Paper presented at the 27th International Conference of the System Dynamics Society

July 26-30, 2009

HOW CAN WE ASSESS WHETHER OUR SIMULATION MODELS IMPROVE THE SYSTEM UNDERSTANDING FOR THE ONES INTERACTING WITH THEM?

Mihaela Tabacaru[1], Birgit Kopainsky[1], Agata Sawicka[2],
Krystyna A. Stave[3], Heather Skaza[3]

[1] System Dynamics Group, Department of Geography, University of Bergen, Postbox 7800, 5020 Bergen, Norway

[2] Norwegian Post and Telecommunications Authority, Post Box 93, 4791 Lillesand, Norway

[3] University of Nevada Las Vegas, 4505 Maryland Parkway, Box 454030, Las Vegas NV 89154-4030 USA

Abstract

Most of the system dynamics studies that evaluate decision making in complex dynamic task focus on the evaluation of performance over repeated trials and on the effectiveness of different instructional strategies as far as performance is concerned. Especially when a strategy seems to yield promising results in terms of performance, it becomes essential to know whether improved performance is due to improved system understanding, i.e. to correct rules or due to other reasons such as trial and error. This paper contributes to the emerging literature in system dynamics about assessing system understanding. Based on the way experts make decisions we develop a step by step guide to evaluate how the understanding of the system develops in the course of subjects interacting with the system through a simulation model. We apply our guide to the reindeer management task and analyze data from previous experiments with the task. This application provides important insights for the further development of the questionnaires that are applied for assessing understanding.

1 Introduction

Dynamic systems such as the economy of a country, production in companies, renewable resources or global warming are difficult to understand and manage successfully. One of the primary goals of system dynamics is to improve decision making in complex dynamic systems. This raises the question of how the effectiveness of any system dynamics intervention can be assessed. Evaluations can basically focus on two main issues:

- Performance, i.e. the results from decision making.
- Understanding, i.e. the rules that lead to decisions.

The majority of the evaluations focus on the first issue and analyze performance in a dynamic task (e.g. Cronin & Gonzales 2007; Cronin et al 2008; Moxnes & Saysel 2009; Sterman & Booth Sweeny 2007). Especially when a strategy seems to yield promising results in terms of performance, it becomes essential to know whether improved performance is due to improved system understanding, i.e. to correct rules or due to other reasons such as trial and error.

This paper contributes to the emerging literature in system dynamics about assessing system understanding (Cavaleri & Sterman 1997, Doyle et al 2008; Jensen & Brehmer 2003, Jensen 2005, Huz et al 1997, Spector et al 2001). Our starting point is the question how we can assess whether interaction with a simulation model improves the understanding of the underlying system. Knowledge per se rarely triggers behavior in everyday life. Knowing about something does not make one behave in a certain way (Ajzen 2002). Knowing how to behave, i.e. knowing how to use the available information is more effective (Böhm & Pfister 2001).

We take this distinction between facts and the processing of facts as a starting point to develop a step by step guide for evaluating understanding of a complex dynamic system. We base the guide on the theoretical foundations of the expert decision making processes (Klein 1997). This theory describes how experts analyze a new situation and derive decision rules. Eliciting the way experts make decisions, i.e. how they use the information available to them is useful in two ways. On the one hand it provides insights into the rules that lead to successful performance. On the other hand it establishes a benchmark for assessing how well people interacting with a simulation understand the underlying system.

We illustrate the theoretical framework and the step by step guide for evaluating understanding of a complex dynamic system with a well tested dynamic decision making task, the reindeer rangeland management task (Moxnes 2004). We have collected preliminary data about subject understanding of the system underlying the task. We analyze these data on the background of our theoretical framework. This will yield a refined and more complete set of questions to ask in further reproductions of the task.

2 Reindeer management task

Throughout the paper we illustrate our approach with the reindeer rangeland management task developed by Moxnes (2004). For this task Sawicka & Kopainsky (2008) have found significant differences in performance when people try to solve the task in its original version (Moxnes 2004) and with simulation-enhanced problem descriptions. Further analyses of the effectiveness of simulation-enhanced problem descriptions need to assess whether such descriptions improve understanding or whether they just provide an additional trial during which the successful strategy can be discovered by chance.

In the reindeer management task, subjects play the role of sole owners of a reindeer herd. They take over the herd and overgrazed rangeland from a previous owner. In the experiment they are responsible for setting the reindeer herd size for each of 15 simulated years. Their goal is to restore the maximum sustainable herd size as quickly as possible. The instructions provide information about the grazing rate of the reindeer and a description of lichen growth dynamics, indicating that the growth rate is a non-linear, inverse U-shape function of lichen density. The instructions also contain a 15-year long historical record on lichen density and reindeer herd size levels.

After reading the instructions subjects proceed to a decision-making interface where they implement their strategies for solving the task. They perform three trials. During each trial they set the herd size for each of the 15 simulated years. The graphs included in the simulator's interface trace their decisions, as well as the development of lichen density.

An adapted version of the instructions to the task (the version used for the data presented in section 5) is reproduced in the appendix.

3 Understanding: theoretical framework

We view learning as becoming expert-like (Ericsson & Smith 1991; Spector 2006). 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 accurately classify problems and quickly choose an appropriate solution strategy; novices, on the other hand, engage in general search techniques such as trial-and-error, or means-ends analysis, taking not only more time to find a solution, but frequently being less successful (Chi et al 1982; Larkin et al 1980). It is also theorized that experts outperform novices thanks to more advanced internal knowledge representations stored in a long-term memory, that allow an expert to categorize problems more precisely and identify their solutions more promptly, without having to go through all the detail solution steps that would otherwise be required of a novice (Sweller 1988; Seel 2003).

