Incorporating Delays in the Decision-Making Interface: 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
Felicjan Rydzak
Centre for Advanced Manufacturing Technologies
Wroclaw University of Technology
ul. Lukasiewicza 3/5, 50-371 Wroclaw, Poland
Tel: +48 71 3204184 / Fax: +48 71 3280670
Email: felicjan.rydzak@pwr.wroc.pl

Delays are one specific factor contributing to misperceptions of dynamics. An experimental study was conducted to investigate how different representations of delays in the decision making interfaces (DMIs) may affect people's ability to manage and understand a dynamic system. A simple production-inventory management game was developed with four distinct DMIs, each featuring the production delay in a different way. Subjects were assigned randomly to use one of the four DMIs and a single-subject, think-aloud experimental protocol was deployed to gather data on the decision making process. No vivid impact of the different representations of the delay in the DMI was observed. However, data gathered through the single subject experimental protocol suggest that the subjects do not follow the anchoring and adjustment rule proposed earlier (see Sterman 1989). Rather, they develop a simplified decision rule that is not robust to changes in the task settings but that is successful in the context of the particular experimental task.

Key words: *decision-making interface, delays, simulation-based learning environment, supply chain management, production management*

Introduction

The fact that people have difficulties in controlling dynamic systems is now rather well documented in the system dynamics literature. Initially, the poor performance was explained in the context of the misperception of feedback (MOF) hypothesis formulated by Sterman (1989, 1994). Drawing on his experimental studies of dynamic decision making as well as the studies conducted by others, Sterman posited that people mismanage dynamic and complex systems because they *"adopt an event-based, open-loop view of causality, ignore feedback processes, fail to appreciate time delays between action and response and in the reporting of information, do not understand stocks and flows, and are insensitive to nonlinearities that may alter the strengths and different feedback loops as a system evolves"* (1994, p. 305). All these misconceptions occur even when complete information about the system is provided; they also seemed to be robust to experience or financial incentives.

The rather bleak outlook on people's ability to perform in dynamic settings was somewhat modified by the subsequent studies (Sengupta and Abdel-Hamid 1993, Machuca et al. 1998, Howie et al. 2000,). It was found that both the design of decision-making interface as well as the content of initial instructions might impact the performance: When aspects critical to understanding the dynamics are emphasized, not only the subject's performance improves, but also their knowledge of the system. Although encouraging, the results are far from satisfactory – one of the main issues being their relative scarcity; the other, the rather poor specificity of the existing recommendations for how to improve decision making in dynamic settings. The prevailing theme has been that people need to know the underlying causal structure before they are able to manage successfully a dynamic system. Such guideline seems quite generic in the context of the MOF hypothesis that stipulates four specific aspects of dynamic systems that are especially difficult for people to handle: (1) feedback relationships, (2) delays, (3) accumulation processes and (4) nonlinearities.

In this paper we report on an experiment conducted to explore effectiveness of the various ways in which one of the four aspects, namely delays, may be presented in the decision-making interface. For the purpose of our study we use an inventory stock control task. The underlying system dynamics model is analogous to the one used in earlier experiments by Diehl and Sterman (1995, Diehl 1992). In our study, we considered only the simplest scenario with one specific delay and with no multiplier feedback effect. Before discussing the experimental design and results, we review some key recommendations from the system dynamics literature on design of environments for dynamic decision-making.

How to improve people's ability to control a dynamic system

The general guideline emerging from the existing system dynamics literature is that providing cues about the system's structure will improve people's ability to manage the system. One of the ways to inform people about the structure of a dynamic environment is to reveal its underlying causal structure. Among the most vocal proponents of this approach was Machuca and associates (Machuca 1992, Machuca 2000, Zamora Gonzalez, Machuca, and del Castillo 2000). Their extensive experimental explorations showed that students provided with Transparent Box Business Simulators (TBBS) outperformed those who used a traditional, black-box style environment (Machuca et al. 1998). Still, the subjects had to be encouraged to use the help function that provided insight into the system's structure (Ibid., p. 5-6).

Similar observations were made by Großler, Maier, and Milling (2000). Also in their LEARN! business simulator information about structure was provided in the form of an on-line help function. Due to infrequent use of the help function, Großler, Maier, and Milling were unable to assess conclusively its effectiveness (Ibid., p. 270). However, it was observed that the subjects who received an explicit presentation of the system's structure before managing the simulated environment tended to perform better (Ibid., p. 269).

Others suggest that direct presentation of the model structure is not necessary and that the important causal relationships may also be conveyed otherwise. Referring to series of studies on people's ability to manage a simple thermo-hydraulic process, Vicente (1996) argues that design of decision-making interface is critical to performance. The interface beyond the physical aspects of the system should also convey the functional characteristics of the system's elements and their interrelationships. The so-called Ecological Design Interface (EID) framework was practically tested in the context of a system dynamics decision problem by Howie et al. (2000) who redesigned the

STRATEGEM interface (Sterman 1987) according to the EID-principles. Subjects who used the EID-based STRATEGEM interface tended to outperform those who played using the original interface. The EID-based STRATEGEM interface did not convey the causal structure of the system in an explicit way. Instead an effort was made to elucidate (through information arrangement, metaphors and animation) all the elements and relationships central to the system's dynamics. Additionally, the interface provided a range of graphs tracing the system's behavior.

While not acknowledged by Howie et al. the inclusion of behavior graphs is a potential confounding factor for their results. In an earlier study Sengupta and Abdel-Hamid (1993) observed an improvement in performance when the traditional text-based and outcome-oriented decision-making interface for managing a simulated software development project was enhanced with a set of behavior-tracing graphs. In the same study Sengupta and Abdel-Hamid tested how providing explicit instruction about key nonlinearities and decision heuristics would impact the subjects' performance: While the subjects in the instruction-receiving group outperformed those who did not receive such an instruction, they still were outperformed by the group with access to behavior graphs.

On the other hand, provision of graphs that trace the system's behavior during the decision making does not seem help people's performance in the renewable resources management tasks explored by Moxnes (2004).[1] In his earlier studies Moxnes (1998) tested also the impact of providing the subjects with an explicit graphical representation of the nonlinear relationship understanding of which is crucial for achieving control over the system. The treatment had a positive effect on the performance of some of the subjects. Others, however, could not interpret the information correctly and thus their decision-making process did not benefit from it. A similar type of finding is reported by Richardson and Rohrbaugh (1990): In their study they presented the subjects with an explicit decision rule that should be followed when playing the STRATEGEM capital acquisition management game. They found that only some of the subjects followed the rule.

The results by Richardson and Rohrbaugh (1990) were subsequently disputed by Bois (2002) who observed an improvement in the subjects' performance with a larger experimental sample. In his study, Bois used the EID-based STRATEGEM interface developed by Howie et al. (2000). In addition to testing the impact of the decision rule, he also tested the impact of an enhanced instruction providing a more detailed explanation of the game environment. The enhanced instructions seemed to have a positive effect both on the subjects' performance and knowledge.

One of the most interesting findings by Bois, however, is his observation that the subjects who reported investing more effort in carrying out the task outperformed the subjects who seemed to be less engaged. The observation seemed to be consistent across all experimental conditions, indicating that the degree of individual motivation may influence people's performance in a significant way regardless of the provided instruction or decision-making interface. A similar observation was made recently by Jensen and Brehmer (2003). Given this, it may be tempting to conclude that when people try hard, they can achieve good results. Since the individual motivation has not been explored in other studies, the conclusion cannot be ruled out completely. However, the body of other results (including other observations by Bois [2002]) suggests that

[1] A similar results were reported by Sawicka, Qian, and Gonzalez 2005 who replicated the experiments by Moxnes (2004) in a different problem context.

performance may also be improved through redesign of the decision-making environment. Still, as illustrated in our short review above the existing evidence is quite mixed: For almost each report of a positive effect of a particular treatment, we can find another study that does not report any significant change in performance given an analogous treatment. The inconsistency of the results may be due to various methodological weaknesses of the conducted studies (see e.g., Großler and Maier (2004)). It may, however, also be due to the rather small number of the observations made. We believe that more research is needed to determine whether and how people's ability to control dynamics systems may be improved. We also believe that it is important that future explorations delineate more clearly the connection between the administered treatments and the dynamic features of the explored system.

