To appear on European Journal of Operational Research

On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems

Raimo P. Hämäläinen[a](corresponding author), Jukka Luoma[b] and Esa Saarinen[c]

[a] Systems Analysis Laboratory, Department of Mathematics and Systems Analysis, Aalto University School of Science, P.O. Box 11100, 00076 Aalto, Finland

[b] Department of Marketing, Aalto University School of Business

[c] Work Psychology and Leadership, Department of Industrial Engineering and Management, Aalto University School of Science

Email addresses: raimo.hamalainen@aalto.fi, jukka.luoma@aalto.fi, esa.saarinen@aalto.fi

Abstract

We point out the need for Behavioral Operational Research (BOR) in advancing the practice of OR. So far, in OR behavioral phenomena have been acknowledged only in behavioral decision theory but behavioral issues are always present when supporting human problem solving by modeling. Behavioral effects can relate to the group interaction and communication when facilitating with OR models as well as to the possibility of procedural mistakes and cognitive biases. As an illustrative example we use well known system dynamics studies related to the understanding of accumulation. We show that one gets completely opposite results depending on the way the phenomenon is described and how the questions are phased and graphs used. The results suggest that OR processes are highly sensitive to various behavioral effects. As a result, we need to pay attention to the way we communicate about models as they are being increasingly used in addressing important problems like climate change.

Introduction

This paper aims at pointing out the importance of behavioral operational research (BOR), defined as the study of behavioral aspects related to the use of operational research (OR) methods in modeling, problem solving and decision support. In operational research the goal is to help people in problem solving but somehow we seem to have omitted the individuals, the problem owners and the OR experts, who are engaged in the process, from the picture. There is a long tradition of discussing best practices in OR (see, e.g., Corbett et al. 1995; Miser, 2001) but it is surprising to note that behavioral research on the process itself and on the role of the analyst and problem owner has been almost completely ignored. Descriptions of case studies are not enough. We also need controlled comparative studies and experiments. We argue that by paying more attention to the analysis of the behavioral human factors related to the use of modeling in problem solving it is possible to integrate the insights of different approaches to improve the OR-practice of model-based problem solving.

OR is used to facilitate thinking and problem solving. How OR methods achieve this is one avenue of inquiry that has not received much attention. What kinds of behavioral biases do OR methods themselves cause or solve is another. Ultimately, problem solving is related to decision-making and behavioral OR aims at helping to make better decisions. There is a strong tradition in research on behavioral decision making but related work on the OR processes in general is still missing. It is interesting to note, however, that in operations management (OM), which uses OR methods to improve operations, the term Behavioral Operations (BO) has already been used for some years (see Bendoly et al., 2006; Loch and Wu, 2007; Gino and Pisano, 2008). First the main interest was in heuristics and biases in decision making but later there has also been interest in system dynamics in operations (Bendoly et al., 2010). Research on judgmental forecasting is also related to BO (see, e.g., Kremer et al, 2011)

In general, behavioral issues in decision making have been widely studied at the individual-, group-, and organizational levels by researchers in judgment and decision making, cognitive psychology, organization theory, game theory and economics. In the field of OR the areas of risk and decision analysis as well as multicriteria decision making have paid attention to behavioral issues rigorously (see, e.g., Winterfeldt and Edwards, 1986; Korhonen and Wallenius, 1996; French et al., 2009; Morton and Fasolo, 2009). Besides the area of decision making little is known about behavioral issues in model-based problem solving. OR is about real-life problem solving and thus, it is in general subject to behavioral issues and effects. We see that it is increasingly important to pay attention to this core area of our discipline. Research on behavioral OR is needed to complement traditional operational research that is limited to developing mathematical methods and optimization techniques. The purpose of behavioral studies is to make better use of OR models and to warn of the behavioral issues that need to be taken into account when using models to support decision making or making predictions.

The well-known results of psychological decision research (see e.g. Kahneman and Tversky, 2000) have already been recognized in models of decision support. However, there is more that we can learn from psychological research. We OR people are likely to have similar cognitive biases in model building as people have in other contexts. Here we can list a few of them. Confirmation bias refers to the phenomenon where people are more responsive to information that confirms rather than challenges their assumptions and in our context our models (Sterman, 2002). People also prefer intuitive alternatives over non-intuitive ones, which is likely to cause OR analysts to favor techniques they have grown to understand intuitively. This can prevent healthy criticism towards models (Simmons & Nelson, 2006). Kahneman (2003a, 2003b) noted that people tend to use simple attribute substitution heuristics to attribute causality. The implication for OR is that successful OR interventions can be erroneously attributed to the model while failure of OR interventions can be attributed to other factors. OR problem solving often involves a team of experts and related group-level processes can generate biased models as well through, for example, the well-known group think mechanism (Janis, 1972). There is also an interesting interplay between model complexity and bounded rationality (Simon, 2000). While models may help overcome some human cognitive capacity limitations (e.g., Simon, 1997), overly complex models may lead to information overloading and thereby incapacity to understand the model mechanics (Starbuck, 2006), which prohibits learning, may cause modeling mistakes, and may distort the communication of the results to the clients. The study of behavioral biases and problems in modeling is a key area of research in behavioral OR.

It is quite surprising that, so far, these issues have received hardly any attention in the OR literature. In the early OR literature there was an active interest in model validity. See, for example, the special issue of *European Journal of Operational Research* in 1983 (Landry and Oral, 1993). At the same time there was related discussion on model validation in system dynamics (Barlas, 1996). These analyses were usually based on the hidden assumption that a valid model will also produce a valid process. This has lead to ignoring the interaction of the analyst and problem owner as well as all the possible behavioral effects. More recently, the index of the *Encyclopedia of Operations Research and Management Science* (Gass and Harris, 2001), for example, does not include topics such as behavioral, biases or heuristics. In the recent special issue of the *Journal of the Operational Research Society* celebrating fifty years of OR, behavioral studies are passed by minor comments and references in the articles by Brailsford et al. (2009), Jackson (2009) and Royston (2009). A notable exception in the early OR literature is the article by Richels (1981). He recommends "user analysis" as a way of improving model-based problem solving. User analysis tries to identify whether and how a model helps decision-makers to focus on the relevant information and generate new insights. It extends the validation and verification and sensitivity analysis activities, which are traditionally considered to demonstrate that a model is "good". Behavioral OR is characteristically a form of user analysis, which focuses on the psychological and social interaction-related aspects of model use in problem solving. There are also more recent signs of positive development. There has been increasing interest in research on group model building and facilitated modeling (French et al., 1998; Rouwette and Vennix, 2006; Franco and Montibeller, 2010; Franco and Rouwette, 2011). These studies have an explicit emphasis on the social aspects and the facilitator skills and styles.

Behavioral issues also relate to the important topic of OR and ethics. So far, discourse about the ethics in OR has assumed that the OR practitioner has cognitive control over all the issues related to ethical matters (Le Menestrel and Van Wassenhove, 2004, see also the special issues of *OMEGA - The International Journal of Management Science*, Le Menestrel and Van Wassenhove, 2009 and *International Transactions in Operational Research,* Wenstop, 2010). Yet there are hidden processes and factors that can also have impacts. The ethical OR practitioner should keep in mind and be aware of the possibility of such unintentional behavioral phenomena (Brockelsby, 2009). The personalities and communication style and perhaps even gender and cultural background of the analyst and the client can play a very big role in important OR interventions (Doktor and Hamilton, 1973; Lu et al., 2001). It is also the responsibility of the OR practitioner to be aware of the possible misunderstandings and communication errors arising from the behavioral and social processes (see, e.g., Drake et al., 2011). For example, models may mask ethical problems by appealing to the authority of OR experts and established conventions (Dahl, 1957). Likewise, the phenomenon of cognitive dissonance may lead the analyst to too soon want to believe that a model captures the essentials of the phenomenon. This may lead to the exclusion of important features from the model (Cooper, 2007). This leads to the natural question of when is a model good enough and how can we know this.