3.1 Experts' decision making process

Decision making in complex dynamic systems is rather a cycle than an event: think a little, act a little, then evaluate outcomes and think and act a little more. It seems we learn to make better decisions by noticing the changes in an environment, storing examples of each situation experienced, and predicting future situations based on past experience.

Experts categorize problems more precisely, managing to find the important aspects that are relevant to a situation. It is as if they could "see" better. Klein's (1997) recognition primed decision making model posits that experts do not chose among alternatives, but rather assess the nature of the situation and, based on this assessment, select an action appropriate to it.

The first step in Klein's recognition model is to classify the situation as typical or novel. To recognize the situation, the decision maker identifies critical cues that mark the type of situation and causal factors that explain what is happening and what is going to happen. Based on this, the expert sets plausible goals and proceeds to selecting an appropriate course of action. Recognition has four aspects (not necessarily in this order) on which we will base our step by step guide for assessing understanding:

1. identification of cues – relevant aspects of the task;
2. formulation of expectations, frequently in ”IF –THEN” type of sentences;
3. definition of goals and
4. design of possible actions/ decision.

3.2 Existing experiences with and approaches to assessing understanding

Not all thinking is conscious or reportable, and thus directly accessible to verbalization. Berry & Broadbent (1984) found, across their whole body of experiments, a significant negative correlation between the ability to perform well and the ability to answer questions about the situation. Actually, for a non-salient task like the reindeer task, learning will be “better through experience, badly tested by questioning, and transfer only weakly to new situations”(Broadbent 1990: 52). In a critique of the beer game, Martin et al (2004) suggested that participants in previous experiments performed poorly simply because they did not have enough practice with the system, giving them little opportunity to learn. Proficient dynamic decision making typically requires *extended practice* with a system before mastering it. Interaction with the system is needed in order to improve understanding. We take this into account when we argue for several trials: repeating measures are more likely to capture the process of learning than just a questionnaire after interacting with a simulation model.

Asking about strategy, questionnaires or multiple-choice questions for finer grain differences are not likely to tell the whole story. Asking about the intended strategy reveals only little about a subject’s understanding because it requires the subject to imagine a plan even if they have none. Multiple choice questions also reveal only a part of a subject’s understanding because of the things that can be omitted by the experimenter, and still be a solution for the subject.

Capturing the understanding of experts in dynamic tasks has received a lot of attention lately, and methodologies were developed for this purpose. DEEP (Spector 2006) or Cognizer (Clarkson & Hodgkinson 2005) are just two of the software solutions that use concept maps drawn and annotated by subjects in order to elicit and then graph how experts think. They are indeed very useful for comparing the degrees of change of the models between instructions, and to compare novice and expert maps.

A slightly different type of software is MITOCAR (Pirnay-Dummer 2006), a software tool that is based on mental model theory (Seel 1991) and that uses natural language expressions as input data for model re-representation, instead of graphical drawings by the subjects. Due to the modular design of MITOCAR the assessment tools (re-representation of models by means of natural language, parsing and graph theory) can be separated from the inferential tools (comparing structures and semantics and both). This opens the MITOCAR technology to use on all kinds of model related data, and answers a direction for future research stated by Luna-Reyes & Andersen (2003), namely:

> “… the conversion of mental data to textual data. Experts have rich stories to tell. These stories are in the form of mental models that exist nowhere on paper, and that, in fact, might never have been verbalized by even the expert himself. Although social scientists and oral historians concern themselves with eliciting stories from their respondents, system dynamicists might very well need to create their own methods for extracting critical dynamic data from the stories that others tell.”

Among the methods used by social scientists is the think-aloud method, a technique that asks subjects to speak out loud as they solve a task. If properly applied, it provides a rich body of expert knowledge text to be interpreted. One limitation is that it reveals only the tactical steps that subjects employ, and not the more general model behind those steps (Goodwin & Johnson-Laird 2006).

4 Step by step guide for the evaluation of understanding

The existing approaches to assessing understanding argue for more intuitive and exploratory techniques. Our guideline sets the stage for tools like MITOCAR by providing the framework and discussing the kind of information that should be considered when evaluating understanding. It includes suggestions to elicit not readily-“verbalizable” knowledge, essential in eliciting the decision making process and in evaluating understanding.

We will illustrate the proposed step by step guide with examples from the reindeer management task.

4.1 Elicit expert understanding

Expertise is domain specific: experts do not use general problem solving techniques when dealing with content customary to their work. In order to have a benchmark, the first step is to constitute a panel of experts to go through all the steps described in section 4.2 to 4.4. The result should be a repertoire of:

- Cues and non-salient factors relevant for the task
- Goals
- Expectations/causal relations
- Actions/decisions

Eliciting expert understanding is closely related to the idea analysis applied in Jensen & Sawicka (2006) and Booth Sweeney & Sterman (2000), to the task analysis step applied e.g. in Jensen & Brehmer (2003) and Jensen (2005), and to eliciting expert conceptualizations of the problem space (Spector 2006).

