Paper presented at the 26th International Conference of the System Dynamics Society, July 20-24, 2008, Athens, Greece

Simulation-enhanced descriptions of dynamic problems: Initial experimental results

Agata Sawicka
University of Agder, Faculty of Engineering and Science
P.O. box 509, NO-4898 Grimstad, Norway
Tel. +47 37 25 33 58, Fax: +47 37 25 30 01
agata.sawicka@uia.no

Birgit Kopainsky
University of Bergen, Department of Geography
P.O. box 7800, NO-5020 Bergen, Norway
Tel. +47 47 62 79 86, Fax: +47 55 58 30 99
birgit.kopainsky@geog.uib.no

System dynamics models are built to assist people in understanding and solving complex and dynamic problems. However, the actual outreach of the models is limited as there is still no effective way identified to present dynamic problems and the associated models to a broader audience. Experimental data as well as cognitive load theory suggest that learning about and performance in complex dynamic systems could be improved by enhancing problem descriptions with simulation elements facilitating interactive exploration of dynamic features of the problem. We replicate experiments by Moxnes (2004) on management of reindeer lichen winter pastures, extending the task instructions by an interactive applet featuring dynamics of non-linear growth rate of lichen. In contrast to previous observations when the subjects misperceived gravely the system's dynamics at the outset, our results suggest that with the interactive applet the misperceptions of dynamics can be reduced already in the first trial.

Introduction

Both theoretical and experimental literature (Forrester 1961, Dörner 1989, Sterman 1994) indicate that people should use computer simulation models to manage dynamic systems successfully. With the introduction of interactive learning environments (ILEs) in the system dynamics (SD) field (Machuca, Ruiz et al. 1998, Spector and Davidsen 1997), the hopes have been raised high that such applications will facilitate dissemination and take up of models. Still, thus far evidence on effectiveness of ILEs is rather mixed (see e.g. Sawicka and Rydzak 2007 for a review, see also Großler and Maier 2004). The group model building remains the most effective way to disseminate the models and their insights (e.g. Andersen et al. 2007, Andersen and Richardson 1997; Luna-Reyes et al. 2006a; Rouwette 2003; Vennix 1996 and 1999; Zagonel 2002). However, given the limited outreach of this approach, there is an urgent need in trying to identify ways in which other dissemination channels may be made more effective.

Most of research on ILEs has focused on design of decision-making interface: The results indicate that ILE interfaces to effectively support learning should go beyond letting the learner input decisions and observe an outcome of the simulation. They need also to support exploration of model structure, and facilitate appreciation of its relationship with the generated behaviour (Spector and Davidsen 1998). Furthermore, ILEs should follow basic usability guidelines (Vicente 1996, Howie, Sy et al. 2000). Still, despite these preliminary principles, one still struggles with development of successful learning environments based on system dynamics models. Although many of the ILEs that provide direct access to the model structure and are developed according to the usability guidelines seem to yield an improved performance and learning, the overall results are mixed (see e.g. Bois 2002, Großler and Maier 2004, or Sawicka and Rydzak 2007 for reviews).

In this paper we report on a pilot experiment where we have modified the initial presentation of the problem rather than the decision making interface, by enhancing it with an interactive simulation applet featuring the key dynamics of the problem situation. In the following section we lay down a brief theoretical rationale for introducing such modification. Next, the experimental study design and results are presented. Its implications and future research directions are discussed in the final section.

The need for augmented descriptions of dynamic problems

Research on problem solving suggests that the main difference between experts and novice problem solvers is in their ability to identify an appropriate solution path: experts are able to classify accurately problems, choosing quickly the best solution strategy; novices, on the other hand engage in general search techniques such as trial-and-error, or means-ends analysis, taking more time to find a solution and being less successful (Chi et al. 1982, Larkin et al. 1980). A tendency to rely on the basic problem solving strategies, such as trial-and-error, or fall a prey various decision biases and simplified heuristics has been frequently observed in the context of dynamic problem solving (Dörner 1989, Sterman 1989, Paich and Sterman 1993, Diehl and Sterman 1995, Brehmer 1992, Moxnes 1998; Jensen and Brehmer 2003, Moxnes 2004), indicating that in the context of dynamic decision making most people do not have expert capacities.

The cognitive structures that allow experts to perform a prompt categorization of a problem, that normally would require several steps, are called *schemes* (Sweller 1988). Schemes, stored in a long-term memory, are developed through learning. Cognitive load theory (CLT, Chandler and Sweller 1991) provides a convenient theoretical framework for conceptualizing the learning process and designing instructional process so that development of schemes is supported in an efficient way. According to CLT learning process is supported by a short-term memory that has a limited capacity (Miller 1956). To utilize this scares resource in a most effective way, the learning material should be designed in such a way so that it induces primarily the so-called *germane cognitive load*, i.e. cognitive load directly associated with development of schemes. For this to occur, care should be taken to reduce as much as possible the so-called *extraneous cognitive load* (induced by structure and format of the learning material) and to adjust the so-called *intrinsic cognitive load* (caused by the inherent difficulty of the learning material) to the knowledge and abilities of the learners. Sawicka and Molkenthin (2005) develop a preliminary system dynamics model of the interplay between the various cognitive loads (for the most updated implementation of the model see Sawicka in press), pointing out how the cognitive load theory guidelines could guide development of system dynamics based learning environments. In a similar way, one could revisit structure and format of tasks used for testing people's ability to manage dynamic systems, probing the question of whether the observed misperceptions of dynamics are due to the inherent inability of people to understand dynamic problems as suggested by many of the studies in the area (e.g., Dörner 1989, Sterman 1987, Sterman 1989, Brehmer and Allard 1991, Paich and Sterman 1993, Diehl and Sterman 1995, Moxnes 1998, Jensen and Brehmer 2003, Moxnes 2004) or whether this misperceptions could be alleviated by a re-design or re-formulation of the task.

