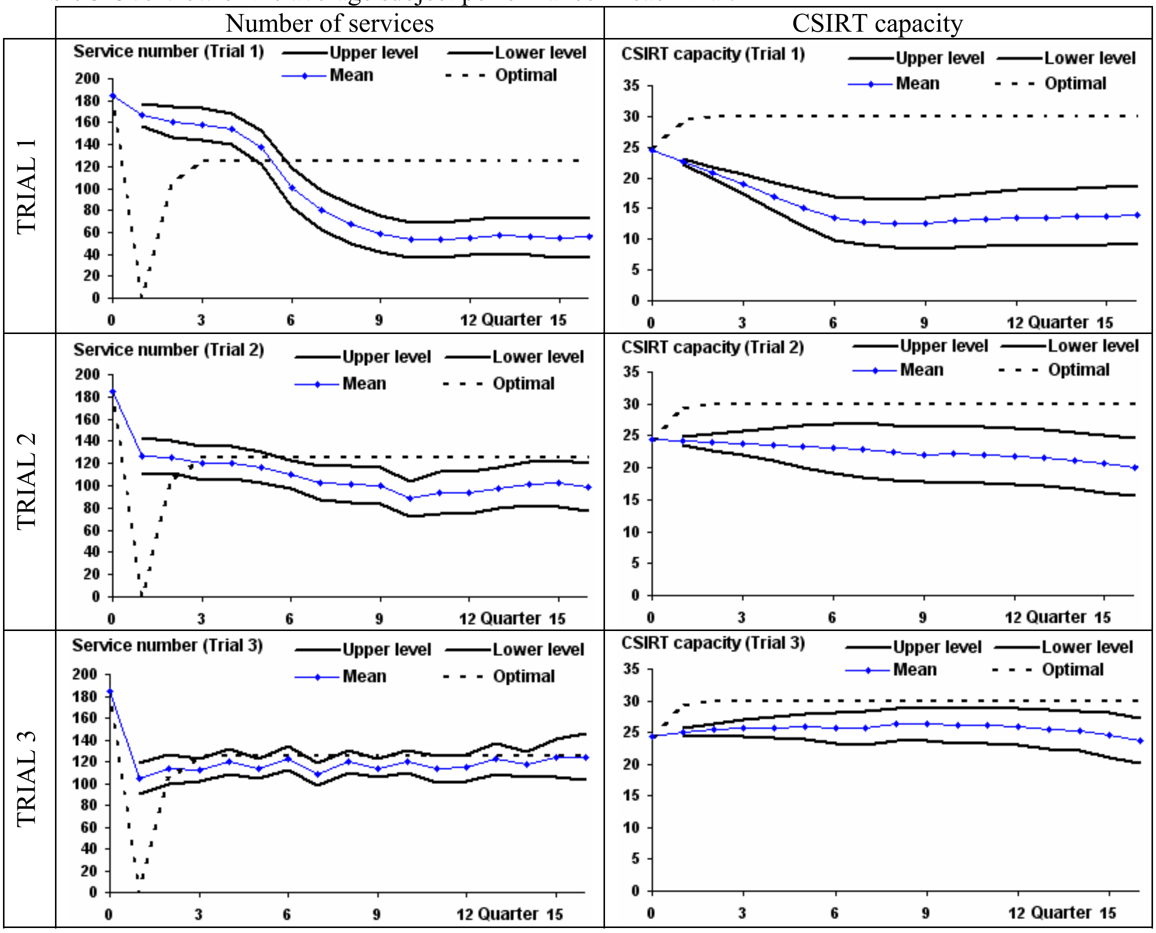
Results overview

Graphs collected in Table 3 show the average number of services and the average level of CSIRT capacity with 95% confidence intervals for each trial. A visual inspection of data suggests that over trials the average subject performance improves, with the greatest improvement occurring between the first and second trial. This is supported by the p-value tests which indicate that there is a significant difference ($\alpha=0.01$) between the average results of the first and second trial ($p_{1-2}=0.000<\alpha$), but no significant difference between the results of the second and third trials ($p_{2-3}=0.976>\alpha$). Comparing the average subject performance to the optimal performance we find a significant difference ($\alpha=0.01$) only between the average subject performance in trial 1 and the optimal solution: $p_{1-opt}=0.000<\alpha$; $p_{2-opt}=0.013>\alpha$; $p_{3-opt}=0.841>\alpha$

Also, the subjects' self-assessment of the performance seems to be consistent with the average group results. Figure 5 shows how the subjects evaluated their own performance in each of the trials when answering the first question in our post-test questionnaire. As we can see, most subjects felt that they performed rather poorly in the first trial and that their performance improved over the trials.

[16] See *Questionnaire.pdf* included in the supplementary materials.
[17] There were a total 14 student groups, most consisting of 3 students, with a few 2-student groups.
[18] See *Assignment.pdf* included in the supplementary materials, see also question 5 in the post-test questionnaire (*Questionnaire.pdf* included in the supplementary materials).

Table 3 Overview of the average subject performance in each trial.[19]



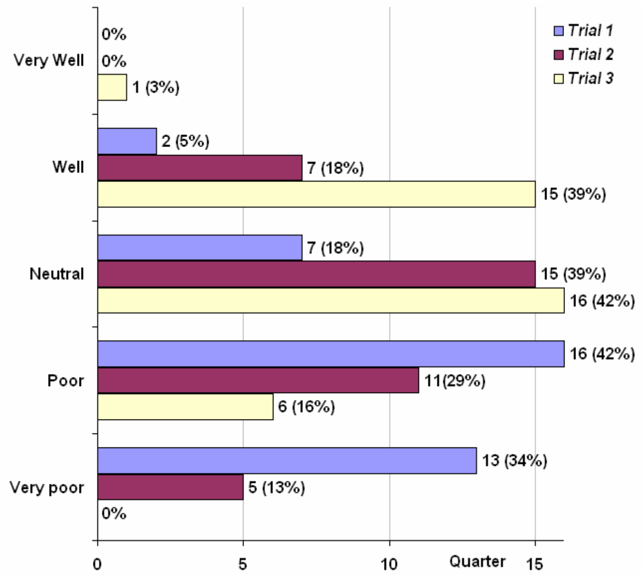


Figure 5 Subjects' assessment of their performance in each trial.[20]

---

[19] The seesaw pattern in the average number of service for Trial 3 is due to one subject going up and down to zero (see [2] Over-utilizers in Table 9, p. 23)

[20] See question 1 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).

In Table 4 we summarize how the subjects commented on their performance (four subjects did not provide comments):

- The majority of the subjects (26/68%) indicated that they used the first trial(s) for an experimental exploration of the simulator. In most cases (17 out of 26) the experimentation was conducted to gain a better understanding of the task. Five subjects experimented in search for the optimum CSIRT capacity level; four reported that between the trials they consciously experimented with different initial values of number of services.[21]
- Over half of the subjects (20 out of 38) indicated whether or not the simulator's behavior surprised them. Only 3 subjects (marked with pink in the surprise behavior column) indicated that they were never surprised by the simulator's behavior. The remaining 17 subjects indicated that they were surprised or confused by the system's behavior; three subjects (marked by the question marks) reported that they initially had no idea how the system would respond to their decisions.
- Half of the subjects (20 out of 38) commented on the relationship between the number of services and the CSIRT capacity. The majority of the subjects indicated that the relationship was difficult to understand (6 out of 19) or that it is linear (6 out of 19). Four subjects indicated they were surprised by the 'disproportional' changes in CSIRT capacity level. Five subjects attributed the lack of immediate adjustment in the capacity following reductions in the number of services to the inherent system delays.

