Balancing Bathtubs in Math Class

Eva Jensen
Department of War Studies
Swedish National Defence College
P.O. Box 27805
SE-115 93 Stockholm, Sweden
Telephone: + 46 8 5534 2610
Fax: + 46 8 5534 2600
E-mail: eva.jensen@fhs.se

What of the knowledge that is required to understand dynamic systems is covered by the traditional school math curriculum? In this study, the Booth Sweeney and Sterman (2000) bathtub tasks and the Jensen and Brehmer (2003) rabbits-and-foxes task were dissected into aspects. Questionnaires with tasks tapping into the identified aspects were administered to first-semester university engineering students and hypotheses were formulated concerning their performance, based on an analysis of the high school mathematics curriculum. The results largely conformed to the expectations. However, many participants misunderstood the questions addressing ability to perform control operation, which is required by the rabbits-and-foxes task. This is something that is not taught in schools, and we do not know what questions to ask because we know too little about how people think about control. We would also need to test the tasks with different participant groups, to learn about what aspects are learnt under what circumstances.

What is the relation between system dynamics and school mathematics? Systems thinking, considering causal loops and nets (maybe using causal loop diagrams), rather than solely one-way causal chains, we seldom see in traditional school mathematics (Ossimitz, 2000).

In system dynamics, the objective is to simulate the behavior of systems. For that purpose, accumulations (stocks) are separated from transactions (flows) of goods (which are sometimes of a more abstract nature), and values are estimated for the parameters determining system behavior.

System dynamics uses stocks and flows instead of difference or differential equations to describe systems. The stock-and-flow diagrams are, however, converted into difference equations by the simulation software. Thus, stock-and-flow diagrams can be seen as a more user-friendly representation than differential (or difference) equations of a dynamic system. They convey the structure and relations of the parts that constitute a system in an illustrative way. They also allow the user to assign values to parameters and define relations without formal knowledge of differential (or difference) equations or any other sophisticated mathematics. This might also be one of its dangers; it is deceptively easy to use. The software constructs all the complex equations and performs all the calculations for the user, but what about the results? How easy are they to interpret?

Is there anything in the representation that suggests how the behavior of flows affects the behavior of stocks, or vice versa? No, this is no part of the stock-and-flow representation. As already mentioned, the stock-and-flow diagrams describe the *structure* of systems, something that could be represented in various ways. System behavior, however, is described by line graphs, a fairly traditional way of describing behavior of variables in time. The bathtub tasks

(Booth Sweeney & Sterman, 2000) test if the participants can figure out, and graphically represents, what is in a container, from information about the behavior in the pipes attached to it.

Flows describe changes to stocks. Inflows add to the stock, and outflows reduce it. The net flow is mathematically the derivative of the stock variable. Similarly, the net change in the stock variable during a specified period of time [t1, t2] is the integral of the net flow during this time period. Thus, stocks and flows is just another way to represent derivatives and integrals. Even if you do not need to be able calculate derivatives and integrals formally, you probably need to understand the principles conceptually, to fully, if at all, appreciate the results from a system dynamics simulation. It seems likely that the ability to solve the bathtub problems should be closely related to knowledge about derivatives and integrals.

Both derivatives and integrals are generally dealt with formally before, or in the process of, developing a conceptual understanding of them. In Sweden these concepts are introduced rather late in the mathematics course in high school (e.g., Björk & Brolin, 2000b), and only to those who choose the full mathematics course offered, which generally are students following the science or technology programs. Integrals are among the last things introduced before the students graduate from high school. This means that it might take some additional mathematics studies at university level for these concepts to be fully mastered.

A body of studies testing the bathtub tasks with different student groups has accumulated, since the problems were first published (Booth Sweeney & Sterman, 2000; Fisher, 2003; Kainz & Ossimitz, 2002; Kapmaier, 2004; Lyneis & Lyneis, 2003; Ossimitz, 2002; Zaraza, 2003). Together they paint a picture in which system classes (Zaraza, 2003), advanced algebra and calculus (Fisher, 2003; Zaraza, 2003), and a major in engineering (Kapmaier, 2004), all contribute favorably to the performance in these tasks.

In the present study, the bathtub tasks are analyzed in terms of what knowledge they require that might be related to what is taught in regular mathematics classes. This might shed some light on why the different student groups performing the task perform the way they do.