The MOF hypothesis stipulates four specific aspects of dynamic systems that are especially difficult for people to handle: (1) feedback relationships, (2) delays, (3) accumulation processes and (4) nonlinearities. Note that each of these aspects tends to be represented differently in the system dynamics notation. Still, most of the studies examining how people's ability to control a dynamic environment may be improved seem to provide recommendations that are independent of the four specific aspects of a dynamic system: reveal the causal structure, trace graphically behavior of the key system variables, provide a decision rule, increase the decision makers' motivation, etc. While providing some guidance, the generic recommendations leave much room to intuition in development of effective decision aids for control of dynamic environments. One way of increasing the guidelines' specificity is to explore whether we can identify ways that are especially effective for conveying each of the four basic aspects of the dynamic environment. The study we present in this paper focuses on delays. We explore four different ways of providing information about accumulation delay. The results of our investigations provide input to the research project that aims at identifying a set of specific guidelines for effective communication of system dynamics models.[2]

Incorporating delays in the decision-making interface

Experimental task analysis

The devastating effects of the delays on people's performance in dynamic settings have been demonstrated in a number of experimental studies (see e.g., Dörner 1975; 1989 (1996), Sterman 1987, Sterman 1989, Brehmer 1992). In all these studies, the subjects were informed about the inherent delays and still failed to adopt their decision rules accordingly. Previous results (see e.g., Sengupta and Abdel-Hamid 1993, Vicente 1996, Howie et al. 2000) suggest that design of decision-making interface may impact people's performance in and understanding of complex settings. Here we explore how the various ways of conveying information about the inherent system delays may improve the way in which people handle them. We conduct our explorations using a simple production-inventory stock control task. The task is similar to this used in Diehl and Sterman (1995, see also Diehl 1992). The subjects manage a simple production system over a number of decision periods. For each trial, the subjects are presented with the current inventory level, sales, production, and costs. Based on this information they are asked to set the production rate for the next decision period. Their objective is to

[2] More information about the project will be provided after the blind review.

minimize cumulative costs arising from off-target inventory levels and production changes.[3]

Diehl and Sterman explored how people's performance varied as a function of task complexity. In their study, they varied the length of the production delay and the strength of the multiplier feedback effect (i.e., the effect of production rate on sales). As we are interested only in the impact of the production delay, we do not consider the multiplier feedback effect (i.e., γ, indicating the strength of the multiplier feedback effect, equals 0 in our case, see Diehl and Sterman 1995). Furthermore, we consider only one delay length. This is because we are not interested in how the delay's length impacts the subjects' performance, but rather how the performance changes given various representations of the production delay and its effects.

We embed our experimental task in a scenario in which the subjects are asked to take on the role of a production manager of one of the divisions of the ABC Manufacturing Co. From the instructions[4] they learn that the production capacity of the division they will be heading has been recently enhanced to prepare for the planned expansion to the new markets. Consequently, the production rate may be increased instantaneously when the need arises, i.e., when the orders from the new market start to arrive. The subjects run three trials involving 32 decisions each. Each trial the subjects are exposed to a random walk of customer order flow. Initially, the orders are at the level of around 100 orders per week; in the first third of the decision period there is a step increase in the customer order rate. The random walk as well as the time and height of the step increase vary across the trials to prevent the subjects from simple reconstruction of decision series that were found successful during the previous trials.[5]

Figure 1 presents the system dynamics model underlying our dynamic decision making environment.

[3] Costs are generated through production adjustments and off-target inventory levels. We assume a linear cost function, were contributions of the absolute inventory deviations and production changes are weighted equally:

$$Cost = \sum_{t=1}^{t=32} (|Inventory_t| + |Production\ Rate\ Change_t|)$$

[4] See Appendix 1.

[5] The three series of orders are provided in Appendix 2.

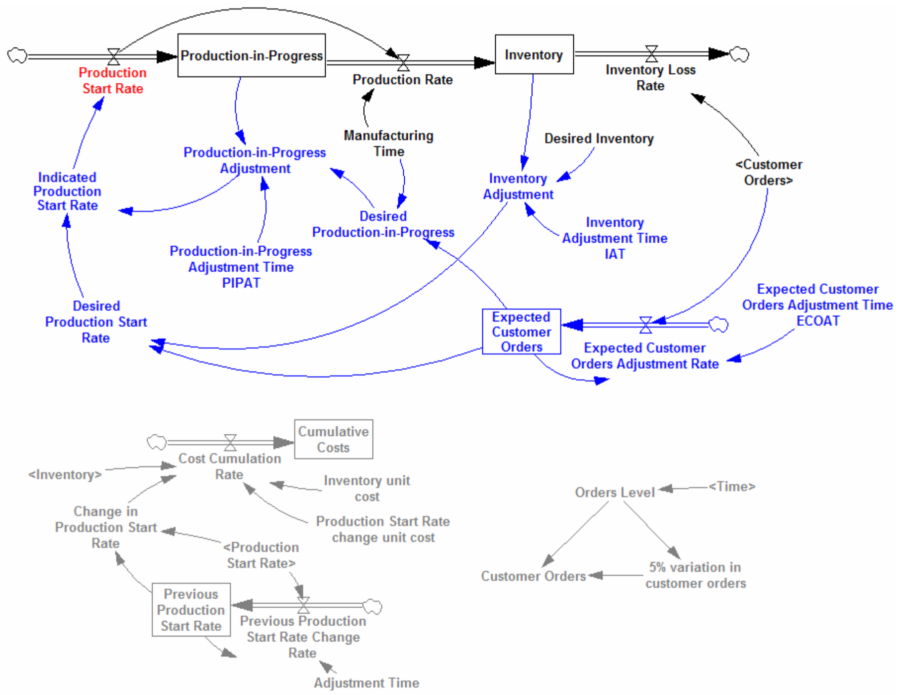


Figure 1 Stock-and-flow structure of the simple production-inventory stock control system as implemented in the experiment.[6]

The various variable types are color-coded in Figure 1:

- the decision variable – *Production Start Rate* – is marked in red,
- variables values of which are available to the subjects during the experiment are marked in black,[7]
- variables necessary to compute the costs and generate the customer order flow are colored grey,
- variables that capture a possible decision rule are marked in blue.

The decision rule incorporated in the model (see the blue variables) is not activated during the experiments and is developed as proposed in Sterman (2000, Chapter 17). It assumes that the optimal production rate is derived based on:

(1) the desired inventory adjustment (i.e., if inventory differs from the desired level, how much it should be adjusted),

(2) the expected customer orders (i.e., what is the expected customer order rate), and

(3) the work in progress level (i.e., what is the volume of production in progress).

As indicated in Figure 1 each of the decision elements (i.e., *Inventory Adjustment*, *Expected Customer Orders*, *Production-in-Progress Adjustment*) has an adjustment time associated with it. Short adjustment times indicate that the production start rates are promptly adjusted to any discrepancies detected in the inventory, customer orders or

[6] Fully documented Vensim model is provided in the supplementary materials accompanying this paper.
[7] *Production-in-Progress* is a special case: While all the pending production rates will always be displayed in the decisions-making interface, the interface will not provide a total volume of the production-in-progress. See also Table 1, p. 9.

production-in-progress levels; long adjustment times indicate that the observed discrepancies are not acted upon immediately.

In Appendix 3 we present a collection of simulations runs that illustrate the impact of the various adjustment times on the cumulative costs and the overall behavior of the system. The case, where the decision maker reacts promptly to the inventory and production-in-progress discrepancies but does not adjust for customer order fluctuations (unless there is a dramatic change in the order level), yields best results (see case D in Appendix 3). This is consistent with the results by Sterman (1989) who formulates the optimal decision making rule as follows:

$$PSR_t = MAX[0, ExpL_t + (I^* - I_t) + (SL^* - SL_t)] \qquad (1)$$

where PSR_t indicates the current production start rate, $ExpL_t$ – the imminent expected losses, I* – the desired inventory, I_t – the current inventory, SL_t – the current supply line and SL*– the desired supply line.

Based on regression analysis, Sterman suggests that poor inventory control may be attributed to the subjects' fixation on the initial inventory level and their failure to account fully for the supply line (Sterman 1989, see pp.334-335). Indeed, our analysis of how the different adjustment times impact the performance, indicates that the most costly oscillations arise when production-in-progress is ignored, see cases B and G in Appendix 3.

The failure to account for the supply line has been seen as indicative of the difficulty people have in assessing the implications of time lags that delay the expected accumulations (see e.g., Diehl and Sterman 1995, p. 211). The focus of this investigation is to gain a better insight into how people deal with time delays and how their ability to correctly assess implications of time delays may be enhanced through an appropriate design of decision-making interface. In the following sections, we discuss the specific interface characteristics expected to affect the performance that were tested in our experiments and the single subject approach applied to gather more detailed data regarding the decision making process.

Experimental treatments: The four decision-making interface designs

As indicated earlier, our investigations were carried out within a research project that aims at identifying a set of guidelines for effective communication of system dynamics models.[8] Identification of such guidelines requires both theoretical and empirical studies. A theoretical framework building upon Cognitive Load Theory (CLT, Sweller 1988, Chandler and Sweller 1991) is proposed by Sawicka (2005). CLT is a theory of the cognitive system that explains the interplay between the learning and instruction processes. Its main premise is that any learning situation should be designed in such a way that the limited human processing resources are used effectively. Presenting the learners with the material that is difficult for them to understand is likely to cause a cognitive overload, hindering any effective learning. However, it is not only the complexity of the material itself that may impact the learning. Also the way in which the material is presented may hamper or aid learning.

The working hypothesis for this study is that people are aware and understand the role of time lags such as the production delay featured in our experimental task (see the

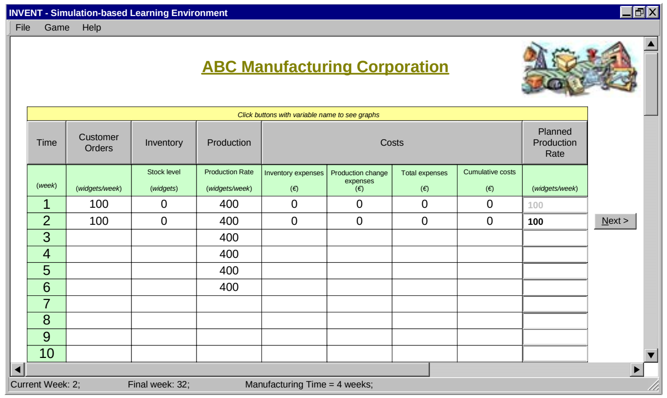
[8] See footnote 2, p. 3

Manufacturing time in Figure 1, p. 6). When asked they would indicate that the delay in production is the inherent part of the system they are supposed to control; still, most of them, when setting the production start rate, will fail to account appropriately for its effects, in particular for the accumulated *Production-in-Progress*, see Figure 1, p. 6. This we would expect to occur when the decision-making interface fails to provide sufficient support for the inference of the current 'production-in-progress' level. For most people setting the production start rate would cause a substantial cognitive burden. Given the stretched cognitive resources, they may easily overlook decision elements that are not explicitly available. According to CLT such problems may be alleviated by redesign of the instruction so that the learning environment provides more support for inferences that are necessary during the problem solving.