Several experiments suggest that both the re-design and re-formulation of the task or the decision making interface may lead to an improved performance and understanding (see e.g., Sengupta and Abdel-Hamid 1993, Howie, Sy et al. 2000, Bois 2002). Still, the results are not consistent, and similar changes made to other tasks fail to yield similar results (see e.g., Moxnes 1998, Jensen and Brehmer 2003, Sawicka and Rydzak 2007). What seems however consistent is the fact that subjects tend to improve their performance as they gain experience through the multiple trials (see e.g., Paich and Sterman 1993, Moxnes 1998; 2004, Jensen and Brehmer 2003, Sawicka, Gonzalez et al. 2005, Sawicka and Rydzak 2007). This suggests that despite the fact that initial instructions always provide all the necessary information about the system, for most subjects the hands-on trial is vital to develop a fuller understanding.

In the context of cognitive load theory, these results suggest that the instructions used for presenting the dynamic tasks fail to support germane processing effectively. This may be either due to insufficient cognitive resources, suggesting that the instructions impose excessive intrinsic and extraneous cognitive loads, or to the fact that the instructions do not stimulate sufficiently germane processing. Such processing seems to be elicited only during the first trials in a simulated environment, when most of the experimental subjects develop their understanding of the system. The main disadvantage of such experience-based learning, is that it frequently leads to simplified, erroneous mental models (Forrester 1961, Sterman 1994). Hence, it is essential that the learning process is facilitated and guided. Drawing on the cognitive load theory recommendations, one should than revisit the initial instructions and make sure that the intrinsic cognitive load they impose is appropriate to the level of prior knowledge that could be expected of learners, that the extraneous cognitive load is minimized and that the germane cognitive load is stimulated (Sweller and Chandler 1994).

Most of the dynamic task instructions consist of textual, written descriptions of the dynamic problems. Given the observation that people acquire understanding of the dynamics through the interaction with a simulator, we suspect that a textual format is not likely to stimulate germane processing for development of schemes about the dynamic issue at hand. Following this intuition, we conducted a pilot experiment to explore how performance in a dynamic task would change if the subjects were presented with instructions enhanced with a simulation-capacity.

Experimental design

For the purpose of our study we use a one-stock reindeer rangeland management task by Erling Moxnes. The task is to restore a highest sustainable reindeer herd size as quickly as possible on an overgrazed lichen pasture. The instructions provide a description of lichen growth dynamics, indicating that the growth rate is a non-linear, inverse U-shape function of lichen density, and a 15-year long historical record on lichen and reindeer herd size levels. Despite this full information, experiments yield a consistently poor user performance, especially in the first trial. This is illustrated in Figure 1 a) and b) where the results from both experiments by Moxnes 2004 are presented.

a) 2002 b) 2003

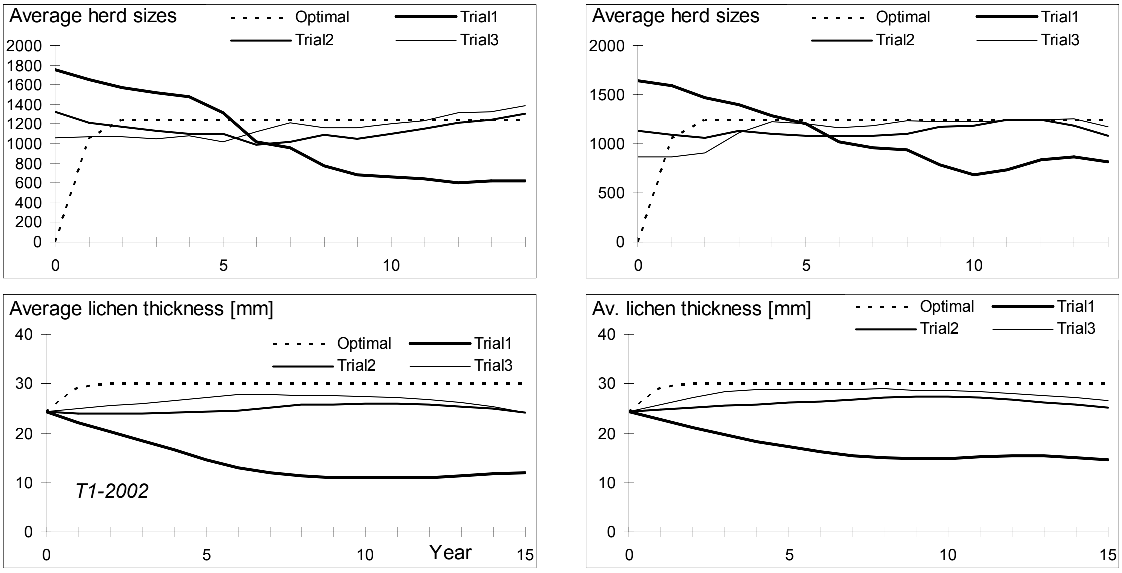


Figure 1: Average performance in experiments from 2002 and 2003 by Moxnes 2004.

The performance was not found to improve greatly when the learners were presented with an explicit illustration of the assumed growth curve (Moxnes 1998). It also remained literary the same when the task was adopted to a different context (Sawicka, Gonzalez et al. 2005, see also Figure 2).