Jensen and Brehmer (2003) investigated how people interpret and make use of information provided in line graphs, the means generally used to communicate simulation results. The line graphs describe the behavior of a rabbit (prey) population and a fox (predator) population. The populations mutually influence each other, and the task is to bring the system to equilibrium by controlling the size of the fox population. People generally experience severe problems when performing this task (Jensen 2003, 2005; Jensen & Brehmer, 2003). Understanding the bathtub problems is probably fundamental, at least to some extent, to understanding the rabbits-and-foxes task. The rabbits and foxes can be viewed as two stocks (bathtubs) with inflows and outflows. In addition, the rabbits-and-foxes task addresses aspects not covered by the bathtub task, but nonetheless important and fundamental to the understanding of dynamic systems.

Hence, the rabbits-and-foxes task can also be dissected into aspects, some addressed by school mathematics, others not. To capture the different aspects of the rabbits-and-foxes task and the bathtub tasks, tasks addressing specific aspects have been constructed, some of them inspired by Kaintz and Ossimitz (2002).

In the present study, questionnaires with tasks tapping into relevant aspects were administered to university students who had recently completed a science or technology program in high school. Hypotheses are formulated concerning the difficulty of the different tasks in regard of the mathematical background of this student group. Hypotheses are also formulated regarding how performance in these tasks will be related to performance in other tasks included in the questionnaires.

One particularly important aspect is the control performance required by the rabbits-and-foxes task. This is something that probably receives very little attention in schools. It is, however, possible that people learn to perform system control in other situations, outside the school curriculum. To control a system may pose quite different demands than merely understanding, representing or predicting system behavior. Perhaps the former are prerequisites for the latter, but at the present time we do not know.

I have used the textbooks on high school mathematics by Björk & Brolin (2000a, b) as my major source of information concerning what is taught and when in the mathematics classrooms of the Swedish high schools. It is not likely that all the participants had been subjected to these textbooks when in high school, and all teachers do not follow the textbook that closely. Students following the same study program at different high schools in Sweden are, however, expected to acquire the same knowledge regardless of where in the country they are studying. This contributes to quite a high degree of homogeneity, both in classroom work and in textbook construction. There is not a great number of textbooks in high school mathematics to choose from, and they all cover largely the same topics in a similar order. The Björk & Brolin (2000a, b) books were chosen because they are fairly typical of Swedish high school mathematics textbooks and they seem to be widely used.

Method

Participants: Twenty first-semester engineering students, nineteen male and one female, at Luleå Technical University at Skellefteå volunteered to participate, and were rewarded with two cinema tickets. Their mean age was 21 years, ranging from 19 to 27. During the initial semester, they generally take courses in scientific method and other courses unrelated to mathematics. They were expected to have concluded their studies at high school recently, in a science or technology program, and not having studied much mathematics yet since then. We made sure they had followed a science or technology program in high school.

Questionnaires: The questions in the questionnaires can be found in the Appendix (supporting material) to this paper. Two different questionnaires were constructed. For some tasks an important transfer could be expected. These tasks were put in different questionnaires. This solution also made it possible to include more tasks without producing unnecessary fatigue on behalf of our participants.

Procedure: The questionnaires, 10 of each set, were administered randomly to the participants. The participant groups receiving the different sets of questionnaires are called Group X and Group Y respectively. The questionnaires were answered individually, but administered in a group session. A fellow student supervised the sessions. The participants were allowed a maximum of two hours to fill in the questionnaire. In addition to solving the problems, they answered questions about their age, gender, how much time they spent on the questions, if they thought the allotted time was enough, and they indicated on a Likert-scale how difficult they found the questions (1 = very easy, 7 = very hard).

Analysis of Required Knowledge, Tasks Selected Testing and Participant's Performance

In the following, the analysis of the knowledge required by the bathtub tasks and the rabbits-and-foxes task is presented, together with the tasks used to test for the different aspects identified, and the resulting performance. The results are summarized in Table 1 at the end of this section (page 14). Only frequencies of correct answers are reported, and no statistical tests have been performed on the answers to the questions. The number of participants, ten in each group, is too low to render any effects, other than really dramatic ones (of which there are none), significant. The study was more a kind pilot study to provide an initial test of the questionnaires. Therefore, the results need to be interpreted with some caution.