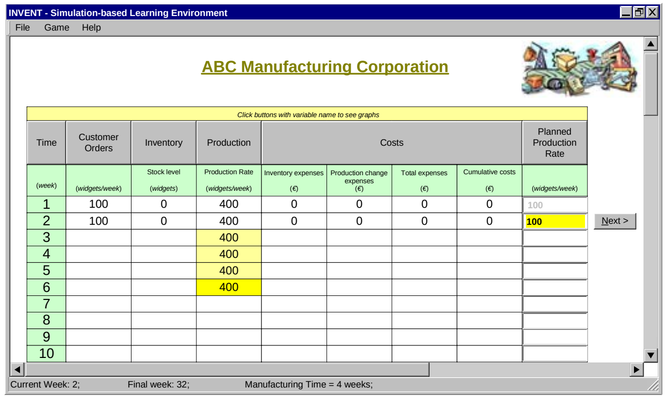
To determine what type of presentation format is likely to be most effective in the current study we explored four different ways of presenting our stock control task and, in particular, the manufacturing delay and the resulting accumulation of the 'production-in-progress.' Table 1 provides an overview of the decision-making interface designs tested.

Table 1 Overview of the experimental conditions: Prototypical interfaces of the dynamic-decision making environment

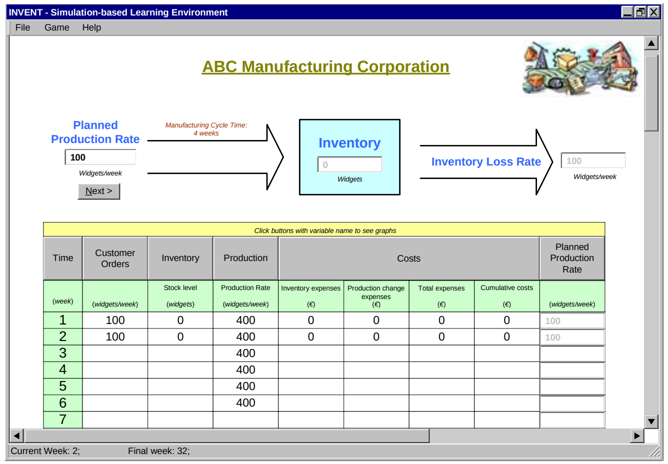
<table>
<tr><td></td><td></td><td colspan="2">PRODUCTION-IN PROGRESS</td></tr>
<tr><td></td><td></td><td>IMPLICIT</td><td>EXPLICIT</td></tr>
<tr><td rowspan="2">SYSTEM STRUCTURE</td><td>IMPLICIT</td><td>DMI_1



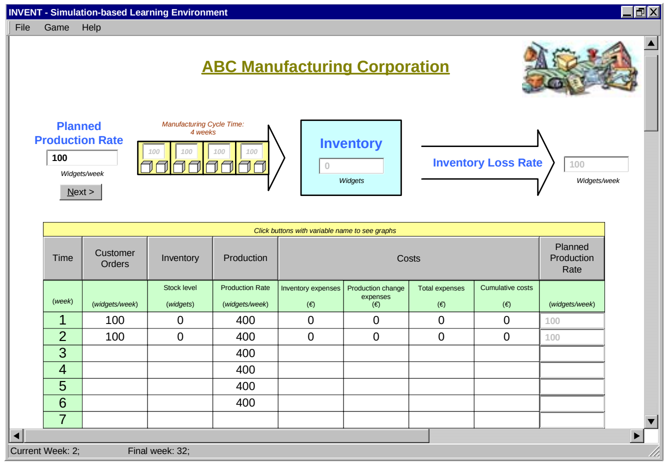
</td><td>DMI_2



</td></tr>
<tr><td>EXPLICIT</td><td>DMI_3



</td><td>DMI_4



</td></tr>
</table>

All four interfaces provide full information required for setting the production rate. They differed, however, with respect to:

- The way in which the decision space in general was presented (text-based vs. diagrammatically-enhanced representation, see rows in Table 1), and
- The way in which the manufacturing delay and the resulting 'production-in-progress' accumulation were presented (a mere enlistment of pending productions vs. an explicit indication of the production-in-progress content, see columns in Table 1).

All four interfaces contain a spreadsheet like report on the problem variables. The report provides the decision-maker with <u>complete information</u> about the current and past results. In addition to the numerical values, the decision-maker may obtain a graphical representation of the results by clicking on the graph icon placed next to each of the column headers.

While the spreadsheet report provides the decision maker with a complete information necessary for setting the production rate, we would expect that this text-based information presentation format alone (DMI1 and DMI2 in Table 1) is more difficult to handle than the diagrammatically-enhanced presentation (DMI 3 and DMI4 in Table 1). The diagrammatically-enhanced format provides the decision-maker with a decision-space template where the various elements are differentiated with respect to their role in the system:[9] The inventory is represented as an accumulation with two associated flows: an inflow due to production and an outflow due to sales. The analyses conducted by Larkin and Simon (1987) indicate that diagrammatically-enhanced problem representations may reduce substantially cognitive effort required to deal with a complex problem. Given CLT (Cognitive Load Theory, Sweller 1988,Chandler and Sweller 1991) principles, the decision-makers should have more cognitive resources at their disposal and, consequently, be able to make better decisions when interacting with the diagrammatically-enhanced interface.[10] Hence, our first experimental hypothesis may be formulated as follows:

> H_1: Diagrammatically-enhanced decision-making interface will have a positive effect on people's performance and understanding of the simple production-inventory stock control task.

Still, even a diagrammatically-enhanced decision-making interface may provide a various degree of support with respect to how one should deal with the effects of the production delay. As indicated earlier (see p. 7), we expect that most people when confronted with the production-inventory stock control task will not have problems with understanding what the production delay is and how it affects the system. However, if the volume of the resulting 'production-in-progress' is not represented clearly enough in the decision-making interface, people will, most likely, fail to incorporate it into their calculations when setting the production rates. This omission may be attributed to the insufficient cognitive resources (CLT, Chandler and Sweller 1991);[11] an explicit indication of the 'production-in-progress' volume is likely to ease the cognitive burden,

---

[9] Such differentiation is also recommended by the Ecological Design Interface framework described in Vicente 1996.

[10] For a system dynamics analysis of the cognitive load dynamics implied by CLT see Sawicka (2005).

providing more support to the process of production setting. Consequently, our second experimental hypothesis may be formulated as follows:

> H_2: Explicit indication of the effects of the production delay, i.e., the 'production-in-progress' volume, will have a positive effect on people's performance and understanding of the simple production-inventory stock control task. In particular, given en explicit indication of the 'production-in-progress' volume, more people will account for 'production-in-progress' when setting their weekly production start rates.

In Table 2 we summarize what results we expect to observe in our study. In accord with H_1 we expect that diagrammatically-enhanced interfaces will result in a better performance and understanding (see rows in Table 2). The performance and understanding will improve further when the effect of the manufacturing delay is indicated explicitly (see H_2 above and columns in Table 2). In the remainder of this section we discuss how we intend to assess performance and understanding in our experimental study.

Table 2 Overview of the performance expected under each of the experimental conditions (for specification of the various decision-making interfaces, DMIs, see Table 1, p. 9)

		PRODUCTION-IN-PROGRESS	
		IMPLICIT	*EXPLICIT*
SYSTEM STRUCTURE	*IMPLICIT*	**DMI_1**	**DMI_2**
	EXPLICIT	**DMI_3**	**DMI_4**

Performance & Understanding

Best *Worst*

Experimental data collection and analysis

The primary source of data in experimental studies on dynamic decision making are logs of the subjects' decisions (see e.g., Sterman 1987, Richardson and Rohrbaugh 1990, Sengupta and Abdel-Hamid 1993, Howie et al. 2000, Moxnes 2004). While the decision logs represent an objective account of the subjects' performance, their analysis alone may not always be sufficient to determine decision rules employed by a particular subject.[11] As indicated in the MOF hypothesis it is possible that people acquire a correct mental model, possibly even develop a correct decision rule, and still mismanage the system due to a limited computing capacity of human mind:

[11] See e.g. Chapter 10 in Cooper, Heron, and Heward 1987

"The robustness of the misperceptions of feedback and the poor performance they cause are due to two basic and related deficiencies in our mental models. First, our cognitive maps of the causal structure of systems are vastly simplified compared to the complexity of the systems themselves. Second, ***we are unable to infer correctly the dynamics of all but the simplest causal maps****."* (Sterman 2000, p. 27, emphasis added)

To identify the performance-understanding discrepancies some researchers enhanced their experimental protocols with various types of questionnaires and workbooks. These instruments help to collect more data on the way the subjects make their decisions and understand the task. However, even with such enhanced data collection procedure, at times it may be difficult to evaluate analytically a particular decision-making process, i.e., to determine reasons that led to specific decisions. For example, the subjects, contemplating several decisions rules, may fail to provide the full account of the process in their workbooks, indicating only one of the intermediary rules. Consequently, discrepancies between the different data sources may arise. To circumvent these problems, the researchers would need data that provide yet a greater insight into the decision-making process. Such data may be gathered through the so-called think-aloud protocol.