4.2 Elicit understanding before the dynamic task

After reading the instructions/the problem description prompt for the four aspects of recognition (cues and non-salient factors, goals, expectations/causal relations, and actions/decisions).

Identification of cues and non-salient factors

- Ask subjects for analogies: e.g. what does this description remind you of?
- Ask subjects to indicate the relevant aspects of the task: variables and activities
- Ask subjects to indicate things *not* stated in the problem scenario that may be relevant to a solution also suggested by Spector (2006). The capability to identify non-salient (i.e. not readily available, non-transparent) factors is one characteristic that makes expert learners differ from novices (Berry & Broadbent 1987).

Definition of goals

- Ask for specific goals

Expectations

- Ask for if-then expectations, i.e. the expected consequences of a change in one variable

4.3 Elicit understanding during the dynamic task (after each trial)

Analyze actions/decisions

- Look at performance to extract the rules used by the subject and confront the subjects with their own decisions (e.g. a drastic decrease in the number of reindeers)
- Ask for causal relations after each try: why do you think this happened? If a relation is identified, prompt for the sign of the relation
- If a relation is identified, prompt for the sign of the relation

Issues to be taken into account for eliciting understanding before and during the task

Subjects are unlikely to take the correct steps towards solving a complex dynamic task after reading the problem description (instructions), simply because they will generally not associate information about e.g. the dependence of current output on the previous level of output (self-generated flow, essential to solving this problem) to such a task. At the same time, many subjects will verbalize strategies they use in order to *control the system*, and not the assumptions behind each action, as pointed out by Broadbent (1990). For example, some of the strategies relevant to control and succeed in the reindeer task would be rules similar to those identified by Fum & Stocco (2003) in a dynamic task:

- Choose randomly a value between x and y to reduce the number of reindeers
- Repeat-Choice
- Stay-on-Hit when the previous choice resulted in a success
- Pivot-Around-Target
- Jump-Up/Down etc.

Being aware of this, always prompt for the “THEN”, i.e., for what the subject expected to achieve by implementing their decision.

4.4 Elicit understanding after the dynamic task

After the task has been completed subjects should be encouraged to mentally reconstruct the system they had been interacting with. Counterfactual thinking (which is very similar to the if-then sentences used before the task) is an effective strategy for this.

Counterfactual thinking concentrates on what could have been different if some details from a past event had been changed. They are a natural response to negative events. Counterfactuals generated following a failure serve the purpose of mentally challenging the causes of the failure and preparing the subject for the next time. The behaviors and intentions thus generated are related to improvement of performance. The functional aspect of counterfactuals is underlined by Epstude & Roese (2008).

Use of counterfactual thinking to prompt for all four aspects of recognition

- Ask subjects to complete the following sentence: "If I had……..then it would have been better."
- Encourage for more than one sentence

Other approaches to prompt for all four aspects of recognition

- Ask to explain to somebody else how to control the system (Stanley et al 1989) *;* this requires that the subjects construct more than a sequence of steps and explain the "why" behind each step

4.5 Examples from the reindeer management task

In this section we illustrate the above sketched step by step guide with the expert decision making process in the reindeer management task.

Examples of cues and non-salient factors

Cues directly extractable from the instructions:

- Reindeer/herd size
- Lichen is a renewable natural resource
- Lichen/lichen density
- Lichen growth rate
- Reindeer eat lichen
- Lichen growth depends on lichen density

Non-salient factors (things not stated in the problem description that are relevant to the solution):

- Lichen density cannot be influenced directly. If I want to adjust lichen density, I have to adjust the number of reindeer
- The number of reindeer should depend on lichen density

Examples of goals

- To achieve, as soon as possible, the maximum sustainable herd size, i.e. the herd size that allows for the highest possible growth of lichen

Examples of if- then expectancies and causal relations

- If I want to maintain/stabilize a given lichen density then grazing rate needs to be equal to the lichen growth rate
- If my pasture is overgrazed I have to reduce the number of reindeer
- If grazing is below lichen growth I can increase the number of reindeer
- Maximum sustainable grazing equals the maximum lichen growth rate
- The higher lichen density, the higher lichen growth. At some point in time, though, this relationship is reversed and becomes: the higher lichen density, the lower lichen growth

Examples of actions/decisions

Expert decisions

- When grazing exceeds lichen growth: Reduce number of reindeer
- Let lichen density recover by considerably reducing the number of reindeer
- When grazing is lower than lichen growth: Increase number of reindeer

Optimal solution of the task

- Reduce number of reindeer to 0 in the first year
- Increase number of reindeer in the second year
- Reach maximum sustainable herd size of 1250 reindeer in the third year

5 How much can we say about understanding in the reindeer management task so far?

The reindeer management task is a well established task that has yielded consistent results in terms of performance with participants with varying background (ranging from university students to participants with substantial professional experience) and varying forms of task illustrations (implicit description vs. explicit illustration of the nonlinear lichen growth curve) (Moxnes 1998, Moxnes 2004). Performance also remained literally the same when the task was adapted to a different context (Sawicka et al 2005).

In more recent replications of the task (Sawicka & Kopainsky under revision) we also introduced a number of questionnaires to gain a better understanding of subjects' profiles and their experiences with the task:

- Immediately after reading the instructions, the subjects worked through the post-instructions questionnaire where they reflected on their understanding of the task and on their intended strategies to solve the task. Questions about understanding of the task and about the intended strategies were open questions.