All tasks, with the solutions, can be found in the Appendix to this paper. I refer the reader to the Appendix for more detailed information about the tasks used, and a strongly recommend the reader too have the Appendix available when reading this paper. The tasks are not included as figures in the paper, because that would have split the text into small fragments scattered between large figures. The numbering of the tasks A1-A20 refers to their place in the Appendix.

The research questions concerned the following topics:

- What pieces of knowledge are required to solve the bathtub tasks and the rabbits-and-foxes-task?
- Which of these pieces belong to the regular high school mathematics curriculum?
- Can the performance of participants who have recently completed high school mathematics, in tasks tapping into the identified knowledge aspects, be predicted from what they have been taught in high school?

Reading and Drawing Graphs

Drawing graphs

One aspect addressed by Kainz and Ossimitz (2002), but not by Booth Sweeney and Sterman (2000), is whether people might experience trouble with the graphical representation. To receive a graphical presentation of a derivative of a function, and to use this information, to integrate it in order to produce the function of which it is the derivative, might be an unfamiliar task to the participants. Kainz and Ossimitz (2002) devised tasks that offered a written description of inflow and outflow of a bathtub during a specified time interval, with the task to depict the graph of the amount of water in the bathtub, or the water flows in and out of it. These tasks were used with slight modifications in the present study (Question A7: producing a flow graph from text, and Question A8: producing stock graphs from text). Kainz and Ossimitz (2002) found that it was significantly more difficult for the participating Austrian Business Administration students to depict the flows than to draw the water stock. This was expected to be true of our participants as well, because students are rarely asked to draw that kind of graphs in Swedish schools either. In addition, the easier bathtub task was inverted, so that instead of producing the stock from information about the flows, the participants were asked to produce a graph of the inflow from information about the stock and the outflow (Question A2). This was expected to be a more difficult task than the original task (Question A1), because it is more alien to tasks found in traditional schoolbooks in mathematics (e.g., Björk & Brolin, 2000a, 2000b). The inverted bathtub task (A2) was expected to be the most difficult one, followed by the original bathtub task (A1), producing flows from text (A7), and producing stock from text (A8), in descending order according to expected difficulty.

In accordance with the hypothesis, fewer participants succeeded in the inverted form of the simpler bathtub task (A2) than in its original form (A1). Five of ten were successful at the former, while eight of ten at the latter. Performance in Question A7, producing flow graphs from text, was as good as in the original easy bathtub task; eight of ten solved that task too. Seven of them were among the eight successful in Question A1, the easy bathtub task. Performance in Question A8, producing stock from text, however, was not as good as expected. Five of ten solved that task, in other words, as many as who solved the inverted bathtub task (A2). Four of these were among the five successful at the inverted bathtub task (A2). The low performance in Question A8, producing stock from text, can probably be explained by a mistake made when the task was constructed. As can be seen in Appendix A8, the author miscalculated the maximum amount of water in the tub, making the stock go way above the provided graphing area. This, in all likelihood, added to the difficulty of the task, but it was, however, solved by half of the participants in the group who received it.

Producing a flow graph from graphical stock information was more difficult for the participants than producing a stock graph from graphical flow information, and than producing flow graphs from written information. People are more familiar with written information, and reading graphs demands some additional effort even from our group of participants.

That producing a graph of the inflow from information about the stock (and outflow) would be more difficult than producing a stock graphs from information about the flows was expected from how these matters are taught in the Swedish high school (Björk & Brolin, 2000a, 2000b). The participants were, however, quite successful in producing flow graphs from text, which was a little unexpected. This is not much practiced in the Swedish high school. It was also the case that participants successful at producing stock graphs tended to be successful at producing flow graphs as well, and vice versa.

The low performance in the task of producing a stock graph from text I attribute to unfortunate task construction.