The think-aloud protocol requires the subjects to verbalize their thoughts and explain their actions as they work through the task and make their decisions. This verbal explication is recorded and provides an additional source of data about the way the subjects dealt with the problem. The think-aloud protocol has not been frequently employed in the context of the dynamic decision-making experiments. The method was deployed by Dörner (1989), one of the pioneers of the experimental dynamic decision making research. The study by Jensen and Brehmer (2003) provides a more recent and – to the best of our knowledge – the only example of the use of think-aloud protocols in the context of the experimental task based on the system dynamics model. One of reason for the relative 'unpopularity' of this method is probably the fact that the think-aloud protocols provide vast amounts of raw data. Under the think-aloud protocol, the researchers have only a limited ability to structure the elicitation before the data collection occurs. Therefore analysis of the gathered data may often be quite cumbersome. Consequently, many may consider the potential additional insight as insufficient to justify the substantial increase in the effort required to analyze the think-aloud protocols.

Another reason for the relative reluctance to employ the think-aloud protocols may be that their deployment requires substantial resources not only for data analysis but also for data collection,[12] effectively precluding their use in a large-size group experiments. While the group-design seem to dominate the experimental studies conducted within the system dynamics field,[13] it is important to be aware of its limitations. For one, the group-design allows us only to learn about the behavior of an 'average' subject. If there are substantial differences between the ways in which various individuals approach the problem, the average results will not allow us to detect these differences; as a result, one can fail to understand the actual decision-making processes. In the context of the dynamic decision making it seems vital that the group analysis of experimental data is

[12] Deployment of think-aloud protocols requires individual data collection for each of the subjects.
[13] With the exception of the study by Jensen and Brehmer (2003), which had included elements of the single-subject design, all the system dynamics studies referred to in this paper implemented the group design.

always combined with inspection of individual cases (typical for the single-subject experimental design, see e.g., Sidman 1960, Barlow and Hersen 1984). The think-aloud protocols give us a unique and arguably best insight into the individual decision-making process. Such insight seems especially useful when the experimental investigations concern aspects of the individual behavior not yet explored in great detail (see e.g., Elashoff and Thoresen 1978, Shaughnessy and Zechmeister 1994).

As indicated earlier, existing system dynamics literature suggests that design of the decision-making interface may facilitate people's performance in and improve their understanding of dynamically complex tasks. However, the available results do not explicate which aspects of the decision-making interface are especially helpful in dealing with particular elements of a dynamically complex task. Our study was aimed to address this deficiency. Being among the firsts of its type, we believe that it may benefit from the single-subject design with a detailed data collection through the think-aloud protocols.

Another important reason for using the think-aloud protocol in our case was the feasibility of acquisition of the necessary data through other, more traditional and less resource-consuming data collection methods. In our study we wanted to identify which elements of the instructions and the decision-making interface facilitated the decision-making process, which elements misguided the process, and which hindered it.

It is difficult to assert up-front to what degree such information might be gathered effectively through the more traditional data collection tools such as workbooks and questionnaires.

Our investigations were quite similar to the regular usability assessments of the human-computer interface designs. The usability engineering literature recommends the think-aloud protocols and direct observations for such assessments. On the other hand, it might turn out that the more traditional data collection methods suffice our purposes. In another experiment on dynamic decision making, conducted by one of the paper's authors, the workbooks and follow-up questionnaire are used to collect information regarding the subjects' evaluation of the decision-making support provided during the experiment. In the discussion section of this paper we will compare the quality of data collected in this other study and with the quality of data gathered in the current study featuring the single-subject design with the think-aloud protocol.

We expected that the direct observation and think-aloud protocols would provide us with a better and more direct insight into two aspects of the decision making process: (1) what features of the decision-making interface (and the instructions) are attended to by the subjects and when, and (2) what are the decision rules employed by the subjects and how they were formulated. In addition to this data, we also deploy the more traditional data collection methods: To obtain a detailed and accurate account of the subjects' performance, we log automatically all their decisions. We also record when the subjects invoked the graphical representations of the system variables. This data will allow us to discuss what impact the over-time behavior graphs had on the decision-making process. The subjects are also provided with the workbooks in which they can conduct any analyses or calculations they wish. Finally, we administer the subjects with a task apprehension test to assess their understanding of the task prior and after their decision-making trials.

Based on these set of data we hoped to gain a detailed insight into the decision making processes of our subjects and to be able to determine how the process is affected by the various elements of the decision-making interface.

Experimental study

Design and procedure

The experimental study was conducted using a simulation-based production game called INVENT, described in the *Experimental task analysis* section, and involved 15 students attending MSc course in Management and Manufacturing Engineering at Wroclaw University of Technology, Poland. There were 8 women and 7 men, 22 to 27 years old. All of participants assessed their knowledge of production management as good (7 participants) or average (8 participants); 6 participants had prior experience in playing business/management simulation games.
The subjects were randomly assigned to one of the DMIs (see Table 1). Following the single subject experimental design principals (see also the Experimental data collection and analysis section), each subject attended the experiment individually. After reading the instructions (see Appendix 1), the subjects were asked to fill out the task apprehension test (see section Experimental data collection and analysis). Next, they played INVENT three times, using a DMI they were assigned to. Each decision and all results were saved in a computer log. During conducting the task the subjects were encouraged to comment on their decisions ('think-aloud protocol'). After each trial an interview was conducted to elicit self-evaluation of the performance. At the end of the experiment each subject filled in the same task apprehension test and evaluated the whole task.

Results

The results of the pre-task questionnaire indicated that most subjects understood the task well and only two subjects failed to provide a correct answer to the question about the manufacturing delay. Still, analysis of the think-aloud protocols and the debriefing interviews conducted following Trial 1 revealed that a number of subjects did not know how to act upon this information (4 out of 15) or have forgotten about it altogether (3 out of 15). Only one subject managed to control the system completely from the start – see S3 in Trial 1 in Table 3 where the results of all subjects are presented.[14] The subjects' performance is compared against the simulated performance assuming the optimal anchoring and adjustment heuristic proposed by Sterman (1989), see equation (1), p. 7. While the results of S3 trace precisely the optimal rule behavior, analysis of the think-aloud protocols suggest that the subject did not take adjust explicitly for the inventory or supply line, as suggested by the optimal rule (see equation (1), p. 7). Rather, his decision rule was anchored in the imminent expected losses (represented by the current customer order level, *CO*) and the losses expected over the next 4 weeks (i.e., duration of the manufacturing delay), and may be formulated as follows:

$$O_t = MAX[0, CO_t + \sum_{i=1..4} (ExpCO_i - PP_i)] \qquad (2)$$

where *ExpCO* indicates the expected level of customer orders and *PP* indicates planned production.

[14] The results of one subject (S2) are excluded from the analysis as this subject did not manage to understand at all the experimental task.

Table 3 Inventory levels – overview of the experimental results.