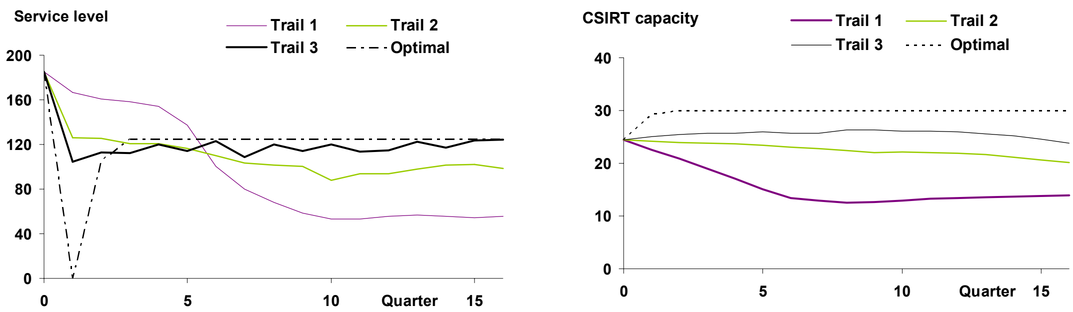


Figure 2: average CSIRT performance

The optimal solution, marked with light grey line in Figure 1 and Figure 2, requires the subjects first derive the lichen growth curve from the historical data, and identify correctly the maximum sustainable growth rate and density and derive the associated maximum sustainable herd size, see Figure 3. To achieve the maximum sustainable condition as quickly as possible in their decision making they need to aim at eliminating the discrepancy between the current and optimal lichen density. The 3 steps in which the maximum sustainable lichen density can be restored are presented in Table 1.

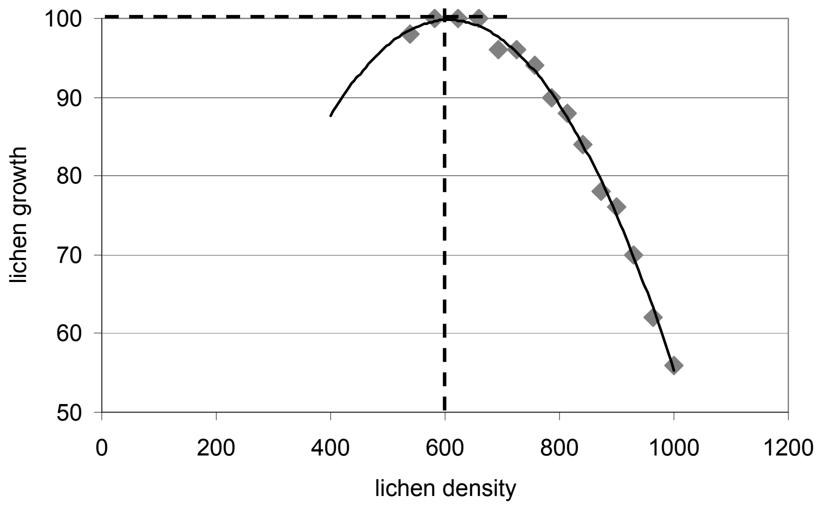


Figure 3: lichen growth curve and optimum conditions

Table 1: calculation of the optimal solution

Decision period:	I	II	III
D : Current lichen density [g/m^2]	488	584,5	600
ΔD : deviation of the current lichen density from the optimum density (D_{opt} =600 [g/m^2]) **ΔD = | D – D_{opt}** |[g/m^2]	112	15,5	0
Desired grazing rate (g) should equal g_{max} (g_{max} = 100 [g/m^2/year]).	0	84,5	100

Current desired grazing rate g [g/m^2/year]: **IF $\Delta D > g_{max}$ THEN g=0 [g/m^2/year]** **ELSE $g = g_{max} - \Delta D$**			
Current desired herd size R considering constant grazing per reindeer per year (gR) **R = g/gR** [reindeer]	0	1056	1250

As reported in Moxnes 2004 and Sawicka, Gonzalez et al. 2005 subjects rarely find the optimal solutions. Their performance is especially poor in the first trial: many increase rather than reduce the initial herd size; many deplete their lichen resource significantly. A common observation is also that the performance improved over the trials. This is consistent with the observations made during other experiments on dynamic decision making (see p. 3), and in the light of the cognitive load theory indicates that a simulation capacity is important for gaining the command of the dynamic system. Consequently, to achieve a satisfactory command of the system, simulation capacity seems to be vital for the phase when the subjects learn about the system. This observation was confirmed by the interview data gathered in Sawicka and Rydzak 2007 where the majority of subjects managing a production-inventory system indicated explicitly that it was the hands-on experience that helped them to understand the system's dynamics.

Building on these observations we believe that it is fundamental that descriptions of dynamic problems are enhanced with simulation capacity. To test whether such enhancement could lead to an improved performance we introduce an interactive explorator allowing the subjects to investigate the dynamics of lichen growth and grazing. Figure 4 presents both the original and the exploration-enhanced conditions.