Reading graphs

The results from the simulation of the rabbits-and-foxes system are presented in line graphs, as results usually are from system dynamics simulations. Hence, the rabbits-and-foxes task requires the ability to interpret and combine information in line graphs. Questions A14-A17 tests the ability to extract information from the combination of a line graph describing the evolution of the rabbit population during a specified time period and a line graph depicting the development of the fox population during the same time period. How the two populations are related to one another is described in qualitative terms, not specifying the exact values of the parameters involved. Question A14 shows an increasing rabbit population and a fox population that first declines moderately and then grows to about the initial size. Question A15 depicts decreasing rabbits and decreasing foxes, and Question A16 illustrates the equilibrium situation with both populations constant. In Question A17, the rabbits first increase while the number of foxes also grows, due to the abundance of food. Eventually, the foxes are numerous enough to turn the rabbit population from increasing to decreasing. The participants were asked to describe what happened to the rabbits and foxes during the pictured time interval, and why.

Earlier results suggest that most people recognize the equilibrium situation in A16 (Jensen, 2003, 2005; Jensen & Brehmer, 2003). The participants were expected to have little trouble describing the behavior of the populations in A14 or A15, but would they combine the information in the two separate graphs, and relate it to one another, in their explanations?

Social science students, who have participated in our previous studies, tend to find it very difficult to produce a reasonable explanation to the development depicted in A17 (Jensen, 2003, 2005; Jensen & Brehmer, 2003). They seem not to manage to combine the behavior depicted in the graphs with the information provided in writing, in order to deduce something coherent. If they were, they would gain information useful for obtaining the goal of equilibrium populations.

All tasks in Questions A14-A17 require the ability to read graphs, but also the ability to combine and relate information in two separate graphs. Question A17, in particular, demands such ability. This is something not very much practiced in high school mathematics (e.g., Björk & Brolin, 2000a, 2000b). The performance was therefore expected to be weak in A17, better in A14 and A15, and rather good in A16. In A14 and A15, the descriptions relating the information in the graphs was expected to be less than perfect for most participants.

Questions A14-A17 produced the expected pattern of performance. All ten participants presented with the equilibrium situation in A16 correctly identified it as such, and four of the ten presented the first increasing then decreasing rabbit population together with an increasing fox population, in A17, gave a correct explanation to these graphs. Seven of ten gave a correct answer to A14, and six of ten to A15, which were both expected to be of moderate and similar difficulty. The explanations provided for the behavior of the rabbit and fox populations were actually better than expected. The information in both graphs was combined in the explanations given by the successful participants, and the success rate in Question A17 was somewhat higher than expected. Of the four successful in A17 only two solved A15, which also was a bit surprising.

Reading line graphs was not expected to be a problem to our participants. To ascertain this, a slightly modified version of the hotel problem developed by Kainz and Ossimitz (2002) was constructed, with the same information in tabular form, Question A13, and in a line graph, Question A12. The graph (A12) or table (A13) presents the number of guests arriving or departing each day during a two-week period. A12 a) and A13 a) asks when there are the most people staying at the hotel, A12 b) and A13 b) on what day most people arrive, and A12 c) and A13 c) when most people depart. No difference in performance was expected between the groups administered the different versions of the task. Kainz and Ossimitz (2002) reported a performance difference in favor of the tabular presentation, but the participants in the present study were expected to be as familiar with line graphs as with tables.

Questions A12 b), A12 c), A13 b) and A13 c) test only for simple point reading in a graph (or a table), while A12 a) and A13 a) require combining information in two graphs (or tables). It is therefore reasonable to expect performance in A13 a) and, in particular A12 a), to be related to performance in Questions A14-A17. Reading a point value in a single graph was hardly expected to be a problem to our group of participants. They were all expected to produce correct answers to A12 b), A12 c), A13 b) and A13 c). Combining information in graphs is not much practiced in high school (e.g., Björk & Brolin, 2000a, 2000b), so they might experience problems doing so, and then, consequently, in Questions A14-A17 as well. This reasoning ought to apply to information presented in tables as well.

In Questions A12 and A13, performance was a little weaker in the line graph condition (A12) than in the table condition (A13). This was not expected for the participants in the present study, but conforms to the results reported by Kainz and Ossimitz (2002). Performance was however high in the b), eight (A12) and nine (A13) of ten, and c), all ten (A12) and nine of ten (A13), parts of Questions A12 and A13, while, as expected, weaker in A12 a) and A13 a), four (A12) and six (A13), of ten for each questionnaire, gave a correct answer.