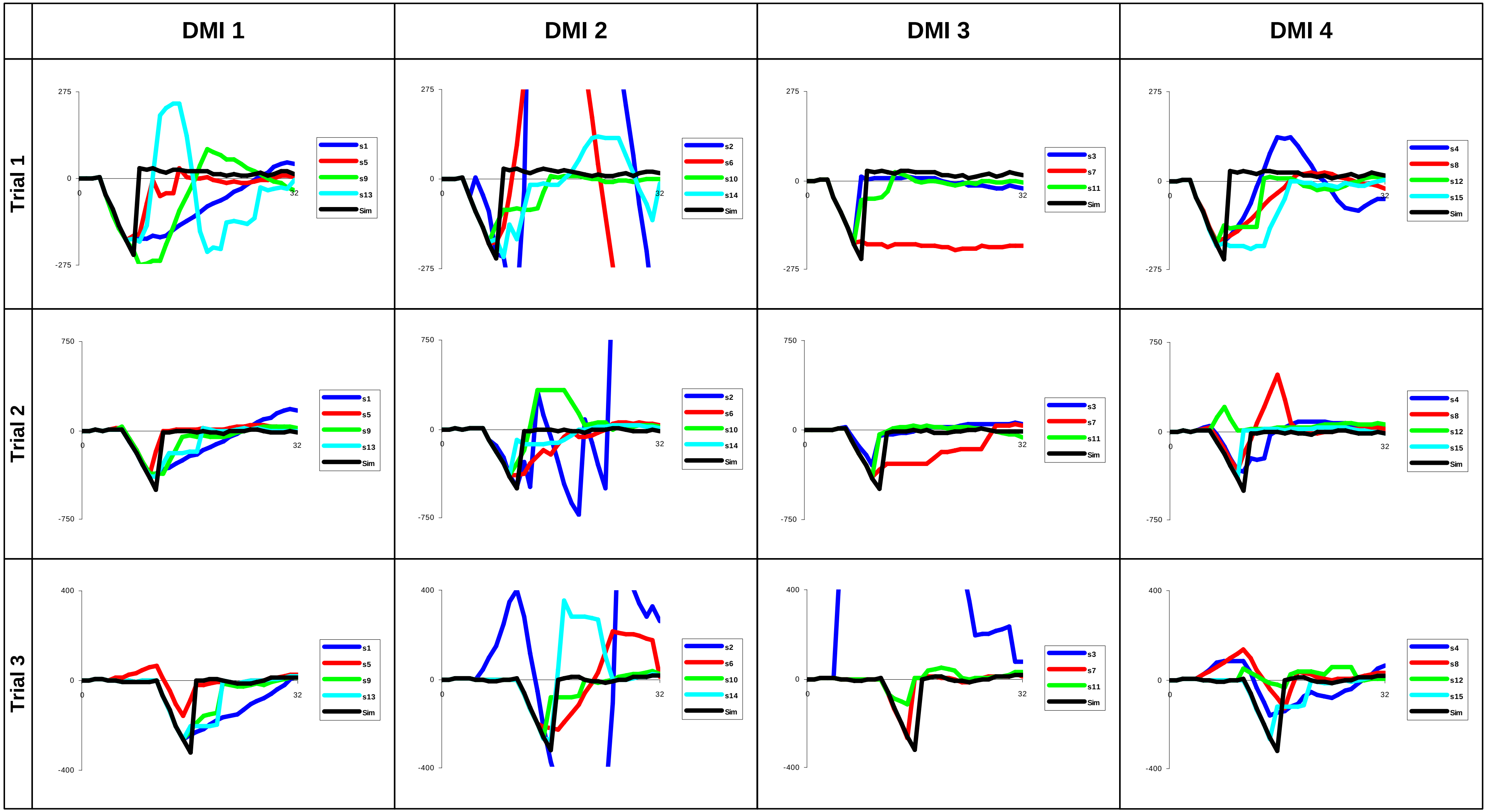


Figure 3 presents a matrix outlining the decision rules followed by the subjects. The two first rows indicate the type of decision rule applied by the subjects following the order increase. In Trial 1, 13 out of 14 subjects did not react to the order level change and set their production rate to approximate the observed customer order, i.e. PP*=CO*. In the course of the trial, 6 of the 13 subjects made an explicit effort to develop a decision rule that would allow them to decrease the inventory. In Trial 2, there are only 6 subjects who do not try to account for the delay, and in Trial 3, there are only 4. Eight out of 11 subjects, who managed to achieve a 0 inventory level at the end of the trial in Trial 3, developed a decision rule similar to this presented in equation (3) above.[15] However, as indicated in Figure 3 only 3 applied the rule properly at once (S5, S7, S8). The other 6 made a mistake, ordering the increased production level twice, misestimating the new order level, or accounting for too little periods with the insufficient production coverage. Among these 6 subjects, 2 subjects (S11, S12) had realised the mistake before the discrepancy between the production supply and customer demand became obvious through the persistent inventory deficit or surplus and used the same rule to adjust their production level further:

$$O_t = CO_t + \sum_{i=1..i} (ExpCO_i - PP_i)]$$

The other 3 subjects (S13, S10, S15) reacted only when it was obvious that there is a sustained inventory discrepancy.

Although our data allow for a detailed analysis of how the individual decision rules were developed such analysis is beyond the scope of the current discussion. What is important to note is that based on the think-aloud protocols, we find that by Trial 3, 10 out of 14 subjects developed and followed consciously the decision rule that allows to successfully control the system, see equation (2).

In Figure 3 we additionally indicate the subjects' self-evaluation following each trial. In the post-trial interviews, we also asked the subjects to assess both their understanding of the system and the perceived task difficulty. With exception of S5 and S13 after Trial 1, all subjects declared that they have understood the system entirely after all trials. The task, on the other hand, was perceived as increasingly easier over the trials, as illustrated in Figure 2.

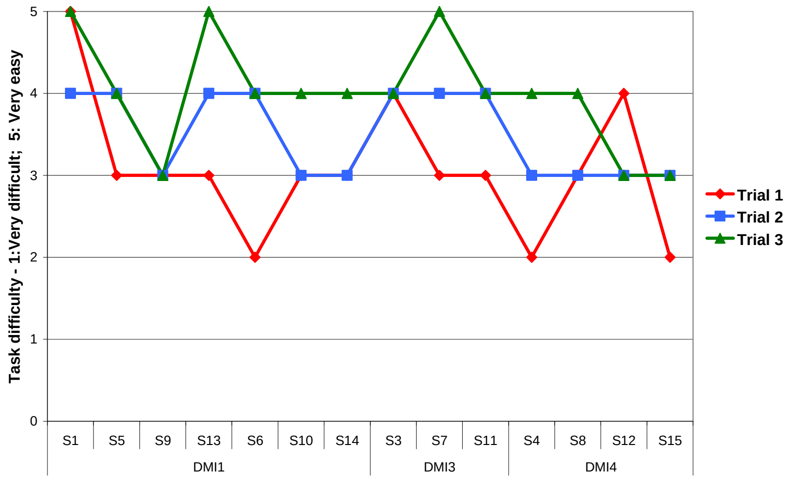


Figure 2 Perceived task difficulty over the three trials.

---

[15] The one subject who controlled INVENT successfully while not following the described decision rule misunderstood the instructions, and tried to achieve a zero inventory level only at the end of the trial rather than as soon as possible.

		TRIAL 1		TRIAL 2		TRIAL 3		*Successful trials*
DMI 1	S1	4		4			5	1
	S5		2		4^{x}		4^{x}	3
	S9	3		3			4	1
	S13		2		4		4^{x}	3
DMI 2	S6	2		4		4		0
	S10		2		2^{x}		3^{x}	3
	S14	2			3		4	2
DMI 3	S3		4^{x}		4^{x}	1^{x}		2
	S7	2			3^{x}		5^{x}	2
	S11		2		3^{x}		5^{x}	3
DMI 4	S4	2		3		3		0
	S8	2			3^{x}		5^{x}	2
	S12		2		2^{x}		2^{x}	3
	S15		2		4^{x}		2^{x}	3
TOTAL		7	7	4	10	3	11	
Final Inventory		$\neq 0$	≈ 0	$\neq 0$	≈ 0	$\neq 0$	≈ 0	
Initial decision rules								
PP*=CO*		7	6	4	2	2	3	
PP*=CO*+4*(CO*-PP) marked by 'x'		0	1	0	8	1	8	
Subsequent decision rules								
CO*+*nxExp*(Δ_Inv.)		0	4	0	8	1	6	
CO* + Δ_Inv		1	3	1	2	0	3	
CO*+/-Δ		6	0	3	0	2	2	

The numbers in the decision rules fields indicate the subject's self-evaluation of the trial:
1 – Very poor; 2 – Poor; 3 – Average; 4 – Good; 5 – Very good

Figure 3 Results matrix giving an overview of the decision making rules applied, the final level of the inventory and the subjects' self-evaluation of performance in each trial.

Finally, we present data illustrating the impact of the various DMIs. Table 4 shows the number of subjects for each DMI that used the decision rule stipulated by equation (2), see p. 14. For each DMI, the overall number of successful subjects is calculated. Most successful trials were conducted using DMI3 rather than DMI4 as initially expected (see Table 2, p. 11).

Table 4 Use of the simplified decision rule $O_t = MAX[0, CO_t + \sum_{i=1..4} (ExpCO_i - PP_i)]$

	Trial 1	*Trial 2*	*Trial 3*	$\sum$	
DMI1	2	2	2	**6**	*Average*
DMI2	0	1	2	**3**	*Poorest*
DMI3	2	3	3	**8**	*Best*
DMI4	0	3	3	**6**	*Average*
$\sum$	**4**	**9**	**10**		

Table 5 presents how many successful trials were observed for the diagrammatically-enhanced interfaces (DMI3 and DMI4) and the text-based interfaces (DMI1 and DMI2), see H_1, p. 10. While in case of diagrammatically-enhanced interfaces (DMI3 & DMI4), more trials were classified as successful, 67% as opposed to 43% for the text-based interfaces, the difference is not statistically significant, χ^2 (1, N=42)=2.40, p=0.121.

Table 5 Successful trials for text-based and diagrammatically-enhanced interfaces.

	Text-based interface [DMI1 & DMI2]	**Diagrammatically-enhanced interface** [DMI3 & DMI4]	$\sum$
Successful decision rule	**9**	**14**	21
Unsuccessful/no decision rule	**12**	**7**	21
Overall success rate	43%	67%	

Table 6 analyses whether an explicit representation of production-in-progress (DMI2 and DMI4) would result in more successful performance, see H_2, p. 10. Contrary to our expectations, explicit representation of production-in-progress did not lead to a better performance. Only 43% of trials with DMI2 and DMI 4 were successful as opposed to 67% observed in case of the other two interfaces. However, the difference between the two interface types is not statistically significant, χ^2 (1, N=42)=2.40, p=0.121.