From a behavioral perspective, when choosing a mix of OR methods, the analyst faces a tradeoff between internal fit and external fit (Mingers and Brockelsby, 1997). On the one hand, the choice of methods has to be made on the basis of the skills, knowledge, personal style and experience of the analyst (Ormerod, 2008). The complexity of models, for example, should be aligned with the cognitive capacity of the OR analysts and his or her client (cf., Simon, 1996). On the other hand, by maximizing internal fit and by staying in the comfort zone, there is a risk of being trapped by a tunnel vision where the specific modeling skills or competences of the expert intentionally or unintentionally create a situation where one approach is seen as the only way to look at the problem. Here, too, the psychological phenomena related to cognitive dissonance become relevant. When the modeler realizes that he/she does not have the skills needed to solve a problem the problem could be ignored as uninteresting. For a summary and interesting readings about cognitive dissonance, see Cooper (2007) and Travis and Aronson (2007).

The importance of behavioral OR is underlined by the fact that today model-based problem solving is an increasingly important approach used when tackling problems of high importance. Issues related to climate change and of natural resource management are examples in which the use of different quantitative and qualitative models is an essential part of the public policy process. We see that the inclusion of behavioral considerations in the practice of OR is a necessity when trying to leverage the benefits and promises of the OR approach[1].

Behavioral research in our sister disciplines, game theory, economics and finance, is already very strong (Camerer et al., 2004, Ackert and Daeves, 2009). It seems that interest in behavioral issues emerges when the basic theoretical core of the research field has matured enough. Economics is a good example. Behavioral economics has become an established topic acknowledged also by theoretical economists. Behavioral research in finance is very active as well (Barberis and Thaler, 2004). Embracing the behavioral perspective helps "generating theoretical insights, making better predictions, and suggesting better policy" (Camerer et al., 2004). If this is true for economics it surely applies to OR as well. Just like economic behavior, OR is a process that involves actual people. OR consists of solving real problems using quantitative and qualitative OR methods. Even if we can prove that there is an optimal policy or decision based on the model of the real world, the process by which the real world is simplified into a model is a process subject to behavioral affects (Lenvinthal, 2011).

Systems thinking (Jackson, 2009), soft OR and the problem structuring methods (PSM) communities have for long criticized OR for being too narrowly concerned with mathematical models only. Extending the OR toolkit, systems thinkers have drawn attention to the sociology and philosophy of modeling and problem solving, (Churchman, 1968; Ulrich, 1994; Brocklesby and Cummings, 1996; Keys 1997; Mingers, 2003; Ormerod, 2008). Soft OR has investigated the possibility of using qualitative methods, including subjective beliefs and values to support decision making (Checkland, 2000; Eden and Ackerman, 2006; Mingers 2011). Problem structuring methods are used as a soft front end of the OR modeling process (Mingers et al., 2009). Cognitive mapping is one popular approach used (Eden, 2004) Yet systems thinking and soft OR, like traditional OR, have remained mainly methodology and tool focused. The criticism of OR for ignoring people issues, has not led to stronger interest in assessing the degree of behavioral realism in the new processes prescribed for model-based problem solving. There are only comparative reports and reviews of observations in different case studies (Mingers and Rosenhead, 2004). To our knowledge the first experimental study of structuring approaches and how they compare with each other has only appeared quite recently (Rouwette et al., 2010). In this experiment the test subjects were

[1] Cf. www.scienceofbetter.org

members of the research organization not real clients. One natural explanation for the lack of behavioral research is the fact that comparative experimental research on problem solving and structuring is very challenging as real clients and real problems can seldom be approached repeatedly using different approaches.

One particular stream of behavioral research that is relevant to model based problem solving has been carried out by researchers in the area of system dynamics (SD). Researchers in this area have analyzed how people understand and make decisions regarding dynamic systems (Forrester, 1958; Sterman, 1989a; Moxnes 2009). Today, mankind is facing climate change as one of the key challenges where understanding decision making in dynamic contexts is essential. Related to this, Sterman (2008) argued in *Science* that the human cognitive biases related to understanding dynamic systems is a major cause of current confusion and controversy around the climate change policy issues. This research on the behavioral issues related to dynamic systems has not received much attention in the OR literature even though some of the seminal papers where published in *Management Science* (e.g., Kleinmunz, 1985 and Sterman, 1987). Studying how people understand and make decisions regarding complex systems is indeed highly relevant to OR in general. Many OR problems are about simulating and optimizing, building understanding of and managing and communicating about dynamic systems. Examples range from the early inventory problems to supply chain management (Silva et al., 2009) and climate policy design (Bell et al., 2003). By the analysis presented later in this paper we hope to attract the interest of the OR community to behavioral research by re-examining a behavioral topic studied actively in the SD community.

Topics for a research agenda in behavioral OR

Human behavior moderates each stage of the OR process (cf. Ravindran et al., 1976) and mediates the progression through stages. The client as well as the analyst is subject to behavioral effects. The personality characteristics of optimism and pessimism have been shown to influence DSS use (Korhonen et al., 2008). Consider for instance the system dynamics modeling process which according to Sterman (2000) consists of problem articulation, constructing dynamic hypotheses, model formulation, model testing, and policy formulation and evaluation (simulation). Problem articulation is susceptible to framing effects (Levin et al., 1998). Constructing dynamic hypotheses may be influenced by attribution errors (Repenning and Sterman, 2002). Model formulation and testing is subject to confirmation biases (Nickerson, 1998). Moreover communicating model results in graphical form may give rise to psychological misinterpretations both by the analyst or the client.

In the illustrative example of the present paper, we adopted an experimental research approach to study the process of communication with and about models. The illustration shows only one way of doing behavioral OR. Other research approaches could be adopted as well. Possibilities for different conceptual or theoretical perspectives include cognitive psychology, organizational behavior and sociological theories. On the methodological front, possibilities include experimental set ups, case studies, comparative case studies, ethnography and surveys. Currently, existing research on the practice of OR uses, typically, case-based reasoning. Few notable exceptions have used interviewing (Corbett et al. 1995), surveys (Rouwette et al. 2009) and group communication analysis (Franco and Rouwette 2011). Beyond the different ways of studying the OR process and practice, the list of OR challenges that could be addressed is quite large as well. Topics that have been studied include facilitated problem structuring (French et al. 1998; Franco and Rouwette 2011) and teaching OR (Grenander 1965). Other topics for behavioral research include communication with and about models (addressed in this study), ethical use of OR and non-expert use of OR methods.

Taking these observations into account the following very diverse list of topics in which behavioral research is clearly needed shows the richness and breadth of the area. Hopefully it will also stimulate interest in BOR by pointing out new research opportunities.

1. **Model building:** Framing of problems; Definition of system boundaries; Reference dependence in model building; Anchoring effect in selecting model scale and reference point; Prospect theory related phenomena in choosing the sign (increasing/decreasing) of variables; Learning in modeling. Satisficing in modeling.

2. **Communication with and about models:** Effects of graphs and scales used; Effects of visual representation of system models (e.g., comparison of system dynamics diagrams versus traditional simulation and systems engineering diagrams); Effects of education and cultural background of problem owners.

3. **Behavioral and cognitive aspects:** How do biases observed in decision making transfer to OR modeling in general; Mental models; Risk attitudes; Cognitive overloading; Oversimplification; Cognitive dissonance; Overconfidence; Effects related to human-computer interaction.

4. **Personal, social and group processes in OR facilitation:** Group processes in model building and facilitation; Social interaction styles and personality; Implicit goals and hidden agendas in modeling (e.g., omission of variables), selection of the scale; Strategic behavior by analyst and stakeholders; Gender differences; Cultural effects; Face to face vs. online interaction.
5. **People in problem solving situations:** How to help people find better strategies; What are the criteria used in problem solving – optimization/satisficing; Learning and belief creation; People in the loop models; Heuristics in model use; Bounded rationality in modeling; Path dependence in problem solving; People behavior in queuing and waiting for service; Crowd behavior in emergency situations.
6. **Comparative analysis of procedures and best practices:** Comparison of different modeling and facilitation approaches; Usefulness of simple versus complex models.
7. **Teaching of OR:** Best practices; Role of software; Developing facilitation and systems intelligence skills.
8. **Ethics and OR:** Behavioral challenges in ethical OR; Unintentional biases in model use; Self leadership skills in OR practice.
9. **Non-expert use of OR methods:** Pitfalls and risks; Is quick learning possible; Collaboration with experts.