In Group X, there was no clear relation between performance in A12 a), most people in hotel (line graph), and A14, increasing rabbits and little change in foxes. Only two of the four successful in A12 a) solved A14, and only three of the seven successful in A14 solved A12 a). As mentioned above, all participants in Group X solved A16, and those who solved A12 a) solved A12 b) and A12 c) as well. There were about as many of those who succeeded at A12 b) and A12 c), most guests arriving and departing, who solved A14 as who did not solve A14.

In Group Y, there was an overlap in performance in Questions A13 a) and A15, and Questions A13 a) and A15. Five of the six successful in A13 a), most people in hotel (table), solved A15, decreasing rabbits and foxes, and five of the six successful in A15 solved A13 a). Of the four successful in A17, the first increasing and then decreasing rabbit population, three solved A13 a). Questions A13 b) and A13 c) were excluded from this analysis, since almost everyone in Group Y solved these tasks.

The participants were quite apt at merely reading information from graphs. They were somewhat less good at combining information in graphs and interpreting the message provided. They were however, and not very surprisingly, better at this than the Social Science students who participated in our prior studies (Jensen 2003, 2005; Jensen & Brehmer, 2003).

Calculating Bathtub Behavior

Booth Sweeney and Sterman (2000) analyzed both their bathtub tasks in terms of the kind of understanding they demanded. This was also used as performance criteria. What they did not do was to relate this required understanding to what is taught in mathematics classes in schools.

If we begin by looking at the easier bathtub task, the one generally solved by a majority of the participants (A1), Booth Sweeney and Sterman (2000) used the following seven criteria for judging performance.

1. When the inflow exceeds the outflow, the stock is rising.
2. When the outflow exceeds the inflow, the stock is falling.

3. The peaks and troughs of the stock occur when the net flow crosses zero (i.e., at $t = 4$, 8, 12, 16).
4. The stock should not show any discontinuous jumps (it is continuous).
5. During each segment the net flow is constant so the stock must be rising (falling) linearly.
6. The slope of the stock during each segment is the net rate (i.e., ± 25 units/time period).
7. The quantity added to (removed from) the stock during each segment is the area enclosed by the net rate (i.e., 25 units/time period • 4 time periods = 100 units, so the stock peaks at 200 units and falls to a minimum of 100 units).

For the difficult bathtub task (A18), Criteria 1, 2, and 4, are the same, while Criterion 3 for the difficult bathtub task is:

3. The peaks and troughs of the stock occur when the net flow crosses zero (i.e., at $t = 2$, 6, 10, 14)

Criterion 5 for the difficult bathtub task states:

5. The slope of the stock at any time is the net rate: Therefore:
   a. When the net flow is positive and falling, the stock is rising at a diminishing rate ($0 < t \leq 2$; $8 < t \leq 10$).
   b. When the net flow is negative and rising, the stock is falling at a decreasing rate ($2 < t \leq 4$; $10 < t \leq 12$).
   c. When the net flow is negative and rising, the stock is falling at a decreasing rate ($4 < t \leq 6$; $12 < t \leq 14$).
   d. When the net flow is positive and rising, the stock is rising at an increasing rate ($6 < t \leq 8$; $14 < t \leq 16$).

Criteria 6 and 7 together correspond to Criterion 6 for the easier bathtub task:

6. The slope of the stock when the net rate is at its maximum is 50 units/period ($t = 0$, 8, 16).
7. The slope of the stock when the net rate is at its minimum is -50 units/period ($t = 4$, 12).

And, Criterion 8 corresponds to Criterion 7 in the easy bathtub task.

8. The quantity added to (removed from) the stock during each segment of two periods is the area enclosed by the net rate (i.e., a triangle with area ± ½ • 50 units/period • 2 periods = ± 50 units). The stock therefore peaks at 150 units and reaches a minimum of 50 units.

Criterion 3, for the difficult bathtub task (A18) demands that you know that the derivative is zero in extreme points, the peaks and troughs of the function. For the easier bathtub task (A1), this kind of knowledge does apply, but you can easily do without it. This is something that is introduced rather late in the Swedish high school (e.g., Björk & Brolin, 2000b). It was therefore expected that most participants in our study would find this difficult. There was no task in the questionnaire addressing this kind of understanding specifically, but if you master Criterion 5 you will probably also master Criterion 3, but this, of course, needs to be empirically confirmed.