- After completing all the trials with the simulator, the subjects filled out the final questionnaire, which elicited basic demographics as well as subjects' general interest in and knowledge of natural resources management, as well as their interest in and experience of the experimental task. In addition, subjects had once again to reflect on their understanding of the task and their intended strategy if they were to have an additional trial. These two questions were identical to the questions asked immediately after the instructions (i.e. before interaction with simulation based activities had taken place) and thus open questions.

5.1 Evaluation of understanding with data from previous reindeer management experiments

In this section we report on data collected with students of an advanced environmental science class in Environmental Studies at the University of Nevada in Las Vegas. The data were collected in the fall semester 2008 for a total of 16 students. The main purpose of this pilot application was to gather experience with data that contains information about subjects' understanding of the system they interact with through simulations. The pilot should also help developing a set of more defined questions that need to be covered in the questionnaires of future replications of the reindeer management task.

The questions asked in the Las Vegas pilot did not cover the full range of questions that need to be asked according to our step by step guide for evaluating understanding of a complex dynamic system. The aspects of recognition that were not prompted for are shaded grey in Table 1 to Table 2.

The tables contain all the elements listed in section 4.5, where we provided examples from the reindeer management task to illustrate our step by step guide for evaluating understanding in complex dynamic systems. The open questions in the questionnaires were double coded. After each aspect of recognition, the tables summarize the implications for understanding (rows labeled e.g. "pre-test evaluation I").

Table 1: Pre-test evaluation of understanding

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
cues and non-salient factors	reindeer/herd size	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen/lichen density	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen is a renewable resource					x											
	lichen growth rate	x					x	x		x	x	x	x	x	x	x	x
	reindeer eat lichen	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen growth depends on lichen density				x	x					x	x					x
pre-test evaluation I	*number of identified cues*	*4*	*3*	*3*	*4*	*5*	*4*	*4*	*3*	*4*	*5*	*5*	*4*	*4*	*4*	*4*	*5*
	I can only adjust lichen density by adjusting the number of reindeer	x		x	x	x	x	x	x	x	x	x	x		x	x	x
	the number of reindeer should depend on lichen density		x	x	x			x	x				x	x		x	
pre-test evaluation II	*number of identified non-salient factors*	*1*	*1*	*2*	*2*	*1*	*1*	*2*	*2*	*1*	*1*	*1*	*2*	*1*	*1*	*2*	*1*
goals	achieve an equilibrium between the reindeer herd size and the lichen so that the lichen density does not decline any more	x	x		x			x	x			x			x	x	x
	increase herd size																
	reduce the herd size so that the lichen density could increase																
	maintain a large herd			x													
	not sure/other					x											
	achieve, as soon as possible, the herd size that allows for the highest growth of lichen						x			x	x		x	x			
pre-test evaluation III	*identified correct goal*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*1*	*1*	*0*	*1*	*1*	*0*	*0*	*0*
causal relations	If I want to maintain/stabilize a given lichen density then grazing rate needs to be equal to the lichen growth rate		x		x	x	x	x	x	x	x		x	x		x	x
	If my pasture is overgrazed I have to reduce the number of reindeer			x						x		x	x			x	x
	If grazing is below lichen growth I can increase the number of reindeer			x								x	x		x		
	Maximum sustainable grazing equals the maximum lichen growth rate					x			x		x	x		x			x
	The higher lichen density, the higher lichen growth. At some point in time, though, this relationship is reversed and becomes: the higher lichen density, the lower lichen growth							x									
pre-test	*number of identified causal relations*	*0*	*1*	*2*	*1*	*2*	*1*	*2*	*2*	*2*	*2*	*3*	*3*	*2*	*1*	*2*	*3*
actions/decisions	explore lichen reindeer relationship				x				x	x				x			
	increase number of reindeer					x	x				x	x					
	decrease number of reindeer		x	x				x					x		x		x
	substantially decrease number of reindeer	x															
pre-test evaluation V	*quality of formulated decision*	*4*	*3*	*3*	*2*	*1*	*1*	*3*	*2*	*2*	*1*	*1*	*3*	*2*	*3*	*-*	*3*

Table 2: Post-test evaluation of understanding

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
cues and non-salient factors	reindeer/herd size	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen/lichen density	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen is a renewable resource					x											
	lichen growth rate	x	x	x		x	x	x	x	x	x	x	x	x	x	x	x
	reindeer eat lichen	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
	lichen growth depends on lichen density				x	x					x	x					x
post-test evaluation I	*number of identified cues*	*4*	*4*	*4*	*4*	*6*	*4*	*4*	*4*	*4*	*5*	*5*	*4*	*4*	*4*	*4*	*5*
	I can only adjust lichen density by adjusting the number of reindeer	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x
	the number of reindeer should depend on lichen density		x	x	x	x		x	x			x	x	x	x	x	
post-test evaluation II	*number of identified non-salient relationships*	*1*	*1*	*2*	*2*	*2*	*1*	*2*	*2*	*1*	*1*	*2*	*2*	*2*	*2*	*2*	*1*
goals	achieve an equilibrium between the reindeer herd size and the lichen so that the lichen density does not decline any more																
	increase herd size																
	reduce the herd size so that the lichen density could increase																
	maintain a large herd																
	not sure/other																
	achieve, as soon as possible, the herd size that allows for the highest growth of lichen																
post-test evaluation III	*formulated correct goal*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
causal relations	If I want to maintain/stabilize a given lichen density then grazing rate needs to be equal to the lichen growth rate		x		x	x	x	x	x	x	x	x	x	x	x	x	x
	If my pasture is overgrazed I have to reduce the number of reindeer		x	x			x	x		x	x	x	x		x	x	x
	If grazing is below lichen growth I can increase the number of reindeer			x				x				x	x		x		x
	Maximum sustainable grazing equals the maximum lichen growth rate			x		x	x		x		x	x		x			x
	The higher lichen density, the higher lichen growth. At some point in time, though, this relationship is reversed and becomes: the higher lichen density, the lower lichen growth							x									
post-test evaluation IV	*number of identified causal relations*	*0*	*2*	*3*	*1*	*2*	*3*	*4*	*2*	*2*	*3*	*4*	*3*	*2*	*3*	*2*	*4*
actions/decisions	explore lichen reindeer relationship	x						x			x	x		x	x		
	increase number of reindeer																
	decrease number of reindeer	x								x							
	substantially decrease number of reindeer		x	x									x			x	x
post-test evaluation V	*quality of formulated decision*	*2*	*4*	*4*	*-*	*-*	*-*	*2*	*-*	*3*	*2*	*2*	*4*	*2*	*2*	*4*	*4*