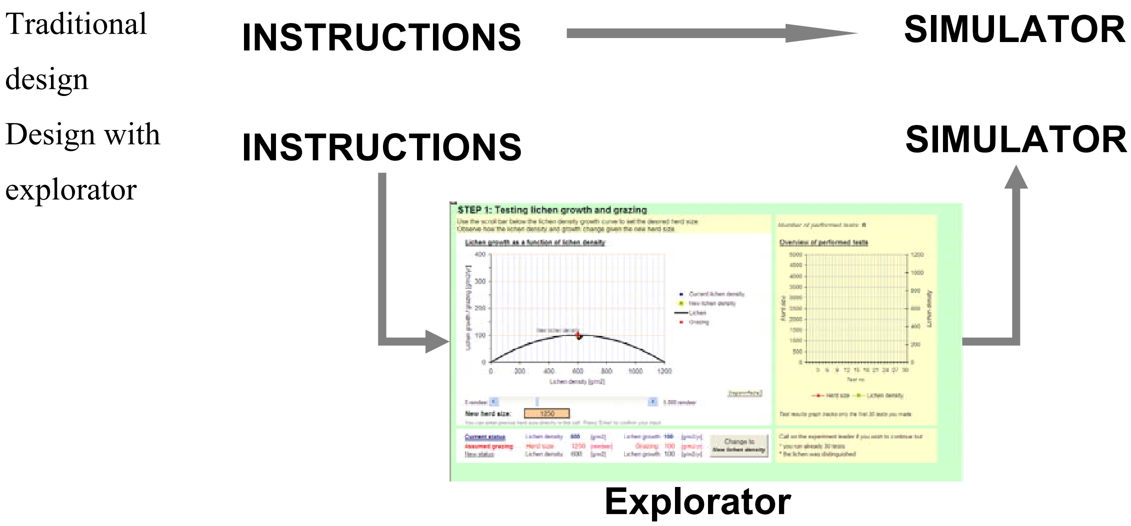


Figure 4: Experiment conditions with traditional design and exploration-enhanced condition

The interactive explorator allows subjects to test out their assumptions about strategies for solving the task described in the instructions. It also allows them to develop some understanding of lichen dynamics and of the problems caused by the nonlinear lichen growth curve. Subjects experience the effects of variations in the herd size on lichen density and lichen growth in situations where the pasture is over- and undergrazed.

As the exploration-enhanced condition differs only with respect to the explorator, and the original design has been shown to yield consistent results (Moxnes 2004, Sawicka, Gonzalez et al. 2005), we pursued a single sample design for this study and will compare the results with those collected earlier by Moxnes (2004, see also Figure 1). In the remainder of this paper we will refer to these results as the control group. With the exploration enhanced condition (in the remainder of the paper referred to as test group) we expect an improved performance already in the first trial.

Subjects and procedure

The pilot experiment reported in this paper was conducted with 8 students taking a course on applied methods in agricultural and regional policy at ETH Zurich in March 2008. The experimental session lasted 2 hours. 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 course grade.

They were also promised that the person who performed best would receive a symbolic prize. This incentive is analogous to the one used by Moxnes (2004, see p. 144).
The experimental session was divided into three stages: First, the subjects read the original task instructions as used in the control group study by Moxnes (2004), and answered a post-instructions questionnaire probing how much mental effort they invested in reading the task and eliciting their current understanding of the task. Next, the subjects explored the lichen growth and grazing dynamics with the explorator. The explorator was provided in three initial states: starting at the maximum sustainable point, starting in an under-grazed situation, and over-grazed situation, see Figure 5.