In summary, any stage of any OR topic is open to behavioral investigations. While there is substantive progress to be made in studying the practice of OR, it is notable that the research methods and conceptual frameworks for extending the research program are well-established in other fields. In the field of strategic management, for instance, research on strategic decision processes has been organized into antecedents (or causes), process characteristics and outcomes (Rajagopalan et al. 1993). Similarly, it is possible to envision a research program into the antecedents, characteristics and outcomes of OR processes. Similarly, we need to evaluate and compare OR methods and processes on criteria such as ability to reach consensus, commit to action or challenge a set of assumptions about a given problem situation (e.g. Franco and Montibeller, 2010; cf. Mitroff and Emshoff 1979; Rajagopalan et al. 1993; Schweiger et al. 1986). To improve the practice it is important to include a behavioral perspective and compare the relative merits of modeling approaches used in different schools of thought in OR.

Behavioral studies in system dynamics

There has been active behavioral research going on in the system dynamics community (Moxnes, 2000). This literature also provides us an illustrative case to demonstrate how the OR process is especially sensitive to behavioral effects. Following this line of research, we investigate how people understand and make decisions regarding dynamic systems, a decision making context which OR practitioners commonly face. Research on how people understand and make decisions regarding dynamic systems has evolved in close relation with the development of system dynamics modeling (Moxnes, 2000). This research dates back to the early analysis of decision heuristics in simulation experiments and in feedback structures (Kleinmuntz, 1985; Kleinmuntz and Thomas, 1987; Sterman, 1987). Studies related to the famous Beer Game (Sterman, 1989a, b; Lee et al., 1997) have found that people attribute the so-called bullwhip effect in supply-chains to volatile end-demand when the actual cause is the system structure with delays in information and material flows. More recently, the issue of understanding dynamics in the climate change context (Sterman, 2008) has made this research area relevant for the climate policy debate as well.

Recently Cronin et al. (2009) continued experiments with the so-called department store task (Sterman, 2002), and they argue that people are unable to understand and apply the principles of accumulation. They claim that people’s inability to understand dynamic systems is a distinct new psychological phenomenon not described before. There are related experiments coming to similar conclusions that reasoning errors arise and affect decision making also in other contexts such as the management of renewable resources (Moxnes, 2000) and alcohol drinking (Moxnes and Jensen, 2009). These authors suggest that this should have important implications to education, policy processes and risk communication (Moxnes, 2000; Sweeney and Sterman, 2000; Sterman, 2008). Indeed, prior research has shown that humans are highly susceptible to using counterproductive heuristics when dealing with dynamic systems. We show that by taking into account the triggers of dysfunctional heuristics the capability of an individual to understand such systems in enhanced.

In this paper, we take the experiment of Cronin et al. (2009) as an illustrative case to show how various behavioral effects can be embedded in and effect OR processes. Our new results with somewhat revised experiments show that the poor performance in the department store task can be attributed to the framing of the problem rather than to people’s poor understanding of the accumulation phenomenon. We argue that in the department store task people’s performance is affected by several cognitive heuristics triggered by a number of factors in the task that camouflage and divert people’s attention from the true stock and flow structure. These factors give false cues and make incorrect answers more cognitively available to the subjects. We show that when these elements are removed, the performance of the subjects in the task improves. Thus, our study extends and reframes the results by Cronin et al. (2009) by highlighting the sensitivity of the OR process to cognitive biases, especially when dealing with dynamic systems. Moreover, the results highlight the importance of paying attention to the way in which modeling results are presented.

The department store task revisited

Sterman (2002) introduced the *department store task* to determine if people are able to understand and apply the principles of accumulation. Accumulation is a dynamic phenomenon referring to the growth of a stock variable when the inflow exceeds the rate of outflow from the stock. Carbon dioxide accumulates in the atmosphere when emissions are greater than the removal of CO_2 from the atmosphere. Bank savings increase when deposits exceed withdrawals. It has been argued that understanding accumulation is a difficult, albeit important cognitive skill when managing and communicating about such dynamic problems in business and policy settings (Sweeney and Sterman, 2000). Sterman (2002) asks that if people have difficulties understanding such a basic dynamic phenomenon, what chance we have in understanding more complex dynamic systems we encounter in real life.

In the task the participants answer a questionnaire which we here call *the original questionnaire*. It includes a graph (Figure 1) showing the number of people entering and leaving a department store during a 30 minute period. In stock and flow terms the curves show the inflow rate and outflow rate of people. These are the *flow* variables. The number of people in the store is the *stock* variable.

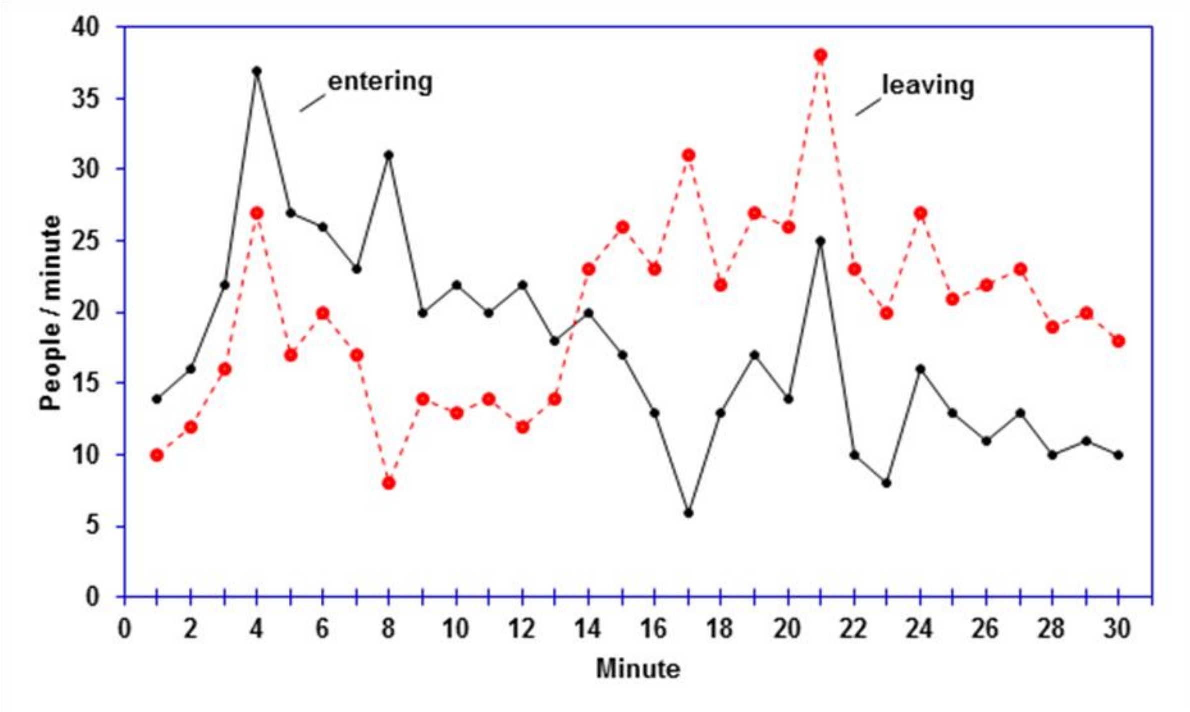

Figure 1. The graph shown to the participants in the original questionnaire representing the number of people entering and leaving the department store, redrawn after Cronin et al. (2009).

The original questionnaire presents four questions. The first two questions relate to the inflow and outflow variables and are said to be used to check if the participant understands graphs. *During which minute did the most people enter the store?* and *During which minute did the most people leave the store?* The following two questions relate to the stock variable. The questions are: *During which minute were the most people in the store?* and *During which minute were the fewest people in the store?* (Cronin et al., 2009, p. 118). The questionnaire asks the participants to write down these times. Alternatively one can also tick a "Can't be determined" box if one thinks that the answer cannot be determined from the information provided.