Table 4 Comments recurring in the subjects' explications of their performance[22]

	Expectation (26/68%)			Surprise / confusing behavior (20)	CSIRT capacity- Number of services relationship (21/50%)			
	To understand (17)	To identify optimum (5)	To find an appropriate start value (4)		*"The more services, the less capacity"*. (6)	Unexpectedly drop in CSIRT capacity (4)	CSIRT capacity regeneration is delayed (4)	Difficult to understand (7)
S1	■			■				■
S2								
S3	■			■				■
S4				■		■		
S5		■						
S6				■		■		
S7				■		■		
S8				■	■			
S9			■	■			■	■
S10			■					
S11	■							
S12			■	■		■		
S13								
S14	■							
S15				?				
S16			■					
S17	■							
S18	■							
S19	■							
S20	■			■				■
S21	■							
S22	■							
S23		■		■	■		■	
S24		■		■	■			
S25	■			■	■			
S26	■			■	■			
S27		■		■				
S28				■	■			
S29				?				
S30								
S31	■			■				
S32	■						■	
S33	■			■				■
S34		■		■				■
S35				■				■
S36								
S37	■						■	
S38	■			?				

[21] The follow-up reports indicate that some of the subject could have used the *Initialize* button to test more start values (see Table 6, p. 20).
[22] See question 1 in the post-test questionnaire (*Questionnaire.pdf* included in the supplementary materials).

When asked about their perceived level of understanding of their task, zero subjects reported to have understood the task *Fully* nor *Not at all* (see Figure 6).[23] Seven subjects felt that they did not understand the task very well; 81% indicated that they understood the task *Well* (13) or *Reasonably well* (17).

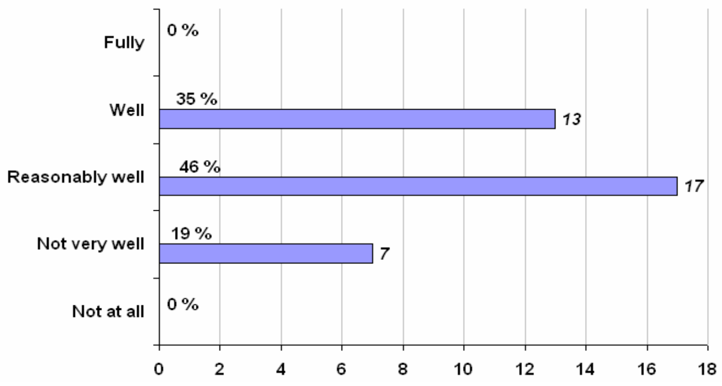


Figure 6 Subjects' perceived understanding of the task.[24]

Most of the subjects provided short descriptions of the task. In Figure 7 we summarize their explications:

- Among the 30 subjects who felt they understood the task well, 25 provided short explications. 17 of the 25 subjects (68%) indicated that their task was to obtain the highest number of services for which the CSIRT capacity is stable. Two indicated that their objective was to obtain an 'optimal'/'sustainable' number of services, five indicated that they were to obtain a 'good' balance between number of services and the CSIRT capacity. Only 3 subjects indicated that they were supposed to reach the highest sustainable number of services in the shortest time.
- Among the 7 subjects who reported not understanding the task very well, five provided additional information. Two of these subjects stated quite correctly that they were supposed to achieve the highest sustainable number of services in the shortest time. One of the subjects stated that he or she did not understand fully the relationship between the number of services and CSIRT capacity. One stated that the main objective was to avoid CSIRT capacity depletion, and one indicated that it was difficult to understand how the two competing goals of providing the highest possible number of services and sustaining the CSIRT capacity could be reconciled.

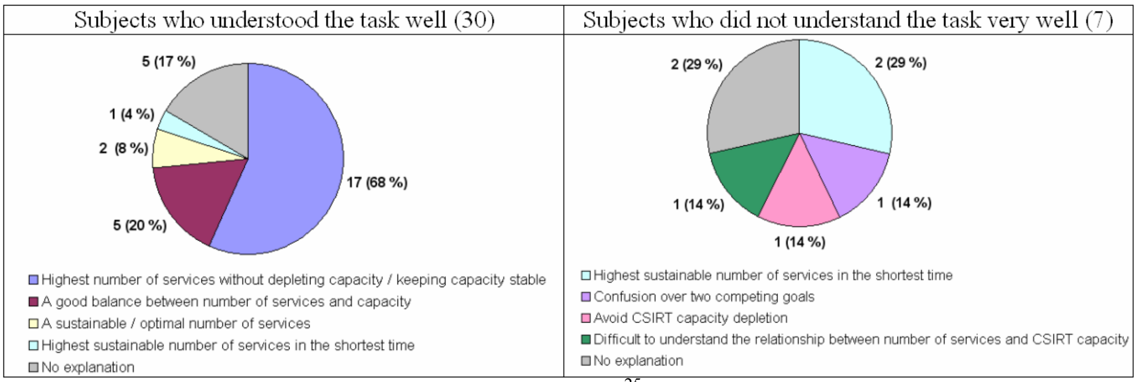


Figure 7 Subjects' interpretation of the experimental task.[25]

[23] One subject did not answer this question.

[24] See question 2 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).

Figure 8 shows how many subjects felt they could control the system:

- Only two subjects reported that they never or only seldom felt that they are able to control the system. One of these subjects commented that the experienced lack of control was due to lack of experience with this type of task.
- Among the eleven subjects who felt they controlled the system only occasionally, three reported achieving more control after the initial trial(s) and four indicated that the system did not behave always as expected.
- In case of the subjects who reported to have control over the system most of the times, the majority (14 out of 20) indicated that they felt in control only after the first trial(s).
- 13% (5 subjects) reported to have a full control of the system: One of these subjects seemed to have misinterpreted the question as referring to the control over the simulator application, two indicated only that they had no problems controlling the system – it is possible that they also misinterpreted the question. Two out of the 5 full-control subjects stated that they felt they controlled system fully after the 1st trial.