Table 6 Successful trials for implicit and explicit representation of production-in-progress

	Implicit Production-in-Progress [DMI1 & DMI3]	**Explicit Production-in-Progress** [DMI2 & DMI4]	$\sum$
Successful decision rule	**14**	**9**	21
Unsuccessful/no decision rule	**7**	**12**	21
Overall success rate	67%	43%	

Discussion

The main purpose of this study was to determine whether particular ways of representing delays may prove to be more successful. We expected that diagrammatic representation of the system structure (see H_1, p. 10) and explicit visualisation of the delay effects (see H_2, p. 11) would result in an improved performance and understanding. Our results are rather weak and mixed. While the diagrammatically-enhanced interfaces indeed yielded more successful trials, the difference is not statistically significant (see Table 5, p. 18, and the associated discussion). On the other hand, contrary to our expectations, we observed more successful trials with the interfaces that did not explicitly visualized the production-in-progress. However, again, the result is not statistically significant (see Table 6, p. 18, and the associated discussion). The later observation is quite interesting, especially given that 4 out of 7 subjects who used either DMI2 or DMI4, providing an explicit production-in-progress representation, indicated that in their opinion the decision making interface could be improved by emphasizing the relationship between the planned and finished production and highlighting the production-in-progress volume. These comments suggest that measures used for representing the production-in-progress may not have been sufficient. Analysis of the think-aloud protocols and interview data shed also new light on the apparent positive impact of the diagrammatically-enhanced interfaces. As indicated above, numerical analysis of the performance logs suggests that, consistent with previous results (see e.g., Machuca et al. 1998, Howie et al. 2000), the diagrammatically-enhanced interfaces yielded more successful results. However, based on our observations and interview protocols, not many subjects using the diagrammatically-enhanced interfaces factually referred to the diagram: None of the 3 subjects using DMI3 commented on the advantage of having the structural diagram available; their comments suggest that they focused exclusively on the tabular representation of the system. Two of the subjects using DMI4 (S8 and S12) admitted that they 'did not notice' the diagram and relied on the tabular representation. Only one of the DMI4 subjects (S4) declared a preference for the diagrammatic representation. Given these data the advantage of the diagrammatically-enhanced interfaces, suggested by the numerical analysis (see Table 5, p. 18), is rather questionable.
The advantage of extended data provided by the single-data set is further illustrated by the insight we gained into the actual decision making process followed by our subjects. As indicated, while by Trial 3, 9 of 14 subjects used a successful rule for controlling the system (see Figure 3, p. 17), none of them followed the rule suggested as optimal by earlier studies (Sterman 1989). The rule developed by our subjects (see equation (2), p. 14) did not involve the concept of supply line, but rather tried to match the upcoming inventory deficits. Furthermore, it is important to note that the rule is not robust and works only for the one-time step increase in the customer orders. As such, it could be deemed as suboptimal. However, given that the subjects were asked to respond to a specific customer order change, it is plausible to accept the rule as appropriate. Further, when comparing the cumulative costs, the subjects seem to outperform the anchoring and adjustment heuristic – their cumulative costs are lower than those generated when the heuristic-based decisions are simulated (see Appendix 4).
Consistent with previous results (see e.g., Sterman 1989, Diehl 1992, Diehl and Sterman 1995), our experimental data suggest that people do have problems dealing with the delays. However, according to our observations most of them learn relatively quickly –

by Trial 2 most of the subjects were able to control the system (see Table 3, p. 15). It is interesting to note that although the subjects by and large provided correct answers to the apprehension tests administered prior the task, during Trial 1 all but one[16] were surprised and confused by the effects the manufacturing delay had on their inventories. Following Trial 1, all but two[17] subjects claimed to have understood the system entirely. At the same time, most felt that the task was difficult or not easy (see Figure 2, p. 16). These results are interesting in several respects.

First, one could expect that a full understanding of the system would yield the fully robust decision rule. This is not confirmed by our results. Successful subjects in our experiment followed a suboptimal, locally successful rule. Development of such simplified rule may be seen as manifestation of fast-and-frugal heuristic (Todd 2001). It would be interesting to test whether the subjects would be able to adopt their decision rule appropriately, given a different customer order flow. It is possible that the claim of 'complete' understanding is exaggerated, and that the subjects, in reality, understand well only the particular case featured by the experimental task. Indeed, such narrow, case-specific understanding is common in complex and dynamic systems and underlies most of the problems with management of these systems (Forrester 1961, Sterman 2000).

Second, it is interesting to note that not only the successful, but also the unsuccessful subjects, who responded intuitively, following the instantaneous inventory deficits/surpluses, claimed to have understood the system entirely. The unsuccessful subjects were also quite satisfied with their performance. Indeed, by Trial 3 they evaluated their performance higher than the successful subjects did, see Figure 4. These results are worrying as they suggest that people, who actually mismanage dynamic systems, are not likely to detect the problems, believing that they not only understand the situation, but also manage it quite well.

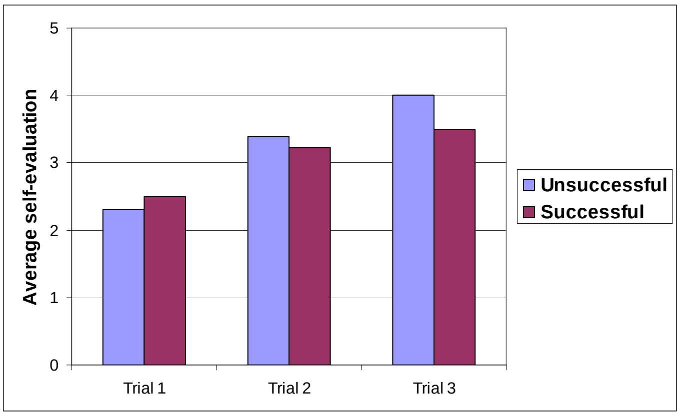


Figure 4 Average performance self-evaluation: 1 – Very poor; 2 – Poor; 3 – Average; 4 – Good; 5 – Very good

Finally, the discrepancy between the good level of understanding detected by the prior-test and the subsequent poor performance in Trial 1 suggests that a descriptive, propositional knowledge of a dynamic system is not a sufficient guarantee for a successful performance. It is encouraging that most of the subjects seemed to realise these dichotomy: following Trial 1, most of them reported to have understood the

---

[16] Subject S3, see DMI3, Table 3, p. 15.

[17] Subjects S5 and S13, see DMI1, Table 3, p. 15.

system entirely, but evaluated the task as not easy (8 subjects) or difficult (3 subjects).[18] Additionally, in the follow-up interviews, when asked about ways to improve the instructions, 4 subjects suggested that a worked-out example or a short training session would be useful. These results suggest that a descriptive presentation of a dynamic problem is not likely to be entirely effective and that a hands-on experience is vital in order to acquire full understanding and command of a dynamic system. This is consistent with earlier results reported by Sawicka, Qian, and Gonzalez (2005).

Further research directions

72% of our subjects developed a successful decision rule for controlling the simple inventory system, albeit only 4 out of 14 succeeded in the first trial, and only one responded appropriately at once. An interesting finding is that all the successful subjects developed independently the same rule, effective in the particular context of the INVENT experiment task, but not robust in general. Further research should be conducted to determine whether the subjects would be able to adjust the rule given a different customer order flow. It would also be interesting to investigate what type of decision rules would be developed if the subjects would not be given any pre-defined customer order flow. Would such task reformulation lead to development of more robust rules?

Results of our experiments suggest that, to control a dynamic system effectively, most people would need, beyond the textual description of the system, some practice. Computer simulators provide a promising tool for this purpose and should be a natural addition to various training programs, both for students and professionals. Still, their use in this context is far from wide spread. One possible reason for this may be that not all simulators seem to be equally successful in promoting a deeper understanding of a dynamic system. An effort is needed to identify what are the features of successful simulators. The issue of representing and communicating the challenging aspects of dynamic systems seems to be vital. Our study investigated different ways of incorporating one of such aspects, namely delays, into the decision making interface. Current results do not allow us to conclude which representation of the delay is most successful. Further research is needed. We would advocate that future experimental investigations are conducted using a single-subject rather than a group design. As demonstrated by our study, the single-subject design allows for a much deeper and more accurate understanding of the decision making process. Such precise understanding is vital especially when the goal is to identify followed decision rules or the elements of the decision-making environment that are likely to impact the decision-making process and understanding.

[18] Only 3 subjects felt that the task was easy (S3, S12) or very easy (S1), see Figure 2, p. 16.

Appendix 1

Below is an English translation of the instructions used for the experiments:

EFFICIENT PRODUCTION MANAGEMENT
A Simulation-Based Challenge

Active practice has always been an important part of the business education. Apprenticeship – during which business students apply theoretical knowledge and gain practical experience – is one of the oldest ways to master business skills. While this form of education provides excellent opportunities for learning, it also poses some challenges. There is a conflict between the need to charge the new apprentice with a challenging task and the fact that this type of tasks – if wrong decisions are taken – may have damaging consequences for the company.

In this context, business simulators provide a very attractive alternative. Such simulators mimic a business reality and allow to charge the business students with even most challenging tasks in the 'safe' environment, without risking a damage to the company in case of students' failures.

While the business simulators may facilitate testing the students' knowledge and abilities, they also play an important role in the business education. They allow the students to test and enhance their skills in the settings that mimic the business reality. The business simulators are typically far less complex than the real life situation. These simplifications are necessary to create a more effective learning environment. Still, it is important to remember that the key aspects of a complex business reality are maintained so that the main mechanisms operating in the simulated environment are analogous to these that operate in the real settings.

INVENT application is a simulation-based tool for mastering your business skills. On the following pages of this booklet you will find a short description outlining a business scenario, your role and task. In the course of the challenge you will use *INVENT* to manage the business situation.

INTRODUCTION

The ABC Manufacturing Co. is a European producer of car components. The company was established in 1980 and until 2004 has operated in 4 different countries of the European Economic Area: Germany, France, Spain and Norway. In each country the company has a manufacturing plant and the associated warehouse that dispatches products to the local customers.

In the light of the 2004 EU enlargement, the management board of ABC Co. decided to enter the market of the new EU member states. Initial negotiations with Polish dealers have been recently completed and the contract will be signed soon. Once the contract is signed, the first orders from Poland are expected. The responsibility for fulfilling these orders will rest with the ABC's branch located in Eastern Germany.