The accumulation phenomenon

To us the understanding of accumulation means simply that one understands that when the inflow exceeds outflow the stock increases and when the outflow exceeds inflow the stock decreases. None of the questions in the original questionnaire addresses this directly. In general, in a stock and flow system, the change in the stock variable at any given time can be computed by integrating the difference between the inflow and the outflow. The level of the stock at a given time can be obtained by adding the positive or negative change in the stock to the initial level of the stock. Due to the fact that this general procedure can always be used to solve the problem, one is likely to try to use it as a starting frame for problem solving also here. However, in the task the graph showing the number of people entering and leaving the department store represents a simpler special case of an accumulation system. In this special case, the above described computation is not required in order to determine the times at which the stock reaches its maximum (t = 13) and minimum (t = 30), see Cronin et al. (2009, p. 117-118). One only needs to observe the fact that the inflow and outflow curves intersect only once. If one is able to realize this, the correct answers are obvious. In such a situation the peak stock is reached at a point where the curves intersect i.e. when outflow becomes higher than inflow. Again to determine the moment when the stock reaches its minimum, one only needs to compare the area lying between the inflow and the outflow curves during the accumulation period, which is before minute 13, and the area lying between the outflow and the inflow curves when the net accumulation is negative, i.e. after minute 13. In the graph the latter is larger which means that the stock is at the lowest level at the end of the day.

In the department store task, focusing to think about accumulation as a general phenomenon does not help to find these easy solutions. On the contrary, taking the basic general approach described above makes the problem very difficult. An attempt to integrate the accumulation of the stock level directly is unsuccessful as the initial stock level is not provided. For determining the moment when the stock reaches its maximum, one should rather only think of the overall characteristics of this particular graph and the fact that in this case there is only one single maximum. Thus, answering the two questions related to accumulation requires an understanding of the principles of accumulation but also *other reasoning skills*.

Cronin et al. (2009) do indeed acknowledge that the department store task shares characteristics with *insight problems* (Mayer, 1995). Cronin et al. (2009, p 129) define them in the following way: “Insight problems are analytically easy—once one recognizes the proper frame to use. Until then, people tend to use a flawed but intuitively appealing (and hence difficult to change) problem frame”. Having recognized this, Cronin et al. still want to leave the possibility for the participants to anchor on a flawed frame in their experiment. This is quite strange. In the department store task people face a challenge of insight rather than a task of understanding accumulation.

The results and conclusions of Cronin et al.

In the Cronin et al. (2009, p. 119) experiment almost everybody was able to locate the maximum inflows (96 %) and outflows (95 %) by locating the related peaks in the respective curves. However, less than half of the participants were able to correctly determine the time at which the number of people in the store reached its maximum (44 %) and minimum (31 %). The authors also carried out experiments with different versions of the task and there were no significant changes in these results. Thus, Cronin et al. (2009, p. 116) argue, the poor performance "is not attributable to an inability to interpret graphs, lack of contextual knowledge, motivation, or cognitive capacity". The overall conclusion made by the authors was that people are not able to understand and apply the principles of accumulation.

Triggers of inappropriate heuristics in the original questionnaire

The questionnaires used by Cronin et al. (2009), including the original questionnaire and its variants, have two main problems. First, the questionnaire does not address people's understanding of accumulation directly. The accumulation related questions are not direct and require extra reasoning effort. Second, the questionnaires include elements that give false cues which mislead the participants. Below is a description of all the features that we think are distracting and mislead the participants.

– The questions relating to the inflow and outflow rate directs the participants' attention to the shapes of the curves, not to the accumulation phenomenon. The answers to the flow-related questions can be seen directly from the graph. The related peaks can be seen clearly from the curves. This primes the subjects to try to find the correct answers to the accumulation-related questions also directly from the graph.

– It is possible that the shapes of the curves trigger what Tversky and Kahneman (1974) call the availability heuristic. The maximum net inflow and the maximum net outflow clearly stand out of the graph (see, Figure 1). As a result they appear suitable solutions due to their availability. Cronin et al. (2009) do, in fact, following Sweeney and Sterman (2000), discuss this, too. They describe this phenomenon with their own new term "correlation heuristic". We think that the well-known availability heuristic covers this phenomenon as well.

– The check box, "Cannot be determined," acts as a cue that primes to think of the possibility that the question is very difficult and cannot be answered based on the data provided on the graph. The easy questions in the questionnaire deal with the maximum and minimum inflow and outflow of people. The accumulation-related questions are more not so straightforward to answer. Because of the easy questions, the participant may expect that also all the other answers should be available without much effort. If the participant is in this mindset it is tempting to check the “Cannot be determined” box when facing a more difficult question.

– The question regarding the minimum level of stock is quite demanding as one needs to compare two areas between the curves. Thus, it is possible that the participants become overwhelmed by the difficulty of this question which can lead them to feel insecure and to underestimate their own skills in the task, and this can lower their performance in the other questions as well.

Given these reasons, we were not convinced that the poor performance reported in the experiment of Cronin et al. was due to the inability of people to understand and apply the principles of accumulation. The alternative explanation is that the “red herrings” described above camouflaged the real phenomenon and diverted the participants attention in the experiment to focus, for example, on the shape of the graphs. Our assumption was that by removing triggers of inappropriate heuristics and by priming people to think about the stock and flow structure of the problem, the performance of the participants in the department store task can be increased. We tested this assumption in a series of four experiments.

Methods

Experiments with revised questionnaires

We repeated the same task by asking the participants directly about accumulation and tried to avoid red herrings priming the participants to think about the problem in a wrong way. We varied the order of the questions, the number of those elements we considered problematic in the original questionnaire and included new elements to manipulate how the participants solve the department store task. Four types of changes were made. First, we added questions asking about the accumulation phenomenon directly. Thus, we addressed the participants’ ability to understand accumulation more directly. Moreover, these questions primed the participants to think about the stock and flow structure of the department store task.

Second, we directed the participants to think about the department store task using explicit reasoning, not only intuition (cf., Kahneman, 2003). We did this by including the following additional question in the task: “Please help the manager of the department store. Provide him with an explanation with a couple of sentences describing how the graph can be used to answer the questions below.” It has been shown that providing written reasons for one’s decision leads to greater use of the available information (see, Milch et al., 2009). The so-called *exposition effect* refers to the phenomenon that people are less influenced by decision problem framing, that is, by the way in which the information is presented, if they are asked to give written reasons for their decisions (Sieck and Yates, 1997).

Third, we tried to reduce the impact of availability heuristic which suggests that the peak inflow rates or net inflow rate refer to the maximum stock level as well by using the graph with smoother curves. In one of the questionnaires, the figure showing the number of people entering and leaving the department store was a revised graph with smoother curves (Figure 2) and not the one with the strongly zigzagging curves of the original questionnaire (Figure 1). These revised curves still represent the same accumulation process as the curves in the original questionnaire. The maximum number of people is at the intersection of the curves, at t = 13, and the fewest at the end of the period, at t = 30, in the same way as in the original task. Finally, we reduced the total cognitive burden of the task by removing one of the accumulation-related questions regarding minimum level of the stock. We expected this to reduce cognitive burden to the point of helping the participants to answer the remaining accumulation-related question correctly.

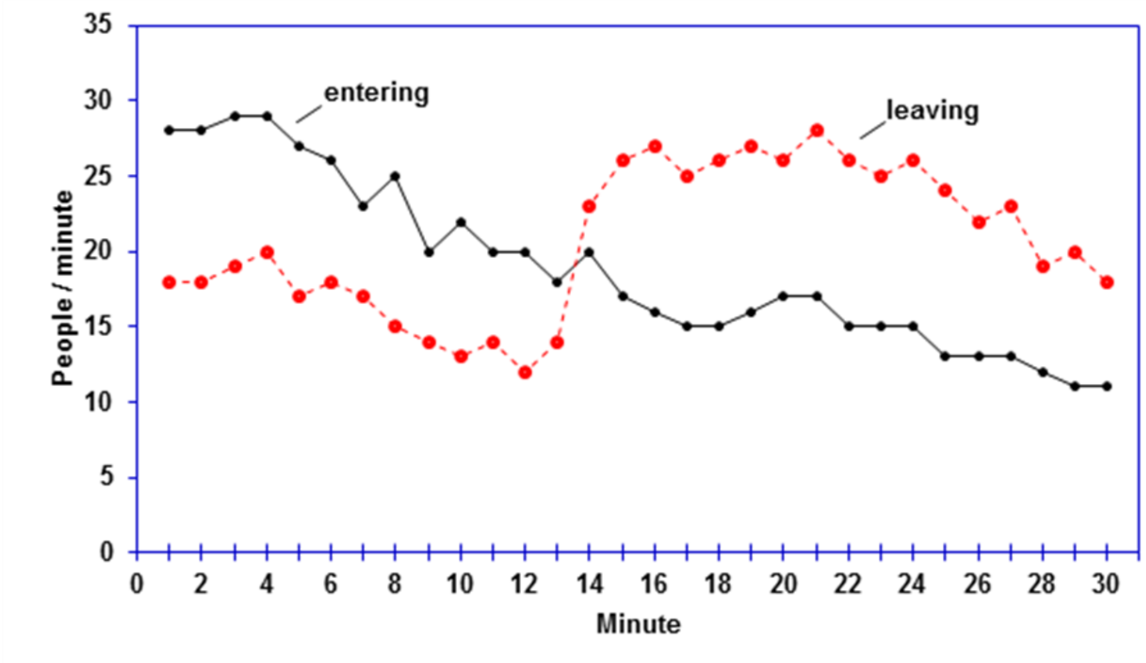

Figure 2. Revised graph with smoother curves for the number of people entering and leaving the department store.