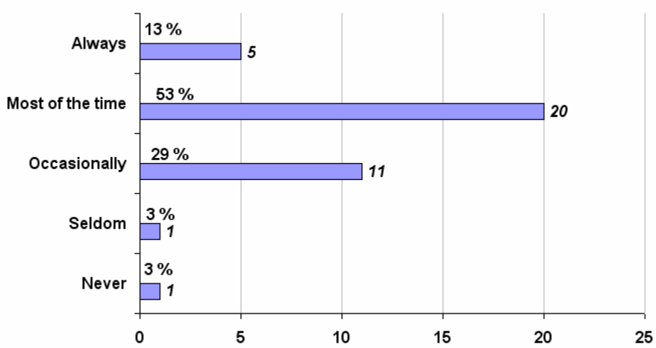


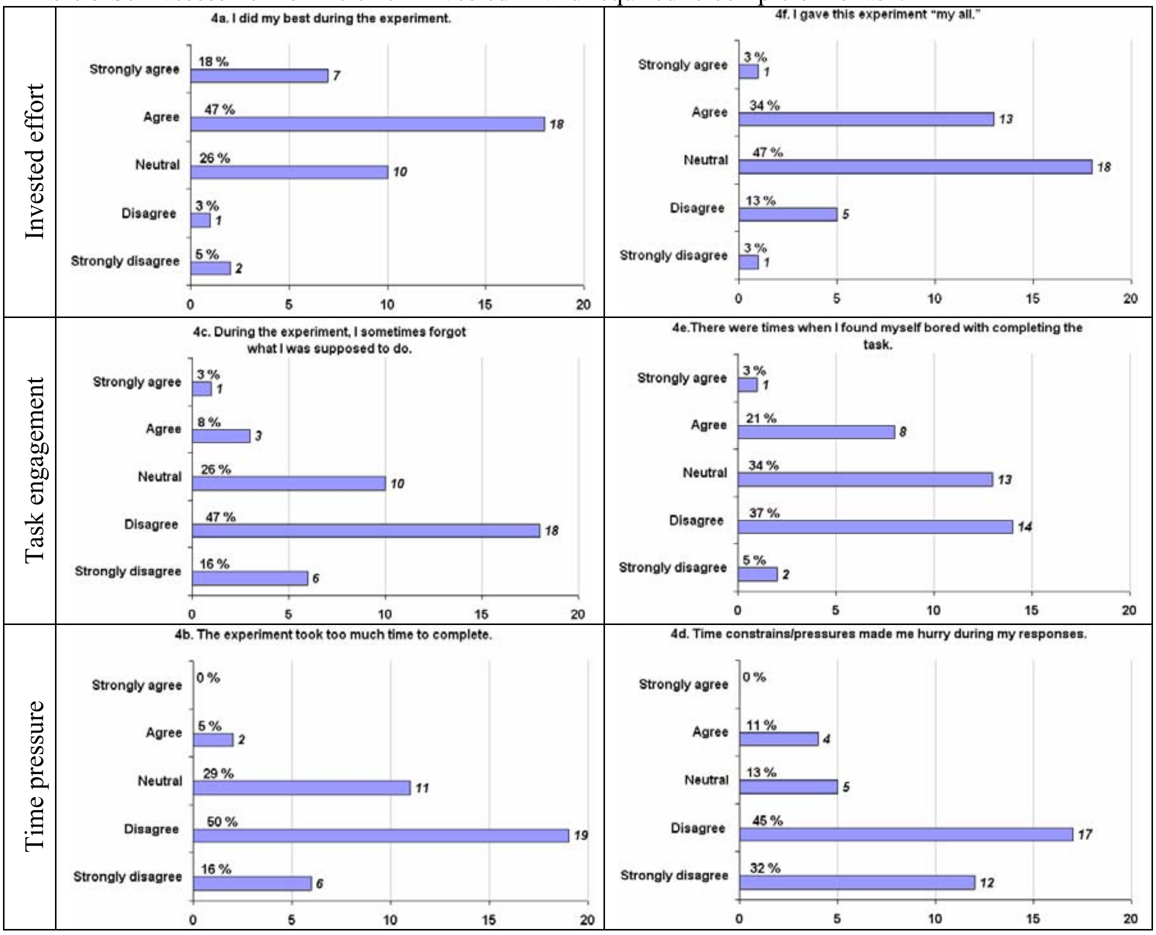
Figure 8 Subjects' perceived degree of control over the system.[26]

In Table 5 we summarize the results of the short effort self-assessment survey. We grouped the six questions into three categories:

- Invested effort: Most of the subjects (25/65%) felt that they did their best during the experiment (*Strongly agree* or *Agree* in question 4a). Only one of these subjects reported not investing the maximum effort (*Disagree* in question 4f). Twelve of the subjects felt they had invested much of the effort in dealing with the task (*Strongly agree* or *Agree* in question 4f); twelve assessed their effort as *Neutral* (question 4f).
- Task engagement: Most of the subjects also found the task only reasonably engaging. Most of the subjects reported to stay focused during the task (63% disagreed or strongly disagreed with the 4c statement). However, only 42% disagreed or strongly disagreed that they at times felt bored with the task (4e).
- Time pressure: None of the subjects had a strong impression of the time pressure (0% for *Strongly agree* in questions 4b and 4d). Two of the subjects felt that the experiment took too much time to complete (question 4b) and four reported being under time pressure. Most of the subjects (66%) felt that the experiment was not too time consuming (question 4b). 77% did not feel themselves as working under time pressure.

[25] Comments given to question 2 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).
[26] See question 3 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).

Table 5 Self-assessment of the effort invested in and required to complete the task.[27]



Finally, Figure 9 presents the subjects evaluation of the simulator. One of the subjects did not answer the question. None of the subjects evaluated the simulator as providing an excellent support to the decision-making process. Most felt that the simulator provided merely sufficient support:

- Two of the subjects evaluated the support as poor. Both seemed to be unable to develop a good understanding of the task given the instructions; one suggested that the instructions were too wordy and suggested that more factors should be reported on in the simulator, the other recommended an oral presentation of the task in plenum.
- Three subjects evaluated the support as insufficient: one indicated that the participants should be alerted to run a regression analysis to identify the 'curve'; one concluded that the experiment was about making decisions *without* all the information and pointed out that the historical data were confusing.
- Twenty subjects (54%) found the simulator as providing sufficient support. Twelve subjects (32%) assessed the simulator as good.
  Twenty six of the 32 subjects commented on the simulator: 3 indicated that the simulator interface was easy to use and 7 pointed out that the graphs were quite helpful. Seven pointed out that they would like to have also a tabular overview of their past decisions. Two of the subjects indicated that the instructions were too wordy and one indicated that the task could be better tackled in groups where one could discuss the course of actions with others. Only three of the subjects wished more information regarding the changes in CSIRT capacity; three others pointed out that information about more 'factors' should be provided.

[27] See question 4 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).

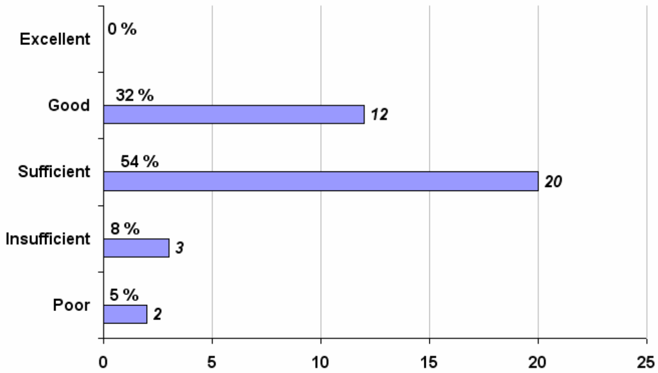


Figure 9 Facilitation of the decision-making process during the experiment.[28]

The subjects' responses to question 5 (Figure 9) indicate that overall the subjects were rather satisfied with the decision-making support they received during the experiment. This evaluation changes quite dramatically in the reports the students prepared following the debriefing session.[29] The reports were prepared in the regular project groups with 2-3 students each (13 groups in total). Figure 10 presents an overview of the type of comments made by the students; with the hindsight of the optimal solution most of comments concerned the experimental instructions.

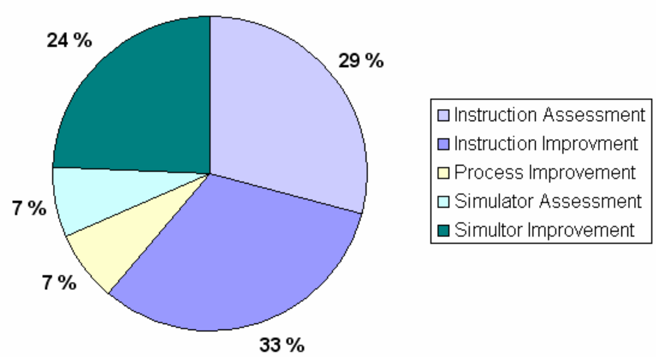


Figure 10 Percentage of comments made regarding instruction assessment and improvement, process improvement, and simulator assessment and improvement.

Assessing the instructions, most of the students criticized the volume of the instructions. Ten out of the 13 groups indicated that there was too much background and irrelevant information provided, four suggested that the instructions would be easier to read if special formatting (such as bullet points, different font types, etc.) were used to clearly indicate the most important information in the handbook, five wished shorter, more to the point instructions.

Suggesting improvements to the instructions, roughly half of the groups (6) indicated that they would like the instructions to specify the task objective more clearly. Only two of the groups indicated that the *Instructions* section in the handbook explained the task well. Two other groups pointed to the first paragraph as an example of a clear instruction. Two groups indicated that the description of CSIRT capacity growth dynamics was useful. One group indicated that the instructions did not make it clear that

---

[28] See question 5 in the questionnaire (*Questionnaire.pdf* included in the supplementary materials).
[29] During the debriefing session we discussed the optimal solution in detail. The presentation featured both Table 1, p. 7, and Table 2, p. 8. See *Assignment.pdf* in the supplementary materials for the assignment text.

there is only one optimal solution. Four groups on the other hand indicated that they considered the description of the utilization rate as confusing.