The major difference in difficulty between the two bathtub tasks is to be found in Criterion 5. This is also supported by results (Both Sweeney & Sterman, 2000; Kapmaier, 2004). Generally speaking, all criteria pose higher demands in the difficult bathtub task than in the easier one. It is Criterion 5, however, that particularly demands an insight into the relation between a function and its derivative, or a function and its corresponding primitive functions, i.e., its integral. Question A20 intended to test this kind of understanding with a more traditional mathematics task. In A20 a) the task is to calculate the velocity after one minute of a vehicle accelerating constantly by 2 m/s^2 from standstill, and A20 asks what distance the vehicle will have traveled then (after having accelerated for one minute). This, at least A20 b), requires an understanding of nonlinear behavior, such as exponential growth, which was tested by Question A3. A3 uses a classic example and asks, in A3 a), when half a pond will be covered by water lilies that double every day if it is completely covered in 30 days, and A3 b) asks when a quarter of the pond was covered. Exponential growth is introduced early in high school mathematics (Björk & Brolin, 2000a) and should therefore not pose any trouble to the participating student group. This kind of understanding might also be related to a feel for the continuous aspect of the rabbits-and-foxes task (see below).

For the difficult bathtub task (A18) Criteria 6 and 7 are closely related to Criterion 5 and to Criterion 6 for the easier bathtub task (A1).

Criterion 8 for the difficult bathtub task (A18) is the parallel to Criterion 7 for the easier bathtub task (A1), and requires fairly similar reasoning. It concerns the principle of integration as area calculation, or a conceptual understanding of mathematical integration. Question A20 b) can be solved by integration as area calculation, but might as well be solved by algebraic integration. Integration, both approached by algebraic and by area calculation, are among the last subjects to be taught in the mathematics courses of the Swedish high school. Therefore, it was considered unlikely that more than perhaps a few participants would correctly solve these tasks.

It was hypothesized that anyone who solves the difficult bathtub task (A18) would also solve the water lilies problem (A3), Question A20 a) velocity and b) distance from acceleration, as well as the easier bathtub task in either form presented (A1 and A2). You may use integration as area calculation to solve the easier bathtub task, but (change in) *stock* = (constant) *flow* • *time* suffices, and is something students acquire even before they enter high school. This was also tested by Question A19, which ought to present no problem to any of the participants. It simply asks the participants to calculate the distance travelled by a car that travel at a constant speed that changes stepwise once during the specified time period (see Appendix A19).

The difficult bathtub task (A18), was solved correctly by four of the participants, two of whom also solved A20 a) and A20 b), a) velocity and b) distance from acceleration. It was the same two participants who solved both subtasks of Question A20. Distance from acceleration (A20 b) was only solved by those two just mentioned, while velocity from acceleration (A20 a) was solved by half (10) of the 20 participants. Thus, calculating distance from acceleration (A20 b) was not easier, as it was expected to be, than solving the difficult bathtub task (A18), and performance in the difficult bathtub task (A18) and calculating velocity and distance from acceleration (A20) was not closely related, in contrast with my hypothesis.

Question A3, the water lilies problem, both a) and b), was solved by 17 of the participants. Either, both A3 a) and A3 b) were correctly answered, or none of them, as was the case of the remaining three participants. The four who succeeded at the difficult bathtub task (A18) were among the successful 17.

All four who solved the difficult bathtub (A18) task also succeeded at the simpler bathtub task, either in the original version (A1) or the inverted version (A2), and, as expected, all participants successfully completed Question A19, the distance from constant speed problem. They had no trouble calculating stock = flow • time, at constant flow.

Criterion 1 and criterion 2 were not tested for in the present study. They were considered trivial to the participating student group. Students who have completed a program in science or technology in high school will most certainly be familiar with the concept of net flow, as the difference between inflow and outflow, and how net flow influences stock. If this assumption were unwarranted, it would betray itself in a remarkably low performance in these tasks.

Criterion 4 is closely intertwined with all the other criteria, and it is very hard to test for specifically in any meaningful way. Any suggested shape of the variations in stock, however bizarre, would be accepted as long as it was continuous. It seems few people make this mistake (Booth Sweeney & Sterman, 2000; Kainz & Ossimitz, 2002; Kapmaier, 2004), and if they do they are, in all likelihood, wrong in most other respects as well.