Procedures

We ran four experiments using eleven different questionnaires, see Table 1. All the experiments were run during classes in the Helsinki University of Technology with different audiences. Participation was voluntary. The questionnaires were collected once everybody had completed the task. The participants used approximately 10 minutes to answer.

Table 1. Description of the questionnaires used in different experiments. The numbers refer to the order in which the question was presented in the questionnaire. A blank cell means that the question or element was not included. The questionnaires Original (a) and (b) are the same as those used by Cronin et al. (2009) in what they call their "baseline experiment"

	Questionnaire										
	I	**II**	**III**	**IVa**	**IVb**	**Va**	**Vb**	**VIa**	**VIb**	**Original (a)**	**Original (b)**
Graph											
Zigzagging curve (Figure 1)		x	x	x	x	x	x	x	x	x	x
Smooth curve (Figure 2)	x										
Additional elements											
"Can't be determined" box				x	x	x	x	x	x	x	x
Written explanation* asked				x	x			x	x		
The order of questions											
Q1. When did the number of people in the store increase and when did it decrease?	1	1									
Q2a.When did the number of people in the store increase?			1								
Q2b. When did the number of people in the store decrease?			2								
Q3a.During which minute did the most people enter the store?				1	2	1	2	1	3	1	3
Q3b.During which minute did the most people leave the store?				2	3	2	3	2	4	2	4
Q4a.During which minute were the most people in the store?	2	2	3	3	1	3	1	3	1	3	1
Q4b. During which minute were the fewest people in the store?	3	3	4					4	2	4	2

The participants, as described in Table 2, represented the typical student body in an engineering school besides in experiment 4 which was conducted in a popular philosophy class with attendance from the general public, too. This is reflected in the wider age range of the participants and in the lower percentage of males than in the other experiments. In Table 2, the questionnaires with the same questions but different orderings of the questions were pooled. For example, Original corresponds to both Original (a) and Original (b). This is why Table 2 shows seven different questionnaires instead of eleven listed in Table 1. Pooling is justified by the results, presented below, that show that ordering of the questions does not have an effect on the percentage of correct answers.

The participants were randomly given one questionnaire out of all the questionnaires used in the experiment. In Experiment 1, the participants were given either questionnaire I or the original questionnaire used by Cronin et al., in Experiment 2 there was only one questionnaire, in Experiment 3, either questionnaire III or the original questionnaire was given and, finally, in Experiment 4, either questionnaire IV, V, VI or the original questionnaire was given. By randomly distributing the questionnaires, we can ensure that the results are due to the treatments in the experiments. Specifically, we want to exclude the alternative explanation that the groups of people who fill out revised questionnaire are more competent in the task. To guarantee that the randomization was successful, we asked the participants to answer some questions regarding themselves. As shown in Table 2, randomization was successful, because within the experiments, the age, gender, and educational background of the participants did not differ across questionnaires.

Table 2. The experiments, questionnaires and the participants.

				Student status / degree								Professional / educational background				
				Percentage								Percentage				
Experiment	Questionnaire	Percentage of males	Mean age (range)	Undergraduate	PhD candidate	Holds MSc or equivalent	Holds PhD, LicSc or equivalent	High school student	Student in a university of applied science	Pre-high school	Other	Technology/planning	Research/education	Management	Art	Other
1 (N = 69)	**All**	**78**	**20 (18-28)**	**99**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**100**	**0**	**0**	**0**	**0**
	I (N = 32)	93	20 (19-22)	97	0	1	0	0	0	0	0	100	0	0	0	0
	Original (N = 37)	65	20 (18-28)	100	0	0	0	0	0	0	0	100	0	0	0	0
		$\chi^2(1, N=67) = 7.73$	*$F(1, N=67) = .94$*				*$\chi^2(1, N=67) = 1.25$*							*$\chi2$ test not applicable*		
		$p=0.01$	*$p=0.34$*				*$p=0.26$*									
2 (N=63)	**II (N=63)**	**83**	**20 (18-41)**	**98**	**0**	**2**	**0**	**0**	**0**	**0**	**0**	**97**	**0**	**3**	**0**	**0**
3 (N=219)	**All**	**77**	**23 (18-37)**	**93**	**0.5**	**6**	**0.5**	**0**	**0**	**0**	**0**	**78**	**5**	**15**	**0**	**2**
	III (N=113)	76	23 (18-30)	94	1	4	1	0	0	0	0	77	5	15	0	3
	Original (106)	78	23 (18-37)	92	0	8	0	0	0	0	0	79	4	16	0	1
		$\chi^2(1, N=67) = 0.11$	*$F(1, N=213)=1.64$*				*$\chi^2(3, N=209) = 3.52$,*							*$\chi^2(3, N=195) = 1.01$*		
		$p=0.74$	$p=0.20$				*$p=0.32$*							*$p=0.80$*		
4 (N = 201)	**All**	**70**	**29 (17-65)**	**64**	**12**	**17**	**2**	**1**	**1**	**0**	**5**	**64**	**8**	**17**	**4**	**7**
	IV (N=51)	63	29 (17-65)	65	4	24	0	2	0	0	6	69	13	20	4	2
	V (N=52)	73	29 (19-62)	60	23	13	0	0	0	0	4	56	12	16	6	10
	VI (N=50)	83	29 (19-61)	63	10	16	4	0	0	0	6	68	11	13	2	6
	Original (N=48)	60	28 (19-57)	69	13	15	2	0	2	0	2	65	4	19	2	10
		$\chi^2(3, N=199) = 7.65$,	*$F(3, N=193) =.08$*				*$\chi^2(12, N=194) = 9.61$,*							*$\chi^2(12, N=194) = 9.61$,*		
		$p = 0.05$	*$p = .97$*				*$p = 0.65$*							*$p = .65$*		

Specifically, the χ2 tests reported in Table 2 confirm that the groups of participants who were given different questionnaires are similar, as measured by the percentages of participants with different student statuses or degrees and professional or educational backgrounds. Specifically, the p-values associated with the χ2 tests are above .10 which leads us to conclude that the professional or educational background was not different across questionnaires (Hair et al., 2005). The mean age is also similar across the questionnaires, as confirmed by one-way analysis of variance. The p-values related to the one-way ANOVA are above .10 showing that the average age of the participants does not vary significantly across questionnaires. Finally, in experiment 1 and experiment 4, the proportion of male participants was different with different questionnaires ($\chi2$ (1, N=67) = 7.73, p=0.01; $\chi2$(3, N=199) = 7.65, p=0.05, respectively). Hence, we also ran separate analyses for males and females. However, for the sake of clarity, we only report the results obtained with the pooled data as the results with the split data were essentially similar to those reported here.

Results

Experiments 1-4

The overall results are presented in Table 3 as the percentage of correct answers with different questionnaires. For comparison the results of the "baseline experiment" of Cronin et al. (2009, see Table 1) are also included. It is quite interesting that our results are almost identical to those of Cronin et al. when the original questionnaire was used. These are shown in the last two columns.

Table 3. The percentage of correct answers to different questions in the different questionnaires.