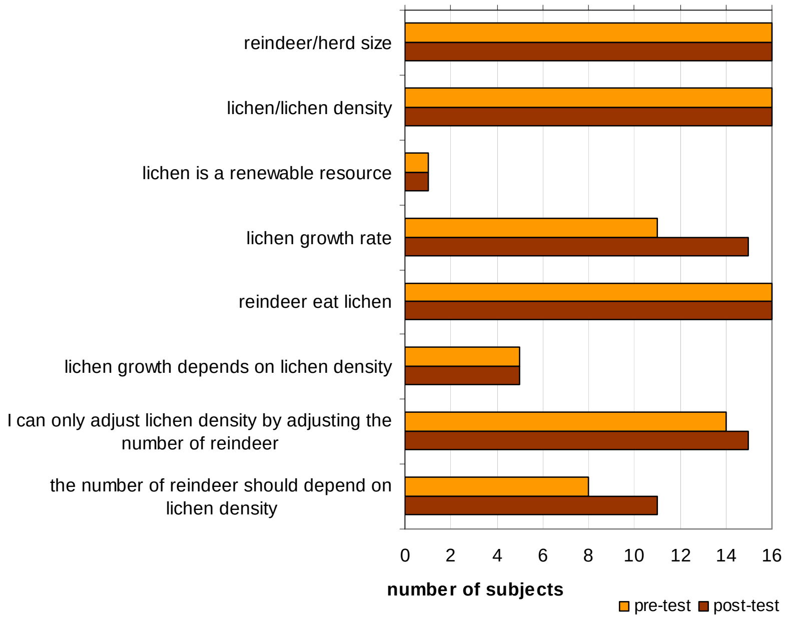
5.2 Findings from pilot evaluation

The findings presented in Table 1 to Table 2 give some indications about understanding in the four aspects of recognition and about the development of understanding in the course the interaction with the simulation-based task.

All the 16 subjects for whom data are available mention the two stocks, reindeer and lichen, when prompted for the cues before they were interacting with the simulator. Out of the six cues identified in the expert decision making process for the reindeer management task an average of four were mentioned with a minimum value of three and a maximum value of five. All subjects mentioned one non-salient factor and about half also mentioned a second salient factor relevant for the reindeer management task.

Figure 1 shows that the number of identified cues and non-salient factors does not change very much in the course of the experiment. Lichen growth rate is the exception; subjects seem to realize the importance of changes in the lichen stock while interacting with the simulator. The other changes concern the two non-salient factors, i.e. the factors that are not explicitly mentioned in the instructions but that have to be derived from the description: the fact that lichen density can only be changed through changing the herd size and the fact that lichen density determines how many reindeer should be kept.

Figure 1: Development of identified cues and non-salient factors in the course of the experiment

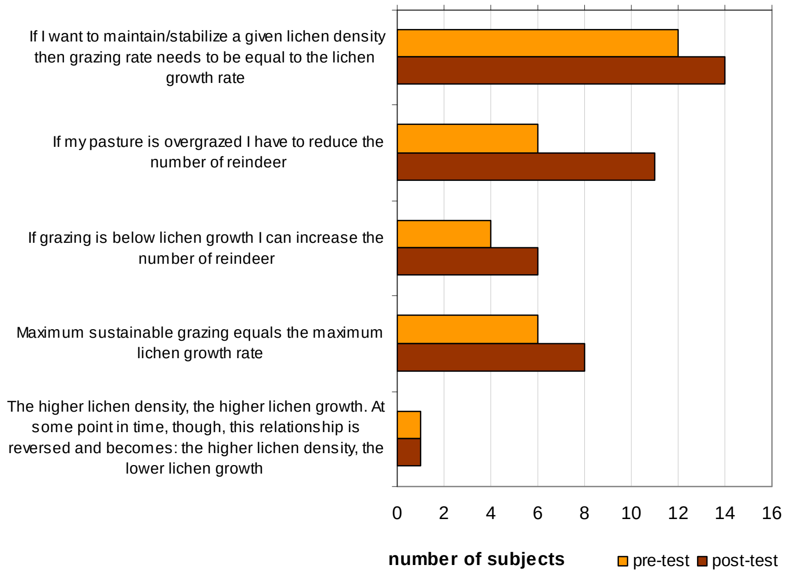


Fewer causal relations were mentioned than cues. Out of the five causal relations identified in the expert decision making process between zero and three were mentioned with most subjects mentioning two. The number of identified causal relations, on the other hand, changed much more in the course of the experiment than then number of cues (Figure 2). These changes affect mainly two aspects of the reindeer management task:

- The equilibrium conditions: grazing and lichen growth need to be equal; maximum grazing is possible at maximum lichen growth
- A more static picture of the reindeer management task: the more reindeer the less lichen; the less reindeer the more lichen (if my pasture is overgrazed I have to reduce the number of reindeer; if grazing is below lichen growth I can increase the number of reindeer).