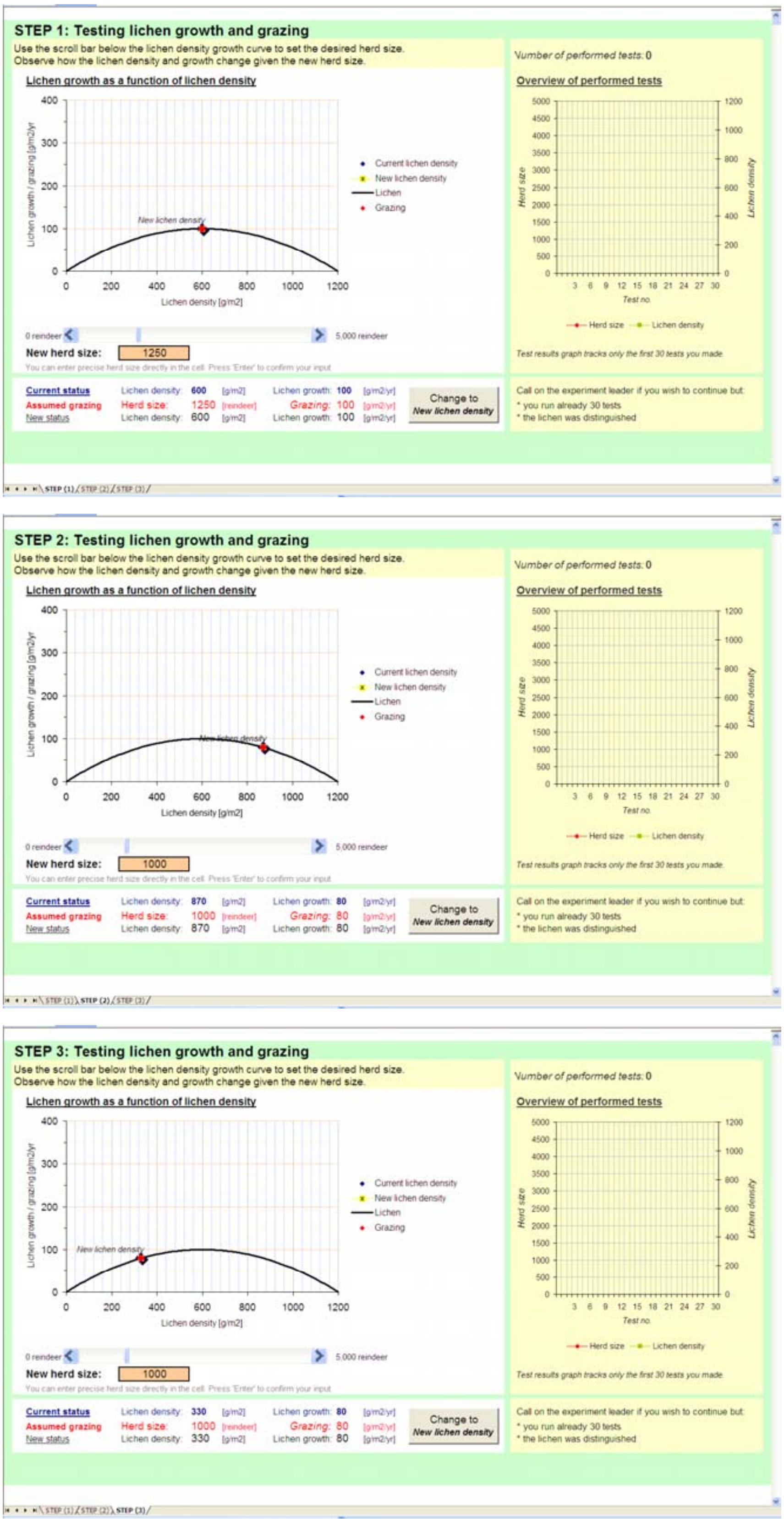


Figure 5: Explorator starting at optimal (step 1), under-grazed (step 2) and over-grazed situations (step 3)

After using the explorator, the subjects performed 15-years long trials using the original simulator by Moxnes (2004), see Figure 6. This was followed by a final questionnaire probing subjects' background as well as their intrinsic motivation in performing the task. All the subjects moved at self-paced speed through the stages. The experimental session was then followed by a short individual interview to clarify any issues outstanding after the initial data review, and the plenary debriefing session where the study background and the subjects' results were discussed.

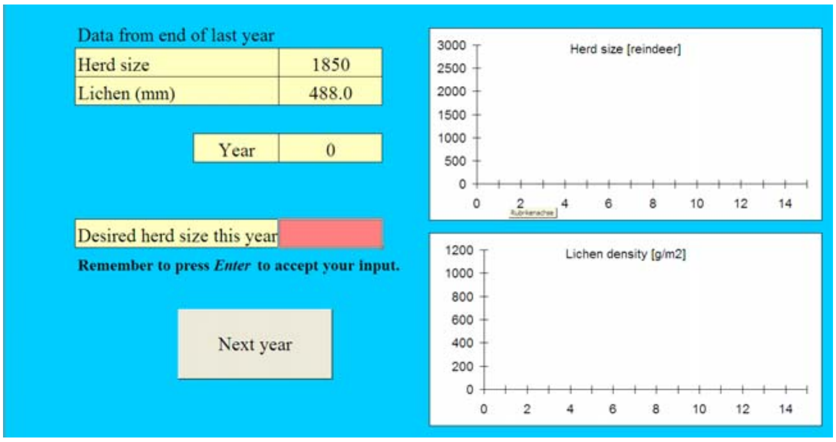


Figure 6: Original simulator

Preliminary results

Table 2 shows individual subject results per trial. All the subjects who decrease their herd size considerably in the first years and built it up again to the maximum sustainable level were classified as successful. As we can see already in the first trial 50% subjects performed the task very well, and by the third trial all the subjects were successful. All but one subject seemed to have understood the task correctly. The one subject misperceived the goal to be reindeer herd maximisation for the period of 15 years only. The subject therefore consistently opted for a herd size of nearly 0 animals towards the end of the trials in order to prepare for 5000 animals in the final year 15.

Table 2 Overview of individual subjects results. Performance of successful subjects is tracked with thicker lines.