Nearly all of the groups (11 out of 13) commented on the provided historical time series. Six groups considered the historical time series useful. Five groups indicated that they would like to have the time series accompanied by the calculations similar to those presented in Table 2 (see p. 8). Two commented that the series suggested the linear relationship between the number of services and the CSIRT capacity level. Two other groups proposed that the instructions should emphasize that the historical development does not produce a sustainable situation.

Six groups commented on the insufficient explication of the CSIRT capacity net growth curve in the instructions. Three groups recommended supplementing the instructions with the net capacity growth curve. Three other wished for a more precise and detailed explication of the role of the nonlinear net growth rate.

Three groups recommended inclusion of the system dynamics model into the instructions. And three groups commented on the questionable realism of the optimal solution requiring all the services to be halted for one quarter.

Two groups commented that they did not feel comfortable with the English-language instructions. One recommended an oral presentation of the task. Another group felt that there were too many papers included in the experimental material package. Yet another felt that it was distracting to enter decisions into the paper decision logs.

Regarding the simulator, most of the groups (7) indicated that it was easy to use and understand. Three groups commented positively on the graphs tracing the number of services and the CSIRT capacity. The majority of the groups (8) indicated, however, that the simulator lacked a detailed overview of the decisions made by the subjects. Most of the groups would like to have available such an overview in the tabular format. Two groups would like to have the net capacity growth rate reported along side other outcome figures. Two other groups would like the report to be extended to include also a change in capacity. Three other groups voiced the need for more detailed reporting without providing any specifics.

Six groups commented on the usability aspects of the simulator's interface pointing out poor choice of color scheme, wording inconsistency, and inconvenience caused by the need to press return prior using the 'Next quarter' button. Three of the groups pointed out that the subjects could perform an unlimited number of trials by re-initializing the simulation with the *Initialize* button.

The issues raised most frequently in the student reports (by 3 or more groups) are summarized in the bar chart presented in Table 6.

Table 6 Issues raised by 3 or more students in the follow-up reports.[30]

		Group 1	2	3	4	5	6	7	8	9	10	11	12	13	
I.A.	Too much irrelevant information	■	■	■		■		■	■	■	■	■		■	**10**
	Historical data useful			■	■					■	■		■	■	**6**
	Information about CSIRT-capacity utilization rate confusing	■				■	■							■	**4**
	Diagram illustrating CSIRT-capacity useful			■							■			■	**3**
I.I.	Provide information about CSIRT capacity net growth curve	■		■		■	■						■	■	**6**
	State the goal more clearly		■			■			■	■			■	■	**6**
	Make clear the need for mathematical calculations	■	■			■		■			■				**5**
	Shorten/compress instruction	■						■				■	■	■	**5**
	Emphasize important information				■					■			■	■	**4**
	Provide SD model					■	■				■				**3**
	The optimal solution should not require the 0-service level					■				■		■			**3**
S.A.	Easy to understand		■	■	■			■	■		■			■	**7**
	Graph of historical data useful				■		■						■		**3**
S.I.	Provide a detailed history decision made	■	■				■	■		■		■	■	■	**8**
	Reporting more factors	■					■				■				**3**
	Get rid of the ¨Initialize¨ button				■	■		■							**3**

Review of the individual performance

As indicated above most subjects felt that their performance improved over the three trials. During the first trial only 2 subjects assessed their performance as good, 7 assessed their performance as average, and 29 as poor or very poor (see Figure 5, p. 13). By the third trial only 6 subjects assess their performance as poor; the rest is split evenly between the subjects who consider their performance as average and the subjects who consider their performance as good (with one giving the performance the maximum mark). In this section we present a graphical record of the individual subjects' performance in each of the trials. For each of the trial we classify the subjects' decision sequence into different categories. We also indicate how much effort the subjects, who were assigned to the particular category, reported to invest in dealing with the task.[31]

Table 7 presents the results of the first trial. For this trial we classify the subjects' behavior into one of the following four categories: (1) successful performers – those who make significant reductions in the number of provided services to restore the CSIRT capacity, (2) gradual adjusters – those who gradually decrease the number of provided services but manage to avoid complete depletion of the CSIRT capacity, (3) explorers – those who try out different policies to test how the system responds, and (4) unsuccessful performers – those who deplete the CSIRT stock. We find no significant (α=0.05) correlation between the subjects' performance in this trial[32] and the reported effort investment[33] (Pearson correlation r=0.14).

The results of the second trial are presented in Table 8. By this trial most of the subjects seemed to understand how to restore the CSIRT capacity (category 1). Still there were

[30] See *Assignment.pdf* in the supplementary materials for the assignment specification.

[31] Elicited through question 4f in the questionnaire (see *Questionnaire.pdf* included in the supplementary materials, see also Table 5, p. 17)

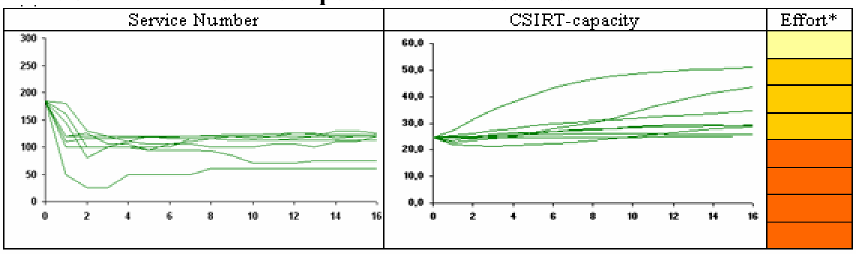
[32] We considered that subjects classified as category 1 to do well.

[33] Subjects who answered *Strongly agree* or *Agree* in question 4f in the post-test questionnaire (see *Questionnaire.pdf* included in the supplementary materials, see also Table 5, p. 17).

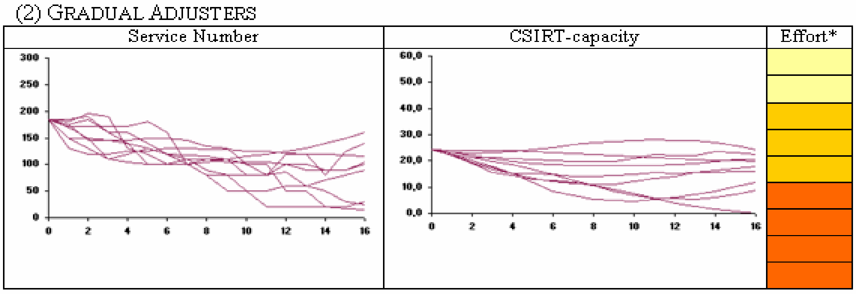
some who explored the system (category 3) and depleted the CSIRT capacity stock (category 4). Others seemed to have problems in identifying the appropriate level of services and locked themselves into the situations that were suboptimal; we classify them as ‘over- and under-utilizers (category 2). Again, we find no significant (α=0.05) correlation between the subjects’ performance in this trial and the reported effort investment (Pearson correlation r=0.11).

Table 9 presents performance of the subjects in the last, third trial. In this trial most of the subjects can be classified either as successful performers (category 1) or over-utilizers (category 2). In the over-utilizers group, all but three subjects made the required significant reduction at the outset. However, instead of maintaining a stable level of services they tended to gradually increase them; consequently, two depleted the capacity just before the trial terminated. Two of the subjects in trial 3 are singled out as under-utilizers who initially reduced the number of services radically, but who failed afterwards to increase sufficiently the service level. Also in this trial we find no significant (α=0.05) correlation between the subjects’ reported effort investment and their performance (Pearson correlation r=0.06).