	Our experiments							**Cronin et al (2009) experiment**
	Questionnaire							
Question	**I (*N* = 32)**	**II (*N* = 63)**	**III (*N* = 113)**	**IV[a] (*N* = 51)**	**V[a] (*N* = 52)**	**VI[a] (*N* = 47)**	**Original[a] (*N* = 193)**	**Original**
Q1. When did the number of people in the store increase and when did it decrease?	81	89		-	-	-	-	-
Q2a. When did the number of people in the store increase?	-	-	75	-	-	-	-	-
Q2b. When did the number of people in the store decrease?	-	-	77	-	-	-	-	-
Q3a. During which minute did the most people enter the store?	-	-	-	92	100	94	97	96
Q3b. During which minute did the most people leave the store?	-	-	-	92	100	94	97	95
Q4a. During which minute were the most people in the store?	88	90	58	55	69	70	50	44
Q4b. During which minute were the fewest people in the store?	72	76	48	-	-	50	34	31

[a] The answers to variants a and b of questionnaires IV, V, VI and Sterman are pooled.

In experiment 4, following Cronin et al. (see, 2009, p. 119), we tested two orderings of the questions: a) questions about the flows first (questionnaires IVa-VIa and Original (a)); b) questions about the stock first (questionnaires IVb-VIb and Original (b)). We compared the results of IVa and IVb, Va and Vb, VIa and VIb, and, finally, Original (a) and Original (b). The results obtained with the two orderings were pooled as the question ordering did not affect the proportion of correct answers. Fischer's exact test for differences in the proportion of correct answers on each of the four questions yielded p=.22 or greater on each questionnaire (IV-VI and the original questionnaire). Henceforth, IV refers to the pooled results of questionnaires IVa and IVb, and "original" refers to the pooled results of questionnaires "original (a)" and "original (b)". The pooled results are shown in Table 3.

The percentage of correct answers to the questions related to flows (Q3a and Q3b) is consistently high across all the questionnaires. However, the results in the accumulation related questions Q4a and Q4b vary with the questionnaire used. The overall percentage of correct answers is higher with all of the revised questionnaires (I-VI) than with the original questionnaire.

The most important result is that people do give correct answers when asked directly about accumulation (questions Q1-Q2 a, b). For instance, a very high percentage of the participants, who filled questionnaire I or II, answered the question "When did the number of people in the store grow and when did it decline?" correctly, 81 and 89 percent, respectively. Moreover, it seems that when people are positively primed and the misleading cues were not in place (questions Q4a and Q4b, questionnaires I, II), up to 90 percent of the people gave correct answers to the difficult questions such as the one "During which minute were the most people in the store?" The corresponding percentage reported by Cronin et al. (2009) was only 44. For the question "During which minute were there the fewest people in the store?" the percentages of correct answers were as high as 72 and 76. This is a very important difference in the results as Cronin et al. report that only 31 percent were able to give the correct answer.

The percentages of correct answers to questions Q3-Q4 a, b are presented in Table 4. These were the original questions used in the Cronin et al. paper. Fischer's exact test is used to determine whether the percentage of correct answers differs statistically for the revised questionnaires and the original questionnaire. The test compares the percentages of correct answers obtained with the revised questionnaires with the respective percentages obtained with the original questionnaire in the same experiment. For example, in Experiment 1, 41 percent of the people who filled the original questionnaire answered correctly to question Q4a. Conversely, 88 percent of the people who filled the revised questionnaire I, gave the correct answer to Q4a. Clearly, the revised questionnaire produced better results. Indeed, the Fisher's exact test yields a p-value close to zero which means that there is small (less than 0. 01 percent) probability that the difference in the percentage of correct answers is due to chance. Thus, the lower the p-value, the higher the probability that, in the experiment under consideration, the revised questionnaire yielded significantly different results than the original questionnaire used in the same experiment.

Table 4. The percentages of correct answers to questions Q3a, b and Q4a, b in the different experiments.

	Experiment 1		**Experiment 2**	**Experiment 3**		**Experiment 4**			
Question	I (N=32)	Original (N=37)	II (N=63)	III (N=113)	Original (N=106)	IV (N=51)	V (N=52)	VI (N=50)	Original (N=48)
Q3a. During which minute did the most people enter the store?		95				92 (p=0.36)	100 (p=0.48)	94 (p=0.62)	98
Q3b. During which minute did the most people leave the store?		95				92 (p=0.36)	100 (p=0.48)	94 (p=0.62)	98
Q4a. During which minute were the most people in the store?	88 (p<0.0001)	41	90	58 (p=0.89)	57	55 (p=0.32)	69 (p=0.02)	70 (p=0.01)	44
Q4b. During which minute were the fewest people in the store?	72 (p<0.0001)	24	76	57 (p=0.22)	48			50 (p=0.05)	31

The result is that the proportion of correct answers on the flow-related questions (Q3a and Q3b) does not vary with the questionnaire type. In experiment 1, questionnaire I yielded a significantly higher percentage of correct answers than the original questionnaire. In experiment 2, no comparisons are made because the original questionnaire was not used. Nevertheless, it is worth noticing that the performance on the accumulation-related questions is very good. The percentages of correct answers are 89 (Q4a) and 75 (Q4b). The questionnaire III did not produce statistically significantly different results than the original questionnaire in experiment 3.

Finally, in experiment 4, those who filled questionnaires V or VI gave more often a correct answer to question Q4a, than those who had the original questionnaire. The percentage of correct answers on the other accumulation-related question (Q4b) was slightly higher for questionnaire VI than for the original questionnaire.

Summary and discussion of results

In our experiments with the revised questionnaires, we asked the participants about the accumulation phenomenon directly. The percentages of correct answers were 81 (experiment 1) and 89 (experiment 2), showing that almost all of the participants were able to understand accumulation.

Also in the absence of the misleading cues people do indeed perform better in the accumulation related questions of the original task. The conclusion is supported most strongly by the results of experiment 1 where the percentage of correct answers to the questions doubled from 44 to 88 and 90 and from 31 to 72 and 76 when the problematic elements were removed (Table 3). These are clearly higher percentages of correct answers than what is typically reported in studies that use the department store task; see Pala and Vennix (2005) for a review.

It is interesting to interpret these observations from the OR perspective. Based on our experimental results, we see that people's poor performance in the department store task (Cronin et al., 2009) does not reflect the existence of a new cognitive bias but, rather, shows the pervasiveness of different heuristics and biases in the context of dynamic systems. Various factors may also undermine the correct understanding of the communication about dynamic systems in the process of OR. Specifically, when communicating about dynamic phenomena in model-based problem solving, people are susceptible to multiple framing and priming effects. As noted, other behavioral phenomena that are likely to influence the OR process include confirmation bias, tendency to prefer intuitive-sounding alternatives, attribute substitution heuristics, group think, information overload, appeal to authority and cognitive dissonance. Psychologists cannot solve these problems for us, which is why they should be addressed from within the OR community. For example, confirmation biases may arise in different ways, and in different stages of the OR process. The effects may arise from the interaction between OR models and human cognition, emotions and group processes. Besides specific biases it is important to understand how people think more generally, and how modeling, for better or worse, supports intuitive, algorithmic and reflective thinking (Stanovich, 2010). These can only be identified and addressed by studying explicitly how they emerge in the OR process.

Conclusions

Operational research is human action that uses quantitative and quantitative methods to facilitate thinking and problem solving. Thus it is never just about models, but about people too. How OR methods facilitate thinking and problem solving, which kinds of cognitive biases and limitations it can help people overcome, and what new behavioral issues it gives rise to are matters that are at the core of OR practice. Yet these topics have received relatively little attention by OR scholars. So far, the two dominant modes of doing academic research on the practice of OR are conceptual discussions and illustrative case studies (e.g., Yunes et al., 2007). Comparative behavioral studies are lacking almost completely. We have argued that by empirically investigating the behavioral aspects of the OR process it is possible to shed new light onto how the practice of OR can be improved. By doing a series of experiments related to the understanding of and communication about dynamic systems, we showed that the communication phase of the OR process can be sensitive to priming and framing effects. By paying attention to these effects, the practitioner can improve the effectiveness of OR interventions.