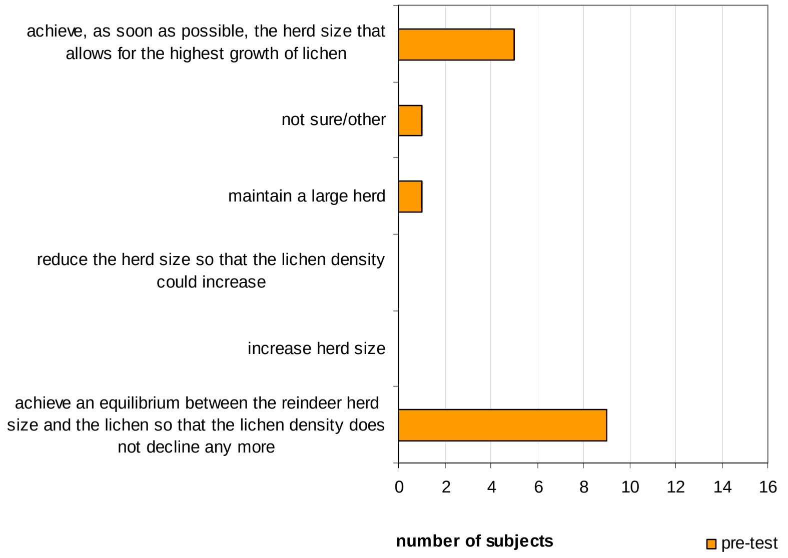
The nonlinear relationship between lichen density and lichen growth is mentioned only by one subject and there are no changes pre- and post-test. This would be the most important element of a more dynamic mental representation of the reindeer management task.

Figure 2: Development of identified causal relations in the course of the experiment



The correct goal of the task would have been “achieve, as soon as possible, the herd size that allows for the highest growth of lichen”. This goal was crossed by about a third of the subjects (five out of 16; Figure 3). A majority of the subjects focused on the stability of lichen (“achieve an equilibrium between the reindeer herd size and the lichen…”). We only asked about the goal in the pre-test. The available data therefore give no indication about changes in the goals.

Figure 3: Identified goals before interacting with the simulation

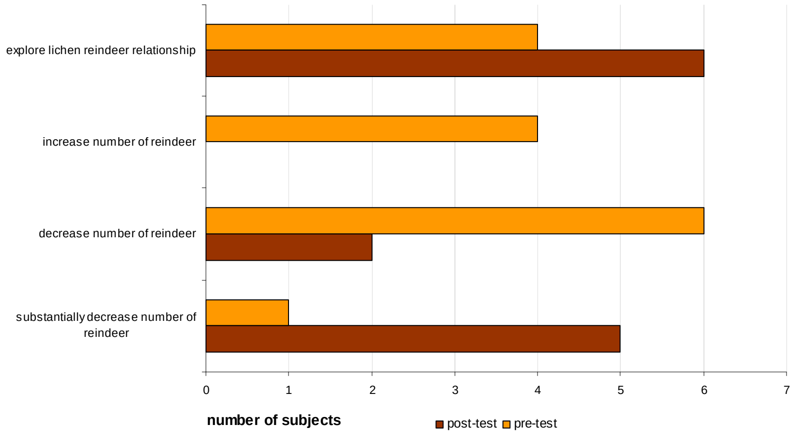


Only a minority of the subjects formulated expert like decisions before interacting with the simulator and on average subjects also did not move towards expert like decisions (Figure 4). A movement towards the expert like decision would imply a movement from increase in herd size to decrease in herd size to the final substantial decrease in herd size.

Exploration is important in trial 1 and becomes important again in the post-test understanding assessment. Those subjects who also with a more correct strategy such as decreasing or substantially decreasing the number of reindeer did not manage to successfully complete the task went back to suggesting a more thorough exploration of the lichen-reindeer relationship.

An increase in the number of reindeer, which is exactly the opposite of what you should be doing, only occurs in the first trial. The expert like decision (substantially decrease number of reindeer) was mentioned by only one subject in the pre-test understanding assessment. After completing the experiment about a third of all subjects suggested to substantially decreasing the number of reindeer to solve the task.

Figure 4: Development of the quality of decisions in the course of the experiment



6 Outlook

The purpose of this paper was to develop a step by step guide for the evaluation of understanding in dynamic decision making tasks. The guidelines add to the existing efforts in this direction in the system dynamics field and set the stage for applying tools like MITOCAR by providing the framework and discussing the kind of information that should be considered when evaluating understanding. The step by step guide was based on the expert decision making process and applied to the reindeer management task. We analyzed data collected in previous experiments with the reindeer management task to test the applicability of the guidelines.