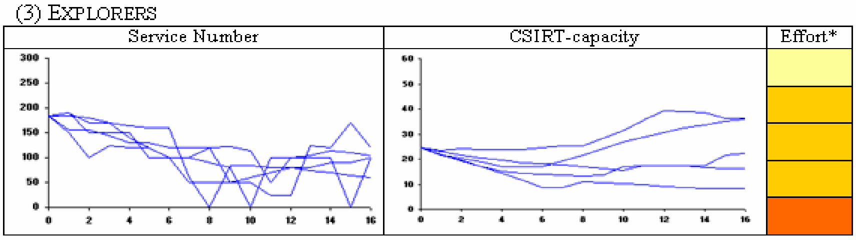
Table 7 Trial 1 – individual performance overview.



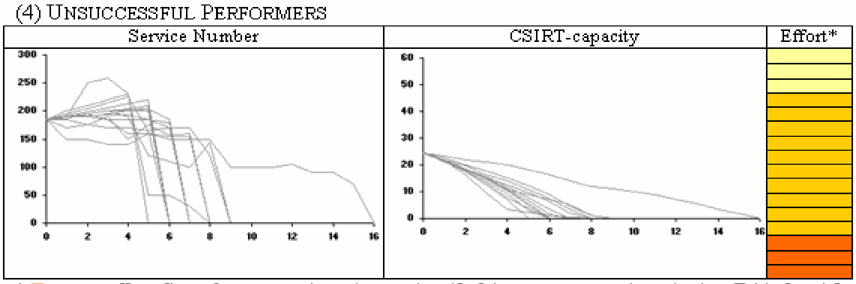
(2) GRADUAL ADJUSTERS



(3) EXPLORERS

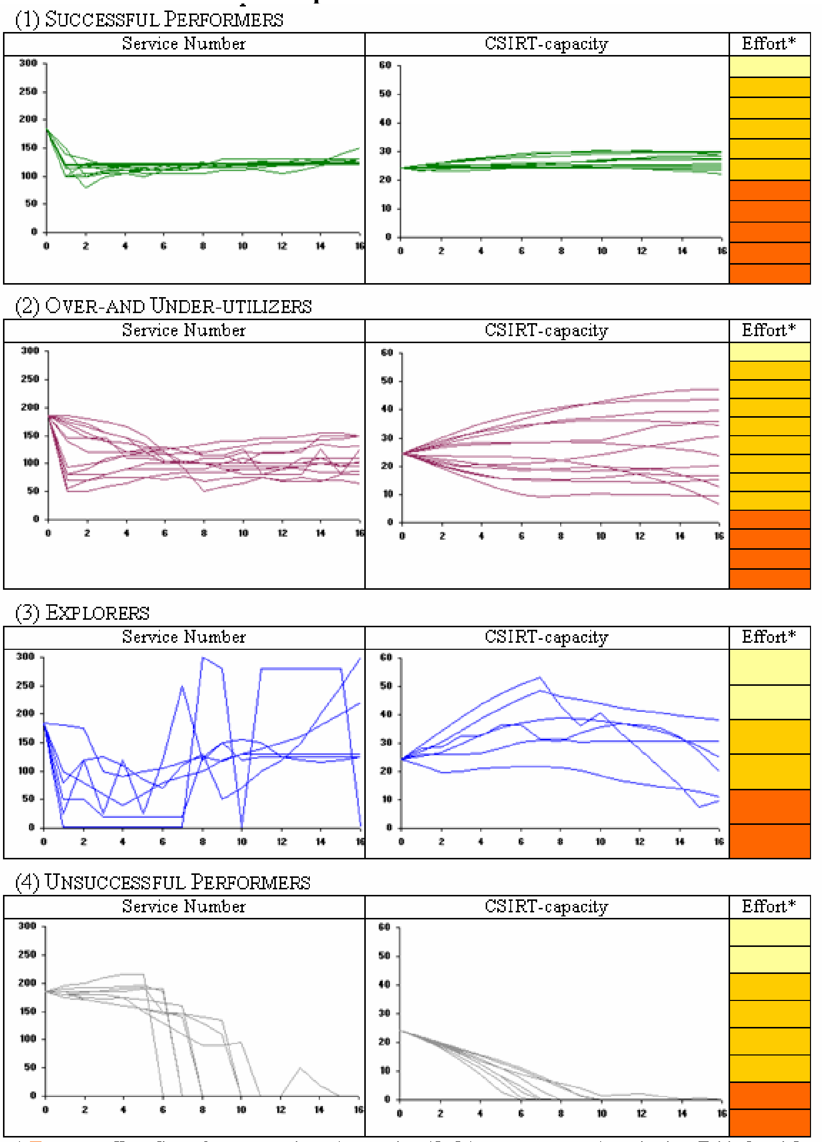


(4) UNSUCCESSFUL PERFORMERS



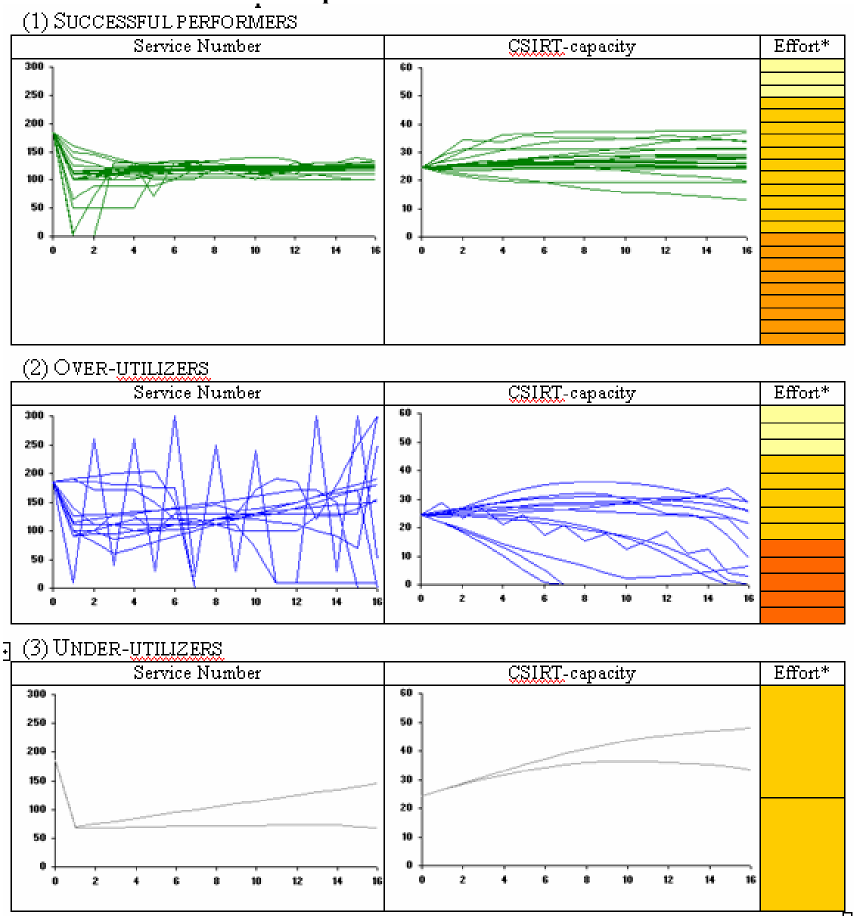
* ■ – most effort; *Strongly agree* or *Agree* in question 4f of the post-test questionnaire (see Table 5, p. 16)
■ – neutral effort; *Neutral* in question 4f of the post-test questionnaire (see Table 5, p. 16)
■ – little effort; *Disagree* or *Strongly disagree* in question 4f of the post-test questionnaire (see Table 5, p. 16)

Table 8 Trial 2 – individual performance overview.