In general, when one would like to start research on behavioral OR a natural idea is to carry out retrospective analyses of existing case studies with a behavioral lens. Unfortunately, the publications and reports of past OR processes do not provide detailed descriptions of how the OR methods were applied. Papers published in the literature do not usually provide critical evaluations of the reasons why the application of the OR method was successful or unsuccessful. That said, understanding when and why a particular OR process produces good outcome would be important. Here, OR researchers could learn from case study researchers in the social sciences where it is usually required that researchers ”describe the case in sufficient descriptive narrative so that readers can experience these happenings vicariously and draw their own conclusions” (Stake, 2005: 450). Such rich descriptions of OR process would help the OR community to draw conclusions about the conditions under which our methods are successful.

Future research could study the OR process at the individual, group and organizational levels of analysis and apply different theoretical perspectives and empirical research methodologies. Despite a few notable studies, this new research area within OR has not been recognized as an integral part of OR. This, we argue, hinders the betterment of OR practice. For instance, one possibility is to do single- (Siggelkow, 2007) or multiple-case studies (Eisenhardt, & Grabner, 2007) that do not necessarily attempt to introduce new methods, but to study how and in which circumstances certain methods produce outcomes that are relevant to practice (e.g., customer satisfaction, learning). By integrating behavioral studies into OR, it is possible to increase the field's relevance to a wider audience – including economics, finance and marketing – which is also interested in improving model-based problem solving.

More broadly, when facilitating an OR process and communicating about models and systems with the clients we also become part of a psychosocial system created by the interaction in the joint problem solving situation. People, with different mental models, expectations and emotional factors bring subjective and active elements into this system, which again becomes a real part of the OR process as has been noted in the OR practice literature (Miser, 2001). This too needs to be taken into account in the facilitation process. We need to take a systems thinking perspective on the whole which consists of the problem, the OR process and the participants in it. The concept of systems intelligence (Saarinen and Hämäläinen, 2004, Luoma et al., 2010) refers to one's abilities to successfully engage with systems both taking into account the explicit and implicit feedback mechanisms to the non-human and human elements in systems. Adopting this perspective could help to see the big picture and create successful communication strategies and ways of acting from within the overall system. People do have these systems skills, and we believe these are crucial in successful joint problem solving processes in OR too. The concern raised by Ackoff (2006) that systems thinking is not used in organizations can also reflect the lack of systems intelligence skills in organizations (see Hämäläinen Saarinen 2008). It would indeed be of great interest to develop a training program to raise the systems intelligence awareness and skills of OR practitioners to improve their support processes and to find more successful model based problem solving and communication strategies. Developing practitioner skills with a behavioral lens will keep OR alive and interesting for our customers and the society at large.

References

Ackert, L., & Deaves, R. (2009). *Behavioral Finance: Psychology, Decision-Making, and Markets* (1st ed.). South-Western College Publishing.

Ackoff, R. L. (2006). Why few organizations adopt systems thinking. *Systems Research and Behavioral Science*, *23*(5), 705-708.

Barberis, N., & Thaler, R. (2003). Chapter 18 A survey of behavioral finance. *Financial Markets and Asset Pricing* (Vol. Volume 1, Part 2, pp. 1053-1128). Elsevier.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review, 12(3), 183-210.*

Bell, M. L., Hobbs, B. F., & Ellis, H. (2003). The use of multi-criteria decision-making methods in the integrated assessment of climate change: implications for IA practitioners. *Socio-Economic Planning Sciences*, *37*(4), 289-316.

Bendoly, E., Donohue, K., & Schultz, K.L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, *24*, 737-752.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, *19*(4), 434-452.

Brailsford, S., Harper, P., & Shaw, D. (2009). Milestones in OR. *Journal of the Operational Research Society*, 60 (Supplement), S1-S4.

Brocklesby, J. (2009). Ethics beyond the model: How social dynamics can interfere with ethical practice in operational research/management science. *Omega*, *37*(6), 1073-1082.

Brocklesby, J., & Cummings, S. (1996). Foucault Plays Habermas: An Alternative Philosophical Underpinning for Critical Systems Thinking. *The Journal of the Operational Research Society, 47*(6), 741-754.

Camerer, C. F, & Loewenstein, G. (2004). Behavioral economics: Past, present, future. Camerer, Colin F., Loewenstein, George, & Rabin, M. (2003). *Advances in Behavioral Economics*. Princeton University Press, 3–51.

Checkland, P. (2000). Thirty Year Retrospective. *Systems Research and Behavioral Science*, 17 (Supplement), S11-S58.

Churchman, C. W. (1968). *The Systems Approach*. Delacorte Press.

Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory*. Sage publications.

Corbett, C. J., Overmeer, W. J. A. M., & Van Wassenhove, L. N. (1995). Strands of practice in OR (the practitioner's dilemma). *European Journal of Operational Research*, *87*(3), 484-499.

Cronin, M. A., Gonzalez, C., & Sterman, John D. (2009). Why don’t well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, *108*(1), 116-130.

Dahl, R. A. 1957. The concept of power. *Behavioral Science*, 2(3), 201–215.

Doktor, R. H., & Hamilton, W. F. (1973). Cognitive style and the acceptance of management science recommendations. *Management Science*, *19*(8), 884–894.

Drake, M. J., Gerde, V. W., & Wasieleski, D. M. (2009). Socially responsible modeling: a stakeholder approach to the implementation of ethical modeling in operations research. *OR Spectrum*, *33*(1), 1-26.

Eden, C. (2004). Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research*, *159*(3), 673–686.

Eden, C., & Ackermann, F. (2006). Where next for problem structuring methods. *Journal of the Operational Research Society*, *57*(7), 766-768.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.

Forrester, J. W. (1958). Industrial Dynamics: A Major Breakthrough for Decision Makers. *Harvard Business Review*, *36*(4), 37-66.

Forrester, J. W. (1994). System dynamics, systems thinking, and soft OR. *System Dynamics Review*, *10*(2-3), 245-256.

Franco, L.A., & Montibeller, G. (2010). Facilitated modeling in operational research. *European Journal of Operational Research*. *205*(3), 489-500.

Franco, L. A., & Rouwette, E. A. J. A. (2011). Decision development in facilitated modelling workshops. *European Journal of Operational Research*, *212*(1), 164-178.

French, S., Maule, J., & Papamichail, N. (2009). *Decision Behaviour, Analysis and Support*. Cambridge University Press.

French, S., Simpson, L., Atherton, E., Belton, V., Dawes, R., Edwards, W., Hämäläinen, R.P., Larichev, O., Lootsma, F., Pearman, A., & Vlek, C. (1998) Problem formulation for multi-criteria decision analysis: Report of a workshop. *Journal of Multi-Criteria Decision Analysis*, 7 (5), 242-262.

Gass, S. I., & Harris, C. M. (2001). *Encyclopedia of operations research and management science*. Springer.

Gino, F., & Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management, 10*(4), 676-691.

Grenander, U. (1965). An Experiment in Teaching Operations Research. *Operations Research, 13*(5), 859-862.

Hair, J. F., Black, B., Babin, B., Anderson, R. E., & Tatham, R. L. (2005). *Multivariate Data Analysis* (6th ed.). Prentice Hall.

Hämäläinen, R. P., & Saarinen, E. (2008). Systems intelligence - the way forward? A note on Ackoff's "Why few organizations adopt systems thinking." *Systems Research and Behavioral Science, 25*(6), 821-825.

Jackson, M. C. (2009). Fifty years of systems thinking for management. *Journal of the Operational Research Society*.

Janis, I. L. (1972). *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Houghton Mifflin.

Kahneman, D. (2003a). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist,* 58(9), 697-720.

Kahneman, D. (2003b). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

Kahneman, D. & Tversky, A. (2000). *Choices, Values and Frames*. Cambridge University Press.

Keys, P. (1997). Approaches to understanding the process of OR: Review, critique and extension. *Omega*, *25*(1), 1-13.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, *31*(6), 680-702.

Kleinmuntz, D. N., & Thomas, J. B. (1987). The value of action and inference in dynamic decision making. *Organizational Behavior and Human Decision Processes*, *39*(3), 341-364.

Korhonen, P., Mano, H., Stenfors, S., & Wallenius, J. (2008). Inherent biases in decision support systems: the influence of optimistic and pessimistic DSS on choice, affect, and attitudes. *Journal of Behavioral Decision Making*, *21*(1), 45-58.