The guidelines differentiate between four aspects of recognition: cues and non-salient factors; causal relations; goals; decisions. A tentative summary of the findings for the four aspects of recognition is that understanding is highest for cues, followed by causal relations and eventually by goals and decisions. Understanding seems to move towards more expert like understanding especially in the case of causal relations while the number of identified cues and non-salient factors remains largely constant. The changes in the quality of subjects' decisions reflect a more ambiguous picture and give no clear indication that completing the experiment helps building a more dynamic representation of the reindeer experiment task.

The available data do not cover all aspects of recognition for all steps (pre, during and post task) so that our results have to be interpreted with care. Rather than giving conclusive insights about understanding, the test application provides directions for future research. As throughout the rest of the paper we apply these directions to the specific case of the reindeer management task. By doing so we hope to be as specific as possible:

- Refine expert decision making process: establish a pool of experts that provide data for the four aspects of recognition
- Develop a full set of questions for evaluating understanding before, during and after the task that fully takes into account the concept of repeating measures

- Develop a full coding system for the answers
- Run experiments to collect complete sets of data
- Use MITOCAR to evaluate data
- Adjust step by step guide depending on the insights from and experiences with the applications

7 References

Ajzen I. 2002. Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior¹. pp. 665-83

Berry D, Broadbent DE. 1984. On the relationship between task performance and associated verbalizable knowledge. *Quarterly journal of experimental psychology* 36A: 209-31

Berry D, Broadbent DE. 1987. The combination of explicit and implicit learning processes in task control. *Psychological research* 49:7-15

Böhm G, Pfister H-R. 2001. Mental representation of global environmental risks. *Research in Social Problems and Public Policy* 9:1-30

Booth Sweeney L, Sterman JD. 2000. Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review* 16:249-86

Broadbent DE. 1990. Effective decisions and their Verbal Justification. *Philosophical Transactions of the Royal Society of London* 327, No. 1241, Human Factors in Hazardous Situations:493-502

Cavaleri S, Sterman JD. 1997. Towards evaluation of systems thinking interventions: a case study. *System Dynamics Review* 13:171-86

Chi MTH, Glaser R, Rees E. 1982. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*, ed. R Sternberg, pp. 7-76. Hillsdale, NJ: Erlbaum

Clarkson GP, Hodgkinson GP. 2005. Introducing Cognizer (TM): A comprehensive computer package for the elicitation and analysis of cause maps. *Organizational Research Methods* 8:317-41

Cronin MA, Gonzales C. 2007. Understanding the building blocks of dynamic systems. *System Dynamics Review* 23:1-17

Cronin MA, Gonzales C, Sterman JD. 2008. Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes* 108:116-30

Doyle JK, Radzicki MJ, Trees WS. 2008. Measuring change in mental models of complex dynamic systems. In *Complex Decision Making*, ed. H Qudrat-Ullah, JM Spector, PI Davidsen, pp. 269-94. Berlin/Heidelberg: Springer

Epstude K, Roese NJ. 2008. The functional theory of counterfactual thinking. *Personality and Social Psychology Review* 12:168-92

Ericsson KA, Smith J, eds. 1991. *Toward a general theory of expertise: Prospects and limits.* . Cambridge, MA: Cambridge University Press

Fum D, Stocco A. 2003. Outcome evaluation and procedural knowledge in implicit learning. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*

Goodwin GP, Johnson-Laird PN. 2006. Reasoning about the relations between relations. *Quarterly journal of experimental psychology* 59:1047-69

Huz S, Andersen DF, Richardson GP, Boothroyd R. 1997. A framework for evaluating systems thinking interventions: An experimental approach to mental health system change. *System Dynamics Review* 13:149-69

Jensen E. 2005. Learning and transfer from a simple dynamic system. *Scandinavian Journal of Psychology* 46:119-31

Jensen E, Brehmer B. 2003. Understanding and control of a simple dynamic system. *System Dynamics Review* 19:119-37

Jensen E, Sawicka A. 2006. What is the use of basic dynamic tasks? *24th International Conference of the System Dynamics Society*. Nijmegen, The Netherlands

Klein G. 1997. The recognition primed decision model:looking back, looking forward. In *Naturalistic decision making,* ed. CE Zsambok, GA Klein: Lawrence Erlbaum Associates

Larkin JH, McDermott J, Simon DP, Simon HA. 1980. Models of competence in solving physics problems. *Cognitive Science* 4:317-45

Luna-Reyes LF, Andersen DL. 2003. Collecting and analyzing qualitative data for system dynamics: methods and models. *System Dynamics Review* 19:271-96

Martin MK, Gonzalez C, Lebiere C. 2004. Learning to make decisions in dynamic environments: ACT-R Plays the beer game, pp. 178-83

Moxnes E. 1998. Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization* 37:107-27

Moxnes E. 2004. Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review* 20:139-62

Moxnes E, Saysel AK. 2009. Misperceptions of global climate change: information policies. *Climatic Change* 93:15-37

Pirnay-Dummer P. 2006. *Expertise und Modellbildung: MITOCAR.* Freiburg University, Freiburg

Sawicka A, Gonzalez JJ, Qian Y. 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, Kopainsky B. 2008. Simulation-enhanced descriptions of dynamic problems: Initial experimental results. In *26th International Conference of the System Dynamics Society*. Athens, Greece

Sawicka A, Kopainsky B. under revision. Simulation-enhanced descriptions of dynamic problems: Initial experimental results about their contribution to performance in dynamic tasks. *System Dynamics Review*