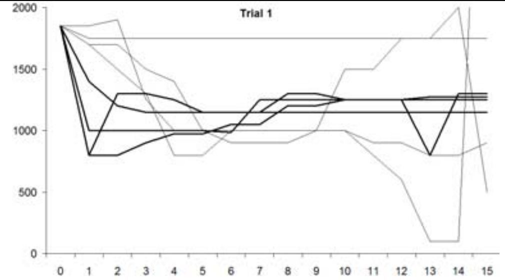
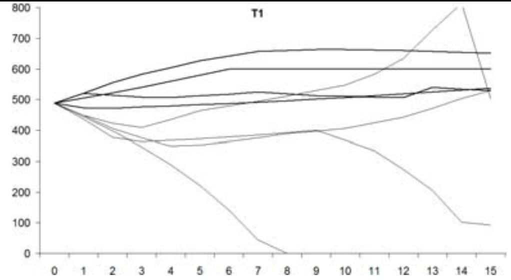
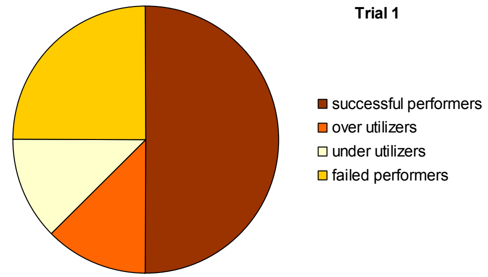
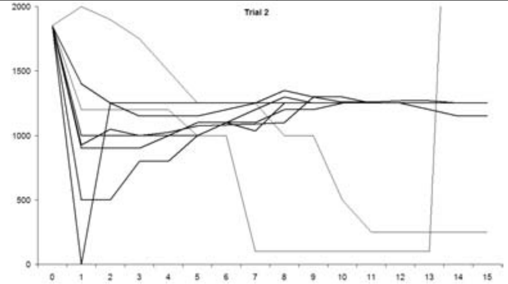
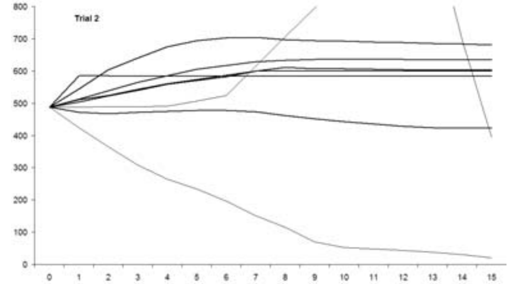
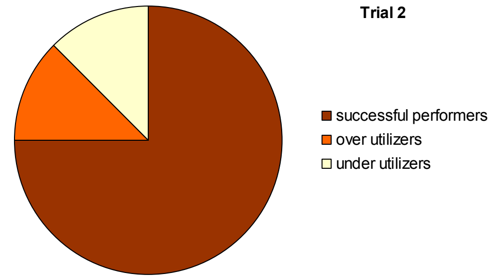
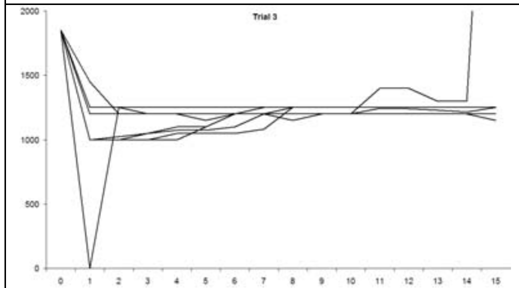
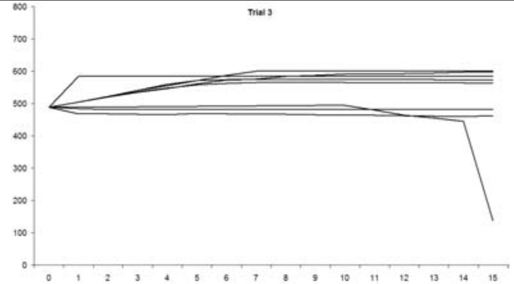
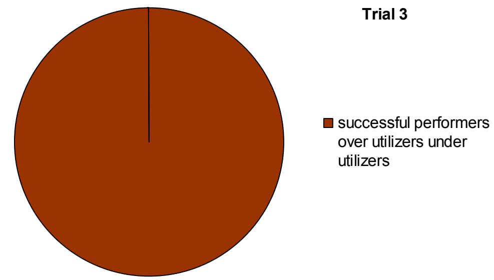
Reindeer herd size	Lichen	Subjects’ classification
		
		
		

Figure 7 shows 95 percent confidence intervals for average herd sizes and lichen density for trial 1 in the test and control groups. The grey, dashed line shows the optimal paths. The widening range of the confidence intervals towards the end of the trials is caused by the subject who misperceived the task and maximized the herd size at year 15, see also Table 2 and the associated discussion. For herd sizes in the test group the optimal solution lies within the confidence interval starting from year 2. In the control group, on the other hand, the average herd size is significantly higher than the optimal in the early years, and in late years, it is significantly lower. Average lichen density is below the optimal level in all years after the initial one in the control group, but reaches the optimum level in year 12 in the control group. This indicates that the subjects in the control group are, on average, not successful in reaching the maximum sustainable herd size and lichen density within 15 years, while the opposite holds true for the test group, i.e. an average subject in the exploration enhanced condition reaches the maximum sustainable herd size within 15 years, and is also more likely to arrive at the maximum sustainable lichen density level.