YOUR ROLE

You are production manager of the East German branch of the ABC Manufacturing Co.

As a production manager you are responsible for setting the weekly production schedule. Typically, such a decision depends on a variety of factors, including production capacity, raw material availability or inventory levels. In this scenario, we assume no limits on the production capacity or raw material. This means that you can acquire as many components and manufacture as many final products as you wish. The production cycle takes always exactly 4 weeks.

However, you must remember that storing the finished products in the warehouse generates costs. Thus ABC applies just-in-time production policy. It manages the production process in such way to meet the customer orders simultaneously reducing the inventory level down to zero. Given a high competitiveness of the market, the orders not fulfilled in the week they are placed will be lost, which is also treated as cost.

Another source of expenses the ABC may incur is production changes. Imagine the customer orders increases and ABC has to hire and train new employees, and buy new equipment in order to fulfill all client requirements. If the demand for ABC's products decreases the company will have to dismiss its workers. All these actions in both cases generate costs.

THE CURRENT SITUATION & THE CHALLENGE

Your task involves setting the weekly production at the East German branch of the ABC Manufacturing Co. You will manage the production line for a period of 32 weeks. During this time, the branch is likely to experience a change in the rate of customer orders due to the new contract with the Polish dealer.

Your main goal is to **minimize cumulative costs** arising from **off-targeted inventory** levels and **production changes**. You need to know the exact nature of these costs:

$\text{Cost}_t = (|\text{Inventory}_t| * €1 + |\text{Production Rate Change}_t| * €1)$
where:

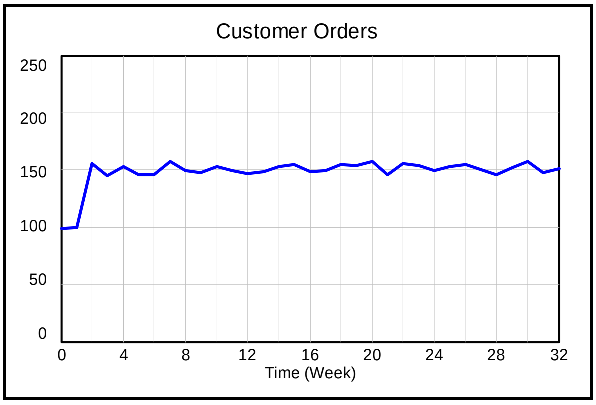
t – the current week,
$|Inventory_t|$ - Inventory Shortage or Surplus in time *t*,
$|Production\ Rate\ Change_t|$ - the difference between Production Start Rate in time *t* and the previous Production Start Rate (in time *t-1*)

The simulation-based environment will provide you with the basic information about the current level of your inventory, the status of the production, and the customer order rate.

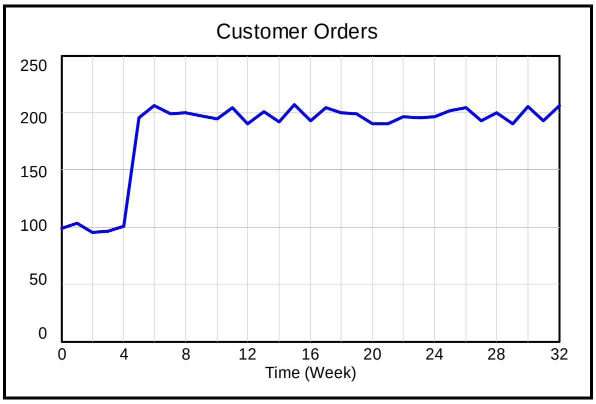
***

The instructor will now lead you through some trial decisions so that you get better acquainted with the interface and game rules. Please feel free to ask any question.

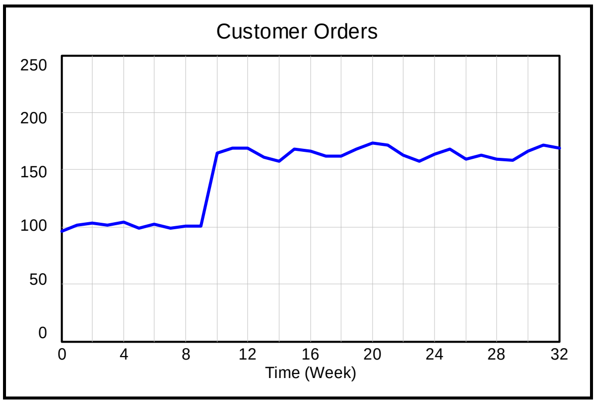
Appendix 2



Time Series 1 Customer Orders series used as input for Trial 1: Increase in the customer orders to 150+/-5% [widgets/week] in week 2 (random seed = 1)



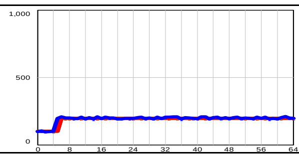
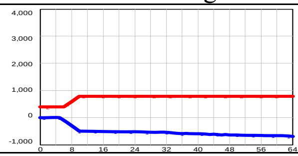
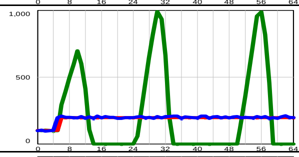
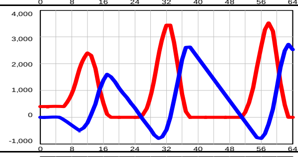
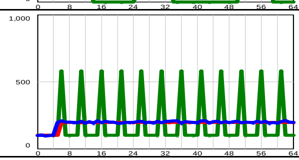
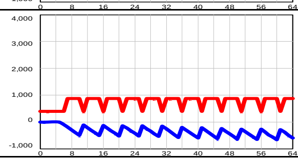
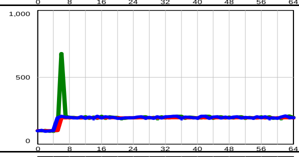
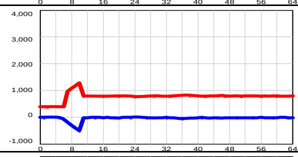
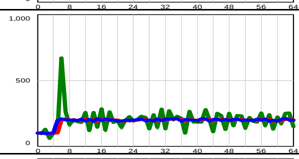
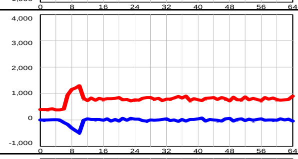
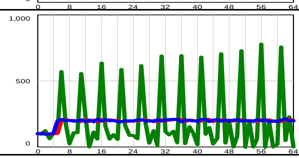
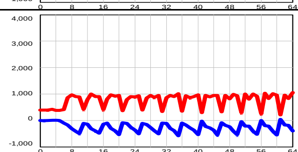
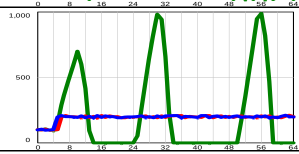
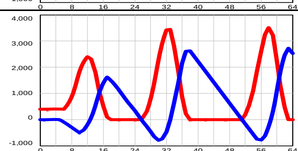
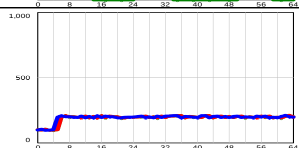
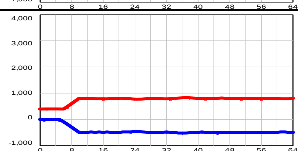
Time Series 2 Customer Orders series used as input for Trial 2: Increase in the customer orders to 200+/-5% [widgets/week] in week 5 (random seed = 2)



Time Series 3 Customer Orders series used as input for Trial 3: Increase in the customer orders to 165+/-5% [widgets/week] in week 10 (random seed = 3)

Appendix 3

Costs depending on attention given to the various decision rule elements: **ECOAT** (*Expected Customer Orders Adjustment Time*) indicates time necessary for the subjects to adjust the level of expected customer orders, **IAT** (*Inventory Adjustment Time*) – the time to adjust for the inventory deviations, **PIPAT** (*Production-in-Progress Adjustment Time*) – the time to adjust for the work-in-progress. Customer Orders are at the level of 100 widgets/wk +/-5% for the first 4 wks; in wk 5, *Customer Orders* increase to 200 widgets/wk +/-5% (see Time Series 2 in Appendix 1). Costs are a linear function of the absolute inventory deviations and production changes (see also footnote 3, p. 3, for more details).

	ECOAT*	IAT	PIPAT	Cumulative Costs	Simulation runs ** Customer Orders & Production Start Rate ***	Inventory & Production-in-Progress ****
A	∞	∞	∞	13 169		
B	∞	1	∞	19 307		
C	∞	∞	1	14 323		
D	∞	1	1	3 036		
E	1	1	1	5 178		
F	1	∞	1	14 800		
G	1	1	∞	19 432		
H	1	∞	∞	12 649		

* *Expected Customer Orders Adjustment Rate* (see Figure 1) is defined as an IF function that ensures that even when ECOAT=∞, the *Expected Customer Orders* level is adjusted instantaneously in response to large changes in the *Customer Orders*. This is implemented because the instructions clearly indicate that a step increase in the customer orders should be expected.