* ■ – most effort; *Strongly agree* or *Agree* in question 4f of the post-test questionnaire (see Table 5, p. 16)
■ – neutral effort; *Neutral* in question 4f of the post-test questionnaire (see Table 5, p. 16)
■ – little effort; *Disagree* or *Strongly disagree* in question 4f of the post-test questionnaire (see Table 5, p. 16)

Table 9 Trial 3 – individual performance overview



* – most effort; *Strongly agree* or *Agree* in question 4f of the post-test questionnaire (see Table 5, p. 16)
– neutral effort; *Neutral* in question 4f of the post-test questionnaire (see Table 5, p. 16)
– little effort; *Disagree* or *Strongly disagree* in question 4f of the post-test questionnaire (see Table 5, p. 16)

Discussion

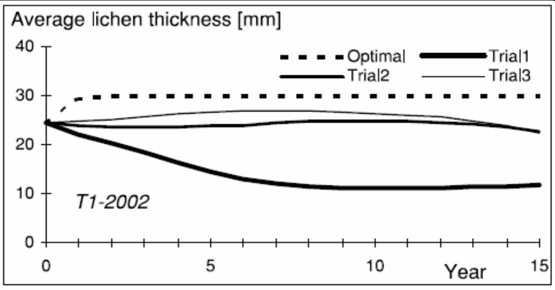
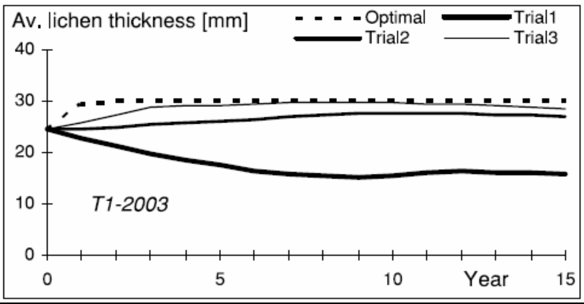
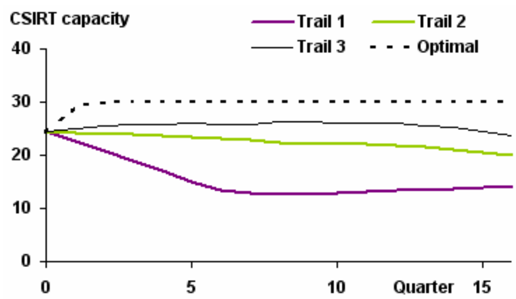
Our main hypothesis for this study (see p. 8) was that with the CSIRT management task we observe mismanagement of renewable resource similar to this observed by Moxnes (2004) in case of the perennial pasture management task. To what degree our results support this hypothesis is discussed first; we compare the average subject behavior and discuss the likely decision rules. Next, we discuss our findings regarding the impact of task presentation and simulator on the decision-making process. Finally, we comment on the quality of collected data.

Managing CSIRT capacity vs. managing perennial pastures

The average versus optimal subject performance

In Table 10 we reproduce the average levels of lichen reported by Moxnes (2004) along side the average levels of CSIRT capacity observed in our study. In all cases, the result for Trial 1 stands out as the most distant from the optimal solution. A rapid improvement occurs between Trial 1 and Trial 2, with a smaller improvement between Trial 2 and Trial 3. The visual analysis of the average results suggests that the behavior observed in case of the CSIRT management task is indeed similar to the behavior observed in case of the perennial pasture management task. Hence, the results seem to support our initial hypothesis regarding the expected behavior patterns (see p. 8).

Table 10 Comparing the average subject performance in the perennial pasture and the CSIRT capacity management tasks.

The perennial pasture management*	The CSIRT capacity management
 	

* The figures are reproduced form Figure 7, p. 148, in Moxnes 2004 with the permission from the author.

The results presented in Table 10 suggest that our average subject, as the average subject from the original study (Moxnes 2004), did not follow the optimal policy in the first trial. Moxnes attributes this to the subjects' failure to develop an appropriate mental representation of the task and suggests that most subjects try to manage the system with a static model in mind that says: *"the more animals, the less lichen, and vice versa"* (2004, p. 151). This mental model is quickly contested by the feedback from the simulator: reductions in the prey population (reindeer herd size) do not necessarily result in an increase in the renewable resource level (lichen). Given such feedback and in absence of any better explanation about the system behavior, most subjects are likely to continue to reduce the herd size. In the meantime, to explain a behavior that is not entirely consistent with the assumed static model, the subjects are likely to develop various auxiliary hypotheses (e.g., that there are unspecified delays in the system or that the behavior is influenced by other unknown factors, etc.).

In our study we also identified some subjects who reported to be confused by the system's behavior, suspected that the system involved some time delays, or seemed to perceive the relationship between the CSIRT capacity and number of services as linear (see Table 4, p. 14). Moxnes suggests the counterintuitive system response lead the subjects to make their decisions using a simple anchoring-and-adjustment rule (2004, p. 152-153). In the following subsections we consider to what degree our results seem to support the Moxnes' hypothesis.

The anchoring-and-adjustment decision rule

The simple anchoring-and-adjustment decision rule proposed by Moxnes (2004, see equations [4] and [5], p. 152) assumes that the adjustments in the desired herd size (the prey population) depend on the perceived discrepancy between the current and desired level of lichen (the renewable resource). The initial analysis of the descriptive data from our experiments suggests that also our subjects adjusted the desired number of services in response to the outcome feedback. However, we do not find evidence that the adjustment followed the rule suggested by Moxnes. The proposed rule requires that the subjects have some desired level of renewable resource they want to reach. The questionnaire responses suggest that our subjects did not have (at least initially) any particular, desired level of CSIRT capacity. Rather, the number of services was adjusted in response to *the observed changes in the CSIRT capacity level*. The negative changes tended to induce reductions in the number of services, the positive or no change most frequently usually led to no-change or to increase in the number of services. Depending on the size of the CSIRT capacity change, the adjustments of the desired number of services were more or less aggressive. Once the CSIRT capacity was stabilized, the subjects tried to increase the number of services to see whether the current CSIRT capacity might sustain it. Over time, more and more subjects realized that a significant reduction in the number of services at the outset is needed and that the number of services may be subsequently increased without depleting the CSIRT capacity.[34] However, it seems that depending on the aggressiveness of their searches, the subjects identified a higher or lower sustainable service level. Most subjects who managed to avoid CSIRT capacity depletion providing a reasonable number of services seemed quite content with their performance (see Figure 8, p. 16). This suggests that the highest number of services that *happened* to allow them to prevent CSIRT capacity depletion was automatically perceived as the highest *feasible* service level. Several subjects indicated that they might be able to identify a higher sustainable service level if they were given a new chance. Even the subject who in the third trial almost precisely hit on the optimal strategy (see Figure 11) expressed doubts whether the *highest* sustainable service level was reached.

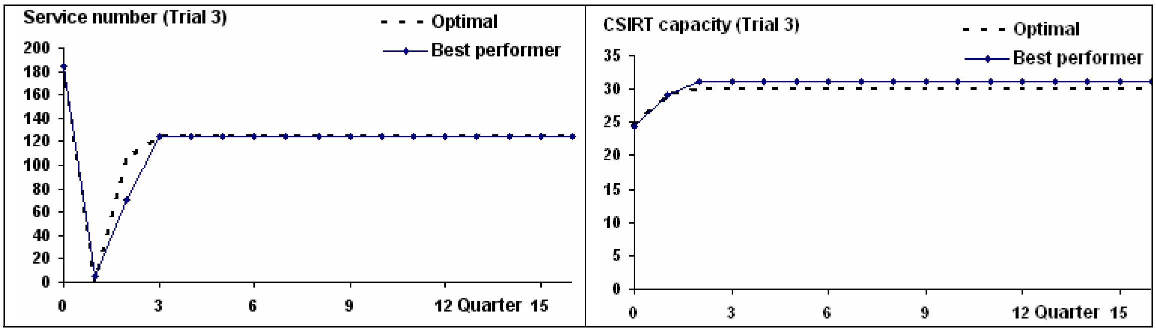


Figure 11 Best performer in trial 3.