a) exploration enhanced condition – test group

b) original experiment design – control group

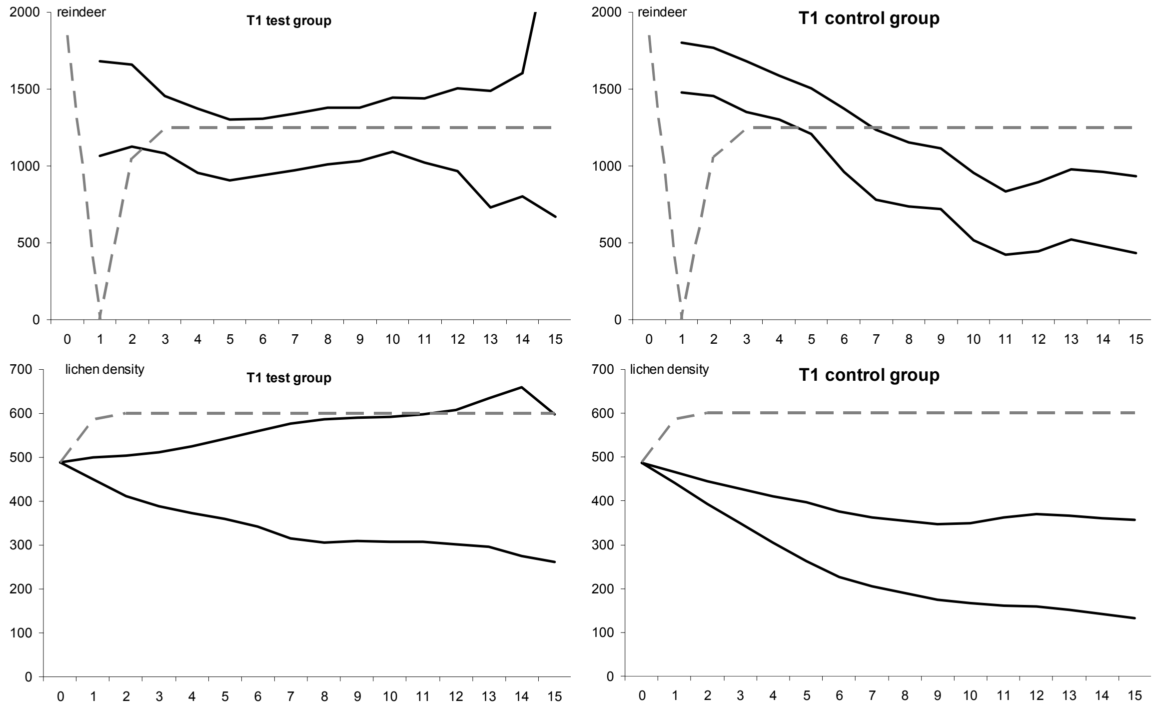


Figure 7: Optimal solution (dashed lines) and 95 percent confidence intervals for the average first trial in the exploration enhanced condition (test group) and the traditional experiment design (control group)

To see if the differences between the test group and the control group are statistically significant at the 10 percent level, we compare lichen thicknesses relative to optimal lichen thicknesses for the two groups. Figure 8 shows confidence intervals for trial 1 in the test group (black, thick lines) and the control group (black, thin lines). Comparing the results year by year we find the p-values shown by the grey, dashed line. The p-value indicates the probability of committing a Type I error (Gujarati 1995). In other words, with higher p-value the more the probability that there is no difference between the results of the test group and the control group. At a ten percent level, the two datasets are, according to Figure 8, significantly different starting from year four.

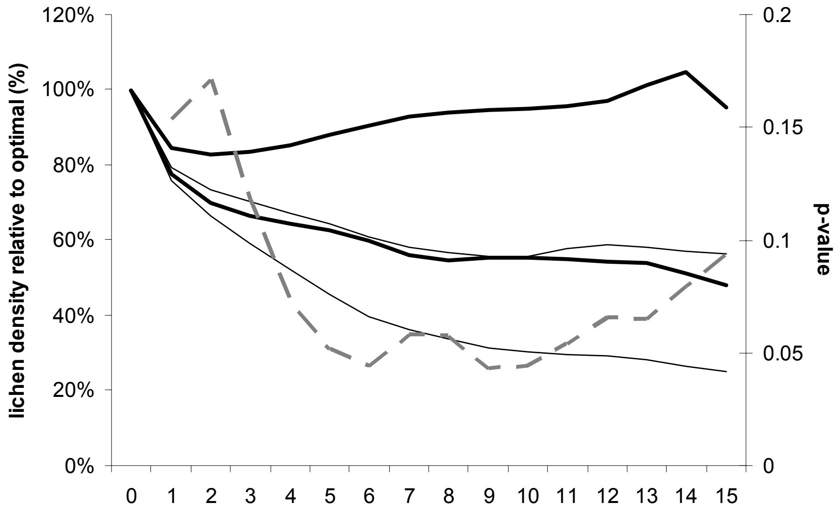


Figure 8: 90 percent confidence intervals for average lichen thicknesses for T1 test group (thick lines) and T1 control group (thin lines)

Discussion

Results of this pilot study indicate that the exploratory-enhanced instructions may facilitate understanding of the reindeer rangeland management task. As illustrated in Figure 7 and Figure 8, there is a significant difference between the way our subjects dealt with the problem in the first trial. Their performance is comparable to the performance typically observed in the second or third trials in previous experiments. This may suggest that the explorator helped the subjects to acquire the understanding of the system that they otherwise had to develop during the initial trials with the test simulator.

This initial result needs however to be taken with caution: First, most of our subjects, as subject majoring in agricultural economics, did have a substantial background in renewable resource management. Indeed, many of them commented that the experience from their coursework was quite useful in tackling the experimental task. This prior-knowledge could have contributed to the outstanding performance observed already in trial 1. On the other hand, previous research by Moxnes indicates that even the subjects with substantial professional experience tend to fail at first trials (Moxnes 1998). Second, the performance data need to be analyzed in the context of qualitative data gathered through the questionnaires administered during the experimental session as well as during the interviews. In these further analyzes we will focus on identifying all factors beyond the explorator that could have contributed to the superior first trial performance. We will also conduct the analyzes of the mental effort measurements gathered throughout the experimental session to see whether these data support our initial intuition that the instructions impose a substantial cognitive burden and the explorator may assist in stimulating the learning process, leading to an improved understanding of the task.

Future directions

The initial analyses of the results of our pilot study indicate that people may be able to better grasp dynamic problems if their descriptions are augmented by simulation-based explorators featuring critical dynamic elements. Wary of the need of further analysis of the current data and replication with a larger, possibly more representative, sample, we still believe that these early results are encouraging. In our future studies we intend to modify further the task instructions so that the explorator becomes an integral part of the problem presentation. It would also be interesting to replicate the results in a different context, for example using the adapted version of the task to the context of management of computer information security response teams presented in Sawicka, Gonzalez et al. 2005.

References

Andersen DF, Vennix JAM, Richardson GP, Rouwette EAJA (2007). "Group model building: problem structuring, policy simulation and decision support." Journal of the Operational Research Society **58**(5): 691-694.

Andersen DF, Richardson GP 1997. "Scripts for group model building." System Dynamics Review **13**(2): 107-129.