The participants had no problem calculating the accumulated stock from information about flow and time duration when the flow was constant, and most of them were familiar with exponential growth. Only a few were, however, able to figure out the behavior of the stock resulting from a non-constant flow (to perform either numerical or graphical integration).

Balancing Predator and Prey

In the rabbits-and-foxes task, the participants receive a full description of the relations describing the system: Every rabbit produces two offspring a year, and every fox eats 4% of the rabbits a year. For every 180 rabbits consumed a new fox is born, and 20% of the fox population dies each year.

A mathematical approach

Given this information it is possible to calculate the population sizes in equilibrium. Then you would know what to strive for. In previous studies (Jensen, 2003; Jensen & Brehmer, 2003) none of our participants proved able to correctly complete the necessary calculation. The participants in these studies had various backgrounds regarding choice of program in high school, and were therefore probably not quite as mathematically proficient as the participants in the present study. Questions A5 and A6 tested if the participants were able to create the equations necessary for the calculations, if induced to do so in a stepwise manner. In equilibrium there are equal numbers of rabbits born and rabbits eaten, and of foxes born and foxes dying. Questions A5 and A6 consisted of four parts a)-d). In a), the participants were asked to express in an algebraic expression (using R for the present number of rabbits, and F for the foxes) the birth rate of the rabbits (in A5) or of the foxes (in A6). In b) they were asked to express the death rate of the rabbits (A5) or the foxes (A6) in an algebraic expression.

Question A5 a) ought to be solved by the majority, and all should get A6 b) correct, while A5 b), and to a larger extent A6 a) might prove somewhat more difficult. All of the tasks were expected to be within the ability of the selected participant group. The expected order of difficulty, from easiest to hardest, were A6 b), A5 a), A5 b), and A6 a). The participants were familiar with the task of constructing equations from written descriptions. They might, however, find the expression describing the number of foxes born a bit complicated.

Questions A5 c) and A6 c) asks how many rabbits (A5, or foxes in A6) are born compared to how many that die, when in equilibrium. This would simply entail stating that the two expressions already produced were equivalent with each other, which was not expected to be a difficult task.

Question A5 d) and A6 d) asks if the information obtained in the previous steps could be used in any way to learn something about the rabbits and/or foxes in equilibrium and, if so, what and how. This might require some familiarity with solving systems of equations, arriving at values for the different parameters in a stepwise manner. This seems not to be particularly well practiced in high school, so this step might prove an obstacle to the participants.

The task of producing the algebraic expressions for the birth and death rates in Questions A5 and A6 produced the expected results, with some distortion that might be attributed to order effects. Question A6 a), the birth rate of the foxes proved by far the most difficult, with only one participant of ten giving the correct answer. The difficult first task probably disheartened the participants into an unexpectedly low performance in task A6 b), where only six of ten were able to produce the very simple answer. Six out of ten also gave a correct answer to the more complicated A5 b), probably after being encouraged by the easy A5 a), which seven of the ten answered correctly. Performance in the a) and b) tasks were somewhat, although not perfectly, related. Of those six successful in A5 b), five were among the successful seven in A5 a), and the only one successful in A6 a) failed A6 b).

Only five out of all twenty answered A5 c) (three) or A6 c) (two) correctly and stated that births equals deaths in equilibrium, and nobody was able to figure out A5 d) or A6 d).

The participants were able to produce simpler algebraic expressions, but not more complex ones, and setting up and solving a simple system of equations was clearly beyond their ability. All this conformed to the expectations, but that only a quarter of the participants were able to state that equilibrium means equal births and deaths, or a birth/death ratio of one, was a bit surprising.

Balance or equilibrium

There is also the aspect of understanding the concept of equilibrium or balance, which ought to be fully grasped by the participants in the present study. As already mentioned, Question A16, presenting line graphs with constant rabbit and fox populations, taps into this, and so does Question A4. In addition Question A4 assesses whether the participants are able to perform adequate calculations to balance births of rabbits with consumption by foxes (See Appendix A4). Questions A12 a) and A13 a), most guests in hotel, are also questions demanding an understanding where to find the equilibrium point, in addition to the requirement of being able to combine information in graphs. Question A16 requires only recognition of the equilibrium situation, Questions A12 a) and A13 a) demands reading off and combining information in graphs or tables, while Question A4 requires calculations to be performed. I expected performance in these tasks to be related and the order of difficulty form easier to harder: A16, A12 a) and A13 a), and A4. All these tasks were expected to be quite easy for our participants.