Analyzing the subjects' responses collected through questionnaires, we did not find any comments indicating that the subjects were aware that the optimal solution requires the service-related utilization rate to equal the maximum net capacity growth rate, or considered the maximum net growth rate at any point in more detail. It seems that while most of the subjects picked up on the key task objectives from the instructions (see Figure 7, p. 15) they gained only a limited understanding of the underlying system's dynamics. The understanding of the dynamics was developed through the trial-and-error explorations with the simulator (see Figure 5, p. 13, and Table 4, p. 14, see also Figure 6, p. 15, and the associated discussions). This was further confirmed by the follow-up reports in which the students were asked to evaluate both the experiment instructions and the simulator with hindsight of the optimal solution.

[34] See Table 7, p. 21, Table 8, p. 22, and Table 9, p. 23

The role of experimental decision-making environment

Previous research results suggest that understanding of and performance in dynamic system may depend on the effort people invest in dealing with a particular problem (Bois 2002, Jensen and Brehmer 2003). We did not find any significant correlation between the reported effort investment and the subjects' performance, and observed only that the performance of most subjects improved over the trials (see Table 7, Table 8, Table 9, pp. 21-23, see also Table 3, p. 13). This is consistent with the subjects' reports that when trying to understand how to manage the system they relied heavily on experimentation (see Table 4, p. 14). While most of the subjects considered the decision-making support during the experiment as sufficient (Figure 9, p. 18) and were quite content with their control of the system (see Figure 8, p. 16), their evaluation changed with the hindsight of the optimal solution. In the reports filled after the debriefing session,[35] most student groups strongly criticized the instructions and recommended various extensions of the simulator (see Table 6, p. 20).

Given the subjects' performance, we expected the student critique to focus on how the nonlinearity of the CSIRT capacity net growth could be better incorporated in the decision-making environment (i.e., the instructions and the simulator). While almost all groups commented on the issue, suggestions for improvements remained quite generic: provide "more descriptive information" about the CSIRT capacity net growth rate, report in "a greater detail" about the CSIRT capacity change in the simulator, etc. (see Table 6, p. 20). On the other hand, all groups elaborated about the problems experienced when dealing with the experiment instructions.

Almost all groups pointed out that the experiment's instructions were difficult to understand. Because our experiment instructions included background information about CSIRTs, they were more complex than the instructions used in the original experiment by Moxnes (2004). Still, we believe that some of the points made by the students are also valid in the context of the original instructions. Many student groups criticized the background information as superfluous but did not seem to have a problem identifying the *Instructions* section of the *Managing CSIRTs* handbook as the kernel of the experiment task instructions. This section almost precisely mirrors[36] the instructions used in the original study by Moxnes (2004). The students' feedback suggests that the instructions failed to state the ultimate task objective in a clear enough manner. Most of the groups indicated that they would like the task goals to be emphasized more clearly in the text. Several groups reported problems with interpreting the 'sustainability' requirement; a couple of groups indicated that it was difficult to deal with the instructions because they were not in their native language. All these comments suggest that the students experienced a significant cognitive burden just trying to understand their task.

The cognitive research indicates that task instructions which induce much cognitive burden are likely to impede learning (see e.g. Chandler and Sweller 1991). For the learning to occur there must be enough cognitive resources available to support the process. If all or most of the resources need to be directed just to understand and follow

[35] During the debriefing we presented the optimal solution, featuring Table 1, p. 7, and Table 2, p. 8.

[36] We added bullet points emphasizing the specification of the CSIRT capacity net growth curve and the service-related utilization rate.

the task instructions, the learning will not be possible.[37] Given this, with the reported difficulties with understanding the basic task objective and following the prescribed process,[38] it may be not surprising that most of the subjects developed an understanding of the system's nature only during the 1st and 2nd trials.

Although our instructions were more complicated than in the Moxnes' study, the instruction's kernel and the decision-making process were almost identical. Therefore, we believe that some subjects in the Moxnes' study could have experienced similar difficulties. Moxnes detected that some of his subjects misunderstood the 'sustainability' requirement. Many of his subjects could have experienced additional cognitive load due to the fact that the instructions were not presented in their native language. Other points made by our students could have not been detected, since the subjects in the original study were not asked explicitly to evaluate the provided task materials.

Would the performance improve significantly if the instructions were provided in the subjects' native language, if they were simplified and condensed to their limits with a clearly outlined objective and if the decision-logging would be fully automated? We believe that such modifications could lead only to a partial improvement in the performance. To improve the performance one needs to understand the role of the renewable resource net growth nonlinearity. The fact that the student follow-up reports did not elaborate in any great detail about how to deal with this issue might be indicative either of the fact that the students still did not fully understand the system's dynamics, or that they could not identify any instruction/simulator features that could help in communicating the message, without a blunt presentation of the calculations featured in Table 1, p. 7, and Table 2, p. 8:

Given some 'improvement' ideas, e.g., removing the information about utilization rate or the description of the CSIRT capacity net growth curve, it is pretty clear that some students still misunderstood the problem or misperceived what information is necessary to solve it successfully. This may be due to the particular group not being motivated to do the assignment well or not paying attention to the discussions in plenum. Alternatively, one could argue that the students failed to fully understand the system during the presentation because the presentation did not engage them in the active explorations of the dynamic system. Our presentation had a traditional slide-show format; although the students were encouraged to engage actively in the discussions, this has not occurred. The hypothesis about the need of active exploration seems consistent not only with our experimental results (most subjects reported to understand the system through its active explorations during the first trials), but also with the results by Moxnes (2004, 1998) and others (Jensen and Brehmer 2003, Dörner 1989 (1996), Crossman and Cooke 1974). Still, the problem is how to facilitate the exploration so that it leads to accurate mental models. We believe that answering this question requires precise understanding of the reasons for the observed failures. In this study, we intended to gain such a more precise understanding based on the data collected through the workbooks and the questionnaires. As discussed in the following section, the results of this effort turned out to be somewhat disappointing.

---

[37] The dynamics of interaction between the instruction and the learning are discussed in more detail in a parallel paper intended for this conference and co-authored by one of this paper's authors (Sawicka and Molkenthin 2005).

[38] Many subjects commented on the inconveniences caused by the requirement to keep the paper log of their decisions. (see Table 6, p. 20)

Collecting data on dynamic decision-making processes

As indicated earlier, we enhanced our experimental procedure with the workbooks and questionnaires. The workbooks were supposed to provide us with an insight into the on-going decision-making process. The questionnaires provided a post-test assessment of the performance and understanding.

Most of the subjects did not use the workbooks actively during the task solving. This may be due to the fact that the instructions only gave the subjects an option to log their calculations and comments there. However, we believe that most of the subjects did not ignore the workbooks just because using them was optional. Rather, it seems likely that using them would disrupt the decision making process. As indicated earlier, in the follow-up reports some of the groups commented on the inconvenience of having to log all their decisions on paper. Providing a written account of the decision-making rules is likely to cause an even greater distraction.

Lack of the insight into the on-going decision-making processes impedes our ability to trace the way in which the subjects developed their understanding of the system. Although in the questionnaire we asked the subjects to comment on all three trials, obviously the comments were biased by the overall subjects' performance. For example, many subjects assessed their performance in the earlier trials in the context of their performance in the subsequent trials; if they were asked right after a particular trial, their assessment could have been different.

Finally, it needs to be noted that the failure of workbooks to provide an effective tool for tracking the decision-making indicates that other methods of data collection should be employed for obtaining more in-depth insight into the way in which the subjects make their decisions.

Future research directions

The results of the experimental study presented in this paper open for us a couple of research avenues.