Seel NM. 1991. *Weltwissen und mentale Modelle*. Göttingen: Hogrefe

Seel NM. 2003. Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning* 1:59-85

Spector JM. 2006. A methodology for assessing learning in complex and ill-structured task domains. *Innovations in Education and Teaching International* 43:109-20

Spector JM, Christensen DL, Sioutine AV, McCormack D. 2001. Models and simulations for learning in complex domains: using causal loop diagrams for assessment and evaluation. *Computers in Human Behavior* 17:517-45

Stanley WB, Mathews RC, Buss RR, Kotler-Cope S. 1989. Insight without awareness: on the interaction of verbalization, instruction and practice in a simulated process control task. *Quarterly Journal of Experimental Psychology* 41A:553-77

Sterman JD, Booth Sweeny L. 2007. Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter. *Climatic Change*:213-38

Sweller J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12:257-85

Appendix: Instructions to the reindeer management task

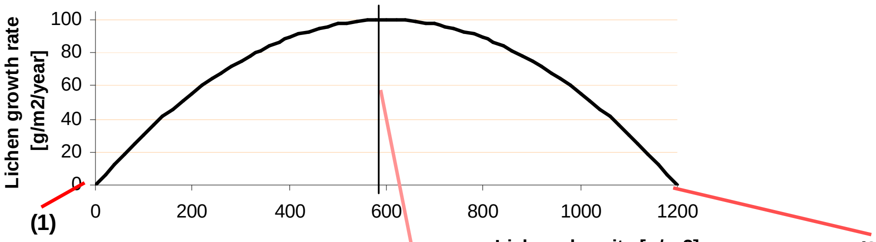
For this activity you will play the role of the manager of a reindeer herd. Your task is to produce as many reindeer as possible. But you must also make sure that the animals do not overgraze the lichen, which is the limiting source of food for the reindeer in winter.

Setting

Your reindeer herd grazes on a pasture used exclusively to feed your herd. Hence its resources will depend only on your decisions regarding the herd size. In summer, food supply is no problem – there is always plenty of grass and herbs. In winter, the food is scare and limited to lichen. If there is no lichen, all the animals will die.

Lichen is a low-growing species that is part plant and part fungus.

Lichen re-grows itself during summer when the reindeer feed on other plants. Lichen grows by propagating its spores. Lichen growth depends on its density and is described by an inverse U-shaped function as illustrated below.



(1)
When there is very little lichen,
i.e. when lichen density is very low,
there are just a few spores and
there will be only little growth.

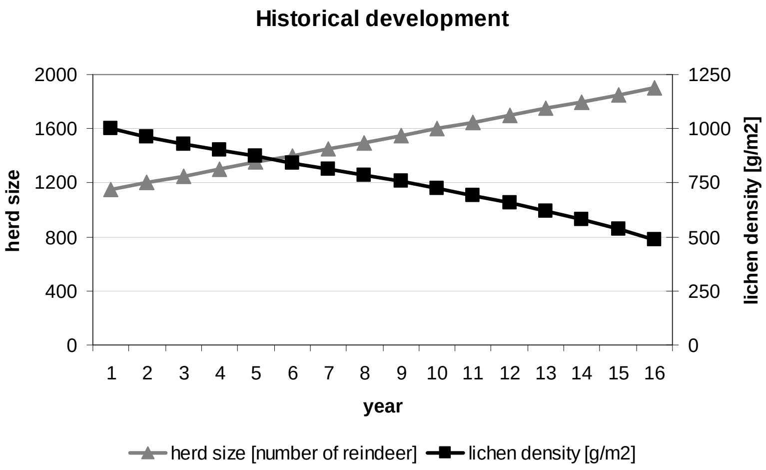
(2)
When there is very much lichen,
i.e. when lichen density is very high,
they start to fold onto themselves
and stop growing.

(3)
In between these extremes,
the growth reaches a maximum.

Grazing by reindeer affects lichen density. It therefore also influences the lichen growth rate. You should assume that 1000 reindeer eat 80g/m^2 of lichen during one winter. So as you can see, the reindeer are dependent upon the lichen, but the lichen is dependent upon the reindeer as well. That means that you have to maintain both the reindeer <u>and</u> the lichen populations together.

Starting point

The previous owner has steadily increased the number of reindeer from 1150 to 1900. As a consequence, the lichen density [g/m^2] has dropped from 1000 to 488 g/m^2. This development is shown in the following diagram and table.



Year	lichen density [g/m^2]	herd size [number of reindeer]
1	1000	1150
2	964	1200
3	930	1250
4	900	1300
5	872	1350
6	842	1400
7	814	1450
8	786	1500
9	756	1550
10	726	1600
11	694	1650
12	658	1700
13	622	1750
14	582	1800
15	538	1850
16	488	1900

Decisions to make

It is your job to decide how to maximize the size of your reindeer herd, while maintaining a manageable lichen density. You cannot control the lichen directly. You can control the number of animals you want to keep on the pasture, and that controls the amount of grazing (food eaten) by the animals.

Each year for 15 years, you will set a desired herd size. You are trying to have the maximum number of animals you can, while also maintaining the lichen at the best density for its growth. You should try to achieve the maximum sustainable herd size as soon as possible.

You can vary the herd size freely: You do not have to think about the sex ratio, the number of calves, losses of animals, or the age structure of the herd.