Korhonen, P., & Wallenius, J. (1996). Behavioural Issues in MCDM: Neglected Research Questions. *Journal of Multi-Criteria Decision Analysis*, *5*(3), 178-182.

Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System Neglect and change detection. *Management Science, 57*(10), 1827-1843.

Landry, M. & Oral, M. (1993). In search of a valid view of model validation for operations research. *European Journal of Operational Research, 66, 161-167.*

Le Menestrel, M., & Van Wassenhove, L. N. (2004). Ethics outside, within, or beyond OR models? *European Journal of Operational Research*, *153*(2), 477-484.

Le Menestrel, M., & Van Wassenhove, L. N. (2009). Ethics in Operations Research and Management Sciences: A never-ending effort to combine rigor and passion. *Omega*, *37*(6), 1039-1043.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, *43*(4), 546-558.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects,. *Organizational Behavior and Human Decision Processes*, *76*(2), 149-188.

Levinthal, D. A. (2011). A behavioral approach to strategy—what's the alternative? *Strategic Management Journal*, *32*(13), 1517–1523.

Loch, C. H., & Wu, Y. (2007). *Behavioral Operations Management*. Now Publishers Inc.

Lu, H. P., Yu, H. J., & Lu, S. S. K. (2001). The effects of cognitive style and model type on DSS acceptance: An empirical study. *European Journal of Operational Research*, *131*(3), 649–663.

Luoma, J., Hämäläinen, R. P., & Saarinen, E. (2010). Acting with systems intelligence: integrating complex responsive processes with the systems perspective. *Journal of the Operational Research Society*, *62*(1), 3-11.

Mayer, R. E. (1995). The search for insight: Grappling with Gestalt psychology's unanswered questions. *The Nature of Insight* (p. 3–32). MIT Press.

Milch, K. F., Weber, E. U., Appelt, K. C., Handgraaf, M. J. J., & Krantz, D. H. (2009). From individual preference construction to group decisions: Framing effects and group processes. *Organizational Behavior and Human Decision Processes*, *108*(2), 242-255.

Mingers, J. (2006). *Realising Systems Thinking: Knowledge and Action in Management Science*. Springer Verlag.

Mingers, J. (2011). Soft OR comes of age--but not everywhere! *Omega*, *39*(6), 729-741.

Mingers, J., Liu, W., & Meng, W. (2009). Using SSM to structure the identification of inputs and outputs in DEA. *Journal of the Operational Research Society*, *60*, 168-179.

Mingers, J., & Brocklesby, J. (1997). Multimethodology: Towards a framework for mixing methodologies. *Omega*, *25*(5), 489-509.

Mingers, J, & Rosenhead, J. (2004). Problem structuring methods in actions. *European Journal of Operational Research*, *152*, 530-554.

Miser, Hugh. 2001 "Practice of operations research and management science" in Gass, S.I, and C. M Harris. 2001. Encyclopedia of operations research and management science. Springer Netherlands. 625-630. (n.d.). .

Mitroff, I. I., & Emshoff, J. R. (1979). On Strategic Assumption-Making: A Dialectical Approach to Policy and Planning. *Academy of Management Review*, *4*(1), 1-12.

Morton, A., & Fasolo, B. (2008). Behavioural decision theory for multi-criteria decision analysis: a guided tour. *Journal of the Operational Research Society*, *60*(2), 268–275.

Moxnes, E. (2000). Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System Dynamics Review*, *16*(4), 325-348.

Moxnes, E., & Jensen, L. (2009). Drunker than intended: Misperceptions and information treatments. *Drug and Alcohol Dependence*, *105*(1-2), 63-70.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175-220.

Ormerod, R., (2008). The Transformation Competence Perspective. Journal of the Operational Research Society, 59 (12), 1573-1590.

Pala, Ö., & Vennix, J. A. M. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review*, *21*(2), 147-172.

Rajagopalan, N., Rasheed, A. M. A., & Datta, D. K. (1993). Strategic decision processes: Critical review and future directions. *Journal of Management*, 19(2), 349-384.

Ravindran, A., Phillips, D. T., & Solberg, J. J. (1976). *Operations research: principles and practice*. Wiley New York.

Repenning, N. P., & Sterman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened. *California management review*, *43*(4), 64–88.

Richels, R. (1981). Building good models is not enough. *Interfaces, 11(4), 48-54.*

Rouwette, E. A. J. A., Bastings, I., & Blokker, H. (2011). A comparison of group model building and strategic options development and analysis. *Group Decision and Negotiation, 20*, 781-803.

Rouwette, E. A. J. A., Korzilius, H., Vennix, J. A. M., & Jacobs, E. (2010). Modeling as persuasion: the impact of group model building on attitudes and behavior. *System Dynamics Review*, 27(1), 1-21.

Rouwette, E.A.J.A., & Vennix, J.A.M. (2006). System dynamics and organizational interventions. *Systems Research and behavioral Science, 23*, 451-466.

Royston, G. (2009). One hundred years of Operational Research in Health—UK 1948–2048. *Journal of the Operational Research Society,* 60 (Supplement), S169-S179

Saarinen, E., & Hämäläinen, R. P. (2004). Systems Intelligence: Connecting Engineering Thinking with Human Sensitivity. *Systems Intelligence: Discovering a Hidden Competence in Human Action and Organizational Life,* Systems Analysis Laboratory Research Reports. Helsinki University of Technology.

Santos, S. P., Belton, V., & Howick, S. (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations & Production Management*, *22*(11), 1246-1272.

Sieck, W., & Yates, J. F. (1997). Exposition Effects on Decision Making: Choice and Confidence in Choice,. *Organizational Behavior and Human Decision Processes, 70*(3), 207-219.

Siggelkow, N. (2007). Persuasion with case studies. *The Academy of Management Journal*, *50*(1), 20–24.

Silva, C. A., Sousa, J. M. C., Runkler, T. A., & Sá da Costa, J. M. G. (2009). Distributed supply chain management using ant colony optimization. *European Journal of Operational Research*, *199*(2), 349-358.

Simmons, J. P. & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General*, 135(3), 409-428.

Simon, H. A. (1996). *The Sciences of the Artificial,* 3rd ed. MIT Press.

Simon, H. A. (2000). Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1(1), 25-39.

Stake, R. E. (2005). Qualitative case studies. *The SAGE Handbook of Qualitative Research*. (N. K. Denzin & Y. S. Lincoln, Eds.). Sage Publications.

Stanovich, K. E. (2010). *What Intelligence Tests Miss: The Psychology of Rational Thought*. Yale University Press.

Sterman, J. D. (1987). Testing Behavioral Simulation Models by Direct Experiment. *Management Science*, *33*(12), 1572-1592.

Sterman, J. D. (1989a). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science*, *35*(3), 321-339.

Sterman, J. D. (1989b). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes*, *43*(3), 301-335.

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw-Hill.

Sterman, J. D. (2002). All models are wrong: reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501-531.

Sterman, J. D. (2008). Economics: Risk Communication on Climate: Mental Models and Mass Balance. *Science*, *322*(5901), 532-533.

Sweeney, L. B., & Sterman, John D. (2000). Bathtub dynamics: initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249-286.

Schweiger, D. M., Sandberg, W. R., & Ragan, J. W. (1986). Group Approaches for Improving Strategic Decision Making: A Comparative Analysis of Dialectical Inquiry, Devil's Advocacy, and Consensus. *Academy of Management Journal*, *29*(1), 51-71.

Starbuck, W. H. (2006). *The Production of Knowledge: The Challenge of Social Science Research*. Oxford University Press.

Travis, C & Aronson, E. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts*. Orlando, FL: Harcourt.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Ulrich, W. (1994). *Critical Heuristics of Social Planning: A New Approach to Practical Philosophy*. Wiley.

Wenstøp, F. (2010). Operations research and ethics: development trends 1966-2009. *International Transactions in Operational Research*, *17*(4), 413-426.

von Winterfeldt, D., & Fasolo, B.. (2009). Structuring decision problems: A case study and reflections for practitioners. *European Journal of Operational Research*, *199*(3), 857-866.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research* (Vol. 1). Cambridge University Press.

Yunes, T. H., Napolitano, D., Scheller-Wolf, A., & Tayur, S. (2007). Building efficient product portfolios at John Deere and Company. *Operations Research*, *55*(4), 615–629.