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.**
Brehmer, B. (1992). "Dynamic decision making: Human control of complex systems. ." Acta Psychologica **81**: 211-241.
Brehmer, B. and R. Allard (1991). Dynamic decisions making: The effects of task complesity and feedback delay. Distributed Decision Making: Cognitive Models for Cooperative Work. J. Rasmussen, J. Laplat and B. Brehmer, John Wiley & Sons. **45:** 223-241.
Chandler, P. and J. Sweller (1991). Cognitive Load Theory and the Format of Instruction. Cognition & Instruction, Lawrence Erlbaum Associates. **8:** 293.
Chi, M. T. H., Glaser, R. & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), Advances in the Psychology of Human Intelligence Vol. 1: 7-76. Hillsdale, NJ: Erlbaum.
Diehl, E. and J. D. Sterman (1995). "Effects of Feedback Complexity on Dynamic Decision Making." Organizational Behavior and Human Decision Processes **62**(2): 198-215.
Dörner, D. (1989). Die Logik des Mißlingens. Reinbek, Rowohlt.
Dörner, D. (1989). Die Logik des Misslingens. Strategisches Denken in komplexen Situationen. Reinbeck, Rowohlt.
Forrester, J. W. (1961). Industrial Dynamics. Cambridge, MA, MIT Press.
Großler, A. and F. H. Maier (2004). "Don't let history repeat itself: Methodological issues concerning the use of simulators in teaching and experimentation " System Dynamics Review **20**(3): 263-274.
Gujarati D.N. (1995). Basic Econometrics. McGraw-Hill, New York etc. 3rd edition
Howie, E., S. Sy, et al. (2000). "Human-computer interface design can reduce misperceptions of feedback." System Dynamics Review **16**: 151-171.
Jensen, E. and B. Brehmer (2003). "Understanding and control of a simple dynamic system." System Dynamics Review **19**(2): 119-137.
Larkin J.H., McDermott J., Simon D.P., Simon H.A. (1980). “Models of competence in solving physics problems.” Cognitive Science 4(4): 317-345
Luna-Reyes LF, Martinez-Moyano IJ, Pardo TA, Cresswell AM, Andersen DF, Richardson GP (2006). “Anatomy of a group model-building intervention: building dynamic theory from case study research.” System Dynamics Review **22**(4): 291-320.
Machuca, J. A. D., J. C. Ruiz, et al. (1998). Our 10 years of work on transparent-box business simulation. 16th International Conference of the System Dynamics Society, Quebec, Canada.
Miller, G. A. (1956). "The magical number seven, plus or minus two: some limits on our capacity for processing information." Psychology Review **63**: 81-97.
Moxnes, E. (1998). "Overexploitation of renewable resources: The role of misperceptions." Journal of Economic Behavior & Organization **37**(1): 107-127.
Moxnes, E. (2004). "Misperceptions of basic dynamics: the case of renewable resource management." System Dynamics Review **20**(2): 139-162.
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.

Rouwette EAJA (2003). Group Model Building as Mutual Persuasion. Wolf Legal Publishers: Nijmegen.
Sawicka, A. (in press). "Dynamics of Cognitive Load Theory: A Model-Based Approach." Computers in Human Behavior.
Sawicka, A., J. J. Gonzalez, et al. (2005). Managing CSIRT Capacity as a Renewable Resource Management Challenge: An Experimental Study. 23rd International Conference of the System Dynamics Society, Boston, GA, USA.
Sawicka, A. and R. Molkenthin (2005). Cognitive Load Dynamics: How to Increase Effectiveness of SD-based Learning Environments. Twenty Third International Conference of the System Dynamics Society, Boston, MA, USA.
Sawicka, A. and F. Rydzak (2007). Incorporating delays in the decision-making interface: An experimental study. 25th International Conference of the System Dynamics Society, Boston, Ma.
Sengupta, K. and T. K. Abdel-Hamid (1993). Alternative Conceptions of Feedback in Dynamic Decision Environments: An Experimental Investigation. Management Science, INFORMS: Institute for Operations Research. **39:** 411.
Spector, J. M. and P. I. Davidsen (1997). "Creating engaging courseware using system dynamics." Computers in Human Behavior **13**(2): 127-155.
Spector, J. M. and P. I. Davidsen (1998). "Constructing learning environments using system dynamics." Journal of Courseware Engineering **1**: 5-12.
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. (1989). MODELING MANAGERIAL BEHAVIOR: MISPERCEPTIONS OF FEEDBACK IN A DYNAMIC DECISION MAKING EXPERIMENT. Management Science, INFORMS: Institute for Operations Research. **35:** 321.
Sterman, J. D. (1994). "Learning in and about complex systems." System Dynamics Review **10**(2-3): 291-330.
Sweller, J. (1988). "Cognitive load during problem solving: Effects on learning." Cognitive Science **12**: 257-285.
Sweller, J. and P. Chandler (1994). "Why some material is difficult to learn." Cognition and Instruction **12**: 185-233.
Vennix JAM (1999). "Group model-building: Tackling messy problems." System Dynamics Review **15**(4): 379-401.
Vennix JAM (1996). Group Model Building: Facilitating Team Learning Using System Dynamics. Wiley: Chichester.
Vicente, K. J. (1996). "Improving dynamic decision making in complex systems through ecological interface design: a research overview." System Dynamics Review **12**: 251-279.
Zagonel AA (2002). Model conceptualiziation in group model building: A review of the literature exploring the tension between representing reality and negotiating a social order. Proceedings of the 20th International System Dynamics Conference.