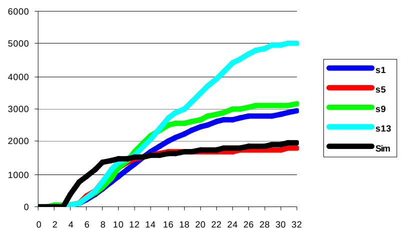
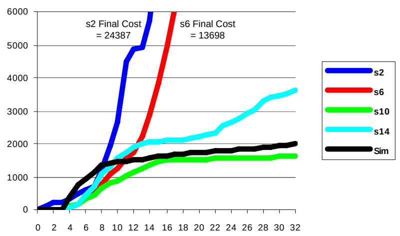
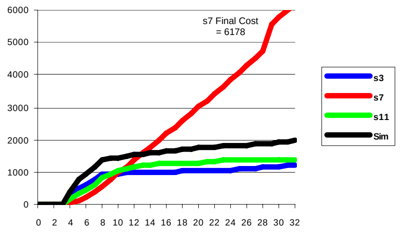
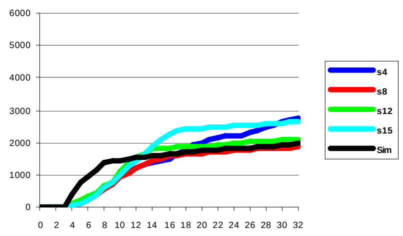
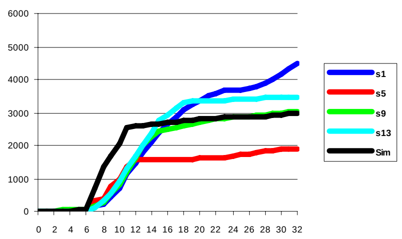
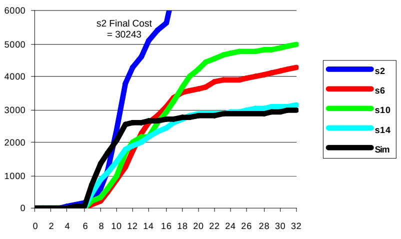
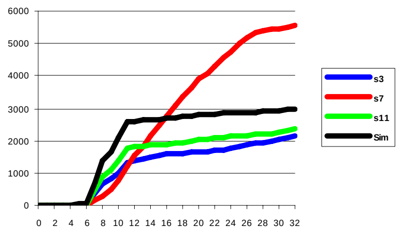
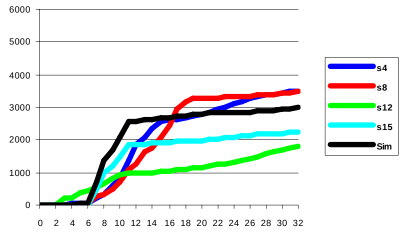
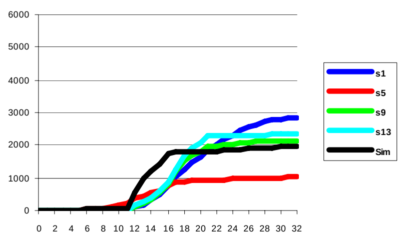
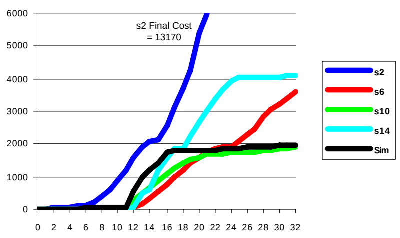
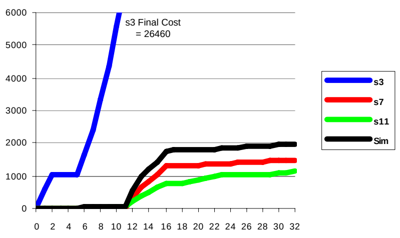
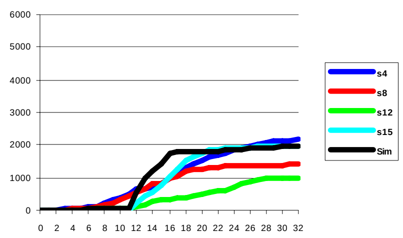
** To obtain a better illustration of the type of behavior that arises in each situation we have doubled the simulation period to *t=64*. The costs are given for the regular trial duration, i.e., *t=32*

*** ❶: Customer Orders, ❷: Expected Customer Orders, ❸: Production Start Rate

**** ❶: Inventory, ❷: Production-in-Progress

Appendix 4

Cummulative costs for all the subjects in all trials.

<table>
<tr><th></th><th>DMI 1</th><th>DMI 2</th><th>DMI 3</th><th>DMI 4</th></tr>
<tr><td>Trial 1</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>Trial 2</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>Trial 3</td>
<td>
</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

Bibliography

Barlow, Daniel H., and Michel Hersen. 1984. *Single Case Experimental Designs: Strategies for Studying Behavior Change*. 2nd ed. New York: Pergamon Press.

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

Brehmer, Brendt. 1992. Dynamic decision making: Human control of complex systems. *Acta Psychologica* 81:211-241.

Chandler, Paul, and John Sweller. 1991. Cognitive Load Theory and the Format of Instruction. In *Cognition & Instruction*: Lawrence Erlbaum Associates.

Cooper, John O., Timothy E. Heron, and William I. Heward. 1987. *Applied Behavior Analysis*. Columbus, OH: Merrill Publishing Company.

Diehl, Ernst, and John D. Sterman. 1995. Effects of Feedback Complexity on Dynamic Decision Making. *Organizational Behavior and Human Decision Processes* 62 (2):198-215.

Diehl, Ernst Walter. 1992. Effects of Feedback Structure on Dynamic Decision Making Sloan School of Management, Massachusetts Institute of Technology, Boston, MA.

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

———. 1989. *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen*. Reinbeck: Rowohlt.

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

Elashoff, Janet D., and Carl E. Thoresen. 1978. Choosing a statistical method for analysis of an intensive experiment. In *Single Subject Research: Strategies for Evaluating Change*, edited by T. R. Kratochwill. London, UK: Academic Press.

Forrester, Jay W. 1961. *Industrial Dynamics*. Cambridge, MA: MIT Press.

Großler, Andreas, and Frank 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.

Großler, Andreas, Frank H. Maier, and Peter M. Milling. 2000. Enhancing learning capabilities by providing transparency in business simulators *Simulation & Gaming* 31 (2):257-278.

Howie, E., S. Sy, L. Ford, and Kim J. Vicente. 2000. Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review* 16:151-171.

Jensen, Eva, and Brendt Brehmer. 2003. Understanding and control of a simple dynamic system. *System Dynamics Review* 19 (2):119-137.

Larkin, Jill H., and Herbert A. Simon. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11:65-99.

Machuca, Jose A. D. 1992. Are we losing one of the best features of system dynamics? *System Dynamics Review* 8 (2):175-177.

———. 2000. Transparent-box business simulators: An aid to manage the complexity of organizations. *Simulation & Gaming* 31 (2):230-239.

Machuca, Jose A. D., J. C. Ruiz, M. A. Domingo, and Maria M. Gonzalez. 1998. Our 10 years of work on transparent-box business simulation. Paper read at 16th International Conference of the System Dynamics Society, at Quebec, Canada.

Moxnes, Erling. 1998. Overexploitation of renewable resources: The role of misperceptions. *Journal of Economic Behavior & Organization* 37 (1):107-127.

———. 2004. Misperceptions of basic dynamics: the case of renewable resource management. *System Dynamics Review* 20 (2):139-162.

Richardson, George P., and J. Rohrbaugh. 1990. Decision making in dynamic environments: Exploring judgments in a system dynamics model-based game. In *Contemporary Issues in Decision Making*, edited by K. Borcherding, O. I. Larichev and D. M. Messick. Amsterdam, Netherlands: North-Holland.

Sawicka, Agata. 2005. Cognitive load dynamics: How to increase effectiveness of SD-based learning environments. Paper read at Submitted to the 23rd International Conference of the System Dynamics Society, at Boston, MA.

Sawicka, Agata, Ying Qian, and Jose Gonzalez. 2005. Managing CSIRT capacity as a renewable resource management challenge: An experimental study. Paper read at Submitted to the 23rd International Conference of the System Dynamics Society, at Boston, MA.

Sengupta, Kishore, and Tarek K. Abdel-Hamid. 1993. Alternative Conceptions of Feedback in Dynamic Decision Environments: An Experimental Investigation. *Management Science* 39 (4):411.

Shaughnessy, John J., and Eugene B. Zechmeister, eds. 1994. *Research Methods in Psychology*. Singapore: McGraw-Hill, Inc.

Sidman, M. 1960. *Tactics of Sceintific Research: Evaluating Experimental Data in Psychology*. New York: Basic Books.

Sterman, John D. 1987. Testing behavioral simulation models by direct experiment. *Management Science* 33 (2):1572-1592.

———. 1989. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes* 43 (3):301-335.

———. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decisin making experiment. *Management Science* 35 (3):321.

———. 1994. Learning in and about complex systems. *System Dynamics Review* 10 (2-3):291-330.

———. 2000. *Business Dynamics. Systems thinking and modeling for a complex world.*: McGraw-Hill.

Sweller, John. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12:257-285.

Todd, Peter M. 2001. Fast and Frugal Heuristics for Environmentally Bounded Minds. In *Bounded Rationality:The Adaptive Toolbox*, edited by G. Gigerenzer and R. Selten. Cambridge, MA: The MIT Press.

Vicente, Kim J. 1996. Improving dynamic decision making in complex systems through ecological interface design: a research overview. *System Dynamics Review* 12:251-279.

Zamora Gonzalez, Maria M., Jose A. D. Machuca, and Jose Carlos Ruiz del Castillo. 2000. SITMECOM 1.0 PC: A transparent-box multifunctional simulator of competing companies. *Simulation & Gaming* 31:240-256.