First, they indicate that similar types of misperceptions and mismanagement of a particular dynamic structure may occur in different task contexts. Findings by Moxnes and Saysel (2004) suggest people manage dynamic challenges better in more familiar task contexts. While the initial performance of our subjects was poor, it improved as they gained experience thorough the trials. Similar observations were made by, for example, Paich and Sterman (1993), Moxnes (1998, 1998, 2000, 2004), Jensen and Brehmer (2003). Also, our results confirm that the improved performance does not seem to be accompanied by a more accurate understanding of the system's behavior. The mental models people develop during their explorations tend to be simplistic; their explanatory power reduced to the limited experience acquired by a particular individual. A tendency to engage in an experienced-based learning and reasoning in uncertain and complex situations has been observed in the cognitive psychology literature (see e.g., Kahneman, Slovic et al. 1982, Hastie and Dawes 2001).[39] The same literature also reports on a range of reasoning fallacies that people are prone to commit when dealing with complex problems (see e.g., Johnson-Laird and Wason 1977, Galotti 1999). Given the nature of dynamics systems, heuristics based on the limited experience are likely to

[39] For a system dynamics analysis of the psychological mechanisms likely to facilitate the way in which people develop their understanding of risk see Gonzalez 2002, Gonzalez and Sawicka 2003, Sawicka 2004.

be faulty; hence, even when people believe they control the system, a small change is likely to leave them helpless. Our results indicate that textual descriptions may not be the most effective way of communicating the system's dynamics. Given such description, most people seem to acquire only a vague understanding of the system. To improve it, they tend to engage in trial-and-error; in most cases, these unguided explorations lead to faulty mental models. Our future research focuses on identifying how the systems' dynamics may be communicated in a more effective manner.[40]

Second, our results suggest that problems experienced by CSIRTs may in part be due to managers falling prey to the misperception of dynamics of the CSIRT capacity growth. Further research is necessary to scrutinize the validity and applicability of the proposed structure. Still, it seems that the proposed case may provide a useful classroom aid to illustrate some challenges in the CSIRT management.

Bibliography

Bois, R. J. (2002). *Decisions within Complex Systems: An Experimental Approach using the STRATEGEM-2 Computer Game*. Rockefeller College of Public Affairs and Policy. Albany, NY, State University of New York. **Ph.D.**

Bolle, F. (1990). High reward experiments without high expenditure for the experimenter? *Journal of Economic Psychology* **11**: 157-167.

Brehmer, B. and R. Allard (1991). Dynamic decision making: The effects of task complexity and feedback delay. In *Distributed Decision Making. Cognitive Models of Cooperative Work*. J. Rasmussen, B. Brehmer and J. Lepleit. Chichester, UK, Wiley**:** 319-334.

CERT/CC (1998). *CSIRT development*.

Chandler, P. and J. Sweller (1991). *Cognitive Load Theory and the Format of Instruction*. Cognition & Instruction, Lawrence Erlbaum Associates. **8:** 293.

Cooper, J. O., T. E. Heron, et al. (1987). *Applied Behavior Analysis*. Columbus, OH: Merrill Publishing Company.

Crossman, E. and J. Cooke (1974). Manual control of slow-response systems. In *The Human Operator in Process Control*. E. Edwards and F. P. Lees. London, UK, Taylor and Francis.

Dörner, D. (1975). Psychologisches Experiment: Wie Menschen eine Welt verbessern wollten und sie dabei zerstörten. *Bild der Wissenschaft* **2**: 48-53.

Dörner, D. (1989 (1996)). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen (The Logic of Failure)*. Reinbeck: Rowohlt.

Galotti, K. M. (1999). *Cognitive Psychology In and Out of the Laboratory*. Pacific Grove, CA: Wadsworth Publishing Company.

Gonzalez, J. J. (2002). Modeling the Erosion of Safe Sex Practices. In Proceedings to the Twentieth International Conference of the System Dynamics Society, Palermo, Italy.

Gonzalez, J. J. and A. Sawicka (2003). The role of learning and risk perception in compliance. In *From Modeling to Managing Security: A System Dynamics Approach*. J. J. Gonzalez. Kristiansand, Norway, Norwegian Academic Press.

Gordon, H. S. (1954). The economic theory of a common property resource: The Fishery. *Journal of Political Economy* **62**.

[40] The research project *Disseminating Insights from Complex Models to a Broader Audience: Case of system dynamics models*, conducted by Agata Sawicka, is a postdoctoral fellowship funded by the Research Council of Norway, grant 160789/V30.

Hardin, G. (1968). The tragedy of the commons. *Science* **162**(13): 1243-1248.
Hardin, G. and J. Baden (1977). *Managing the Commons*. New York, NY: Freeman.
Hastie, R. and R. M. Dawes (2001). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making.*: Sage Publications.
Howie, E., S. Sy, et al. (2000). Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review* **16**: 151-171.
Jensen, E. (2005). Learning and transfer from a simple dynamic system. *Scandinavian Journal of Psychology* **46**: 119-131.
Jensen, E. and B. Brehmer (2003). Understanding and control of a simple dynamic system. *System Dynamics Review* **19**(2): 119-137.
Johnson-Laird, P. N. and P. C. Wason, Eds. (1977). *Thinking: Readings in Cognitvie Science*. Cambridge, UK, Cambridge University Press.
Kahneman, D., P. Slovic, et al. (1982). *Judgment under Uncertainty: Heuristics and Biases*. New York, NY: Cambridge University Press.
Killcrece, G., K. P. Kossakowski, et al. (2003). *State of the Practice of Computer Security Incident Response Teams (CSIRTs).*
Kneese, A. V. and J. L. Sweeney (2002). *XX*. Amsterdam, Netherlands: North-Holland.
Lipson, H. F. (2000). *Survivability - A new secuirty paradigm for protecting highly distributed mission-critical systems*. IFIP WG 10.4. Summer 2000 Meeting, 28 June - 2 July.
Melara, C., J. M. Sarriegi, et al. (2003). A system dynamics model of an insider attack to information systems. In Proceedings to the Proceedings of the 21st International Conference of the System Dynamics Society, 2003 New York, USA.
Moxnes, E. (1998). *Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics.* Management Science, INFORMS: Institute for Operations Research. **44:** 1234.
Moxnes, E. (1998). Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization* **37**(1): 107-127.
Moxnes, E. (2000). Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review* **16**(4): 325-348.
Moxnes, E. (2004). Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review* **20**(2): 139-162.
Moxnes, E. and A. K. Saysel (2004). Mispercpetions of global climate change: Information policies. In Proceedings to the 2004 International Conference of System Dynamics Society, Oxford, UK.
Paich, M. and J. D. Sterman (1993). *Boom, Bust, and Failures to Learn in Experimental Markets.* Management Science, INFORMS: Institute for Operations Research. **39:** 1439.
Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Hants, UK: Ashgate Publishing Ltd.
Repenning, N. P. and J. D. Sterman (2001). Nobody ever gets credit for fixing problems that never happened: Creating and sustaining process improvement. *California Management Review* **43**(4).
Sawicka, A. (2004). *Dynamcis of Security Compliance: Case of IT-based Work Environments*. Department of Information Science and Media Studies. Bergen, Norway, University of Bergen. **Ph.D.**
Schneier, B. (2000). *Secrets and Lies: Digital Security in a Networked World*. New York: John Wiley & Sons, Inc.

Sidman, M. (1960). *Tactics of Sceintific Research: Evaluating Experimental Data in Psychology*. New York: Basic Books.
Smith, D. (1994). Forming an incident response team. In Proceedings to the FIRST Annual Conference, Brisbane, Australia.
Sterman, J. D. (1987). Testing behavioral simulation models by direct experiment. *Management Science* **33**(2): 1572-1592.
Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* **43**(3): 301-335.
Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review* **10**: 291-330.
Sterman, J. D. (1997). Superstitious learning. *The Systems Thinker* **8**(5): 1-5.
Sterman, J. D. (2000). *Business Dynamics. Systems thinking and modeling for a complex world.*: McGraw-Hill.
van Wyk, K. R. and R. Forno (2001). *Incident Response*. Sebastopol, CA: O'Reilly & Associates.
Wack, J. P. (1991). *Establishing a Computer Security Incident Response Capability (CSIRC)*, National Institutes of Standards and Technology.
West-Brown, M. J., D. Stikvoort, et al. (2003). *Handbook for Computer Security Incident Response Teams (CSIRTs)*.