Question A4, calculating the rabbit population necessary to compensate for the number killed by foxes, was correctly solved by 11 of the 20 participants. This task was no more difficult than Question A12 a) and A13 a) that received 10 of 20 answers correct. There was not a big overlap in performance, however. Only six of those successful in Question A4 were successful in Question 12 a) or 13 a). As mentioned earlier, all participants gave correct answers to A16.

The participants had no problem recognizing an equilibrium situation presented to them. Calculating the number of rabbits needed to produce enough offspring to compensate for the number killed by foxes, or realizing that arrivals equals departures when the lines intersect, were accomplished by half (although not entirely the same half) of the participants. This was a little weaker performance than expected.

The continuous aspect

An important factor in the solution of the rabbits-and-foxes task is to understand the continuous aspect of the system. Both populations, rabbits and foxes, evolve continuously during the year (roughly simulated by calculating the development in one-month time-steps in the simulation). Participants in previous studies, however, frequently made the mistake of assuming the simulation to evolve in discrete one-year time-steps (Jensen, 2003; Jensen & Brehmer, 2003). Booth Sweeney and Sterman (2000) have named this behavior spreadsheet thinking.

I found no suitable way to test for this understanding in isolation, using the rabbits-and-foxes example, but an understanding of the continuous, non-linear development of the predator and prey populations, ought to be related to an understanding of exponential growth, tested by Question A3 (the water lilies problem), by Question A20 b) (calculation distance from acceleration), and by the difficult bathtub task (Question A18) (results reported in the Calculating Bathtub Behavior section above). It should also be related to the ability to give nuanced answers to Questions A14 and A15 and a comprehensive explanation to Question A17 (combining information in a rabbit and a fox graph; results reported under Reading graphs above).

Of the four successful in the difficult bathtub task, all were successful in A14, A15, or A17 (combining rabbit and fox graphs), while among the three failing A3 (the water lilies problem) one succeeded in A14, increasing rabbits and little change in foxes, and one, quite surprisingly, in A17, the first increasing and then decreasing rabbit population.

The participants, or at least about half of them, could be said to have a feeling for the continuous aspect. They told a coherent story in time when describing the interactions of the rabbits and foxes depicted in line graphs in Questions A14, A15, and A17.

The control aspect

To control an evolving process is not part of the mathematics curriculum of the Swedish high school (e.g., Björk & Brolin, 2000a, 2000b). It might be introduced in technology classes, but hardly beyond the utterly basic. Direct control, like regulating the heat of the stove when cooking, is part of everyday life, and as far as no substantial delays are part of the process, and the output is reasonably linearly related to the input, people are quite adapt at learning to perform appropriate input (e.g., Crossman & Cooke, 1974; Moray, 1987). We found, in our earlier studies (Jensen, 2003, 2005; Jensen & Brehmer, 2003), that the participants met an increasing rabbit population by increasing the fox population, and a declining rabbit population by reducing the foxes. Question A9 was intended to make sure that our participants mastered this, and to introduce them to the tasks following.

Questions A9-A11 introduce the rabbits-and-foxes system with the general description, including parameter values. In Question A9, the task for the participants was to keep the rabbit population within certain limits, by adjusting the fox population to any size considered suitable. They were presented with the population sizes one and two years ago, together with the present numbers. They were asked to decide on the appropriate size of the fox population in the present situation. The rabbits were more numerous than desired, so any request for a larger fox population (within reasonable limits) would pass as a correct answer. All participants were expected to be able to produce a correct solution.

All participants performed A9, while Group X received A10 and Group Y received A11.

In A11, the participants were again asked to decide on the desirable number of foxes based on information about the present situation as well as the situation one and two years ago. This time the rabbit population had been drastically reduced. Rules had changed, however, so that hunting now regulated the fox population. This meant that the only way to increase the fox population, once reduced by hunting, was to keep the puppies born. The participants were asked about what size of the fox population they desired under the present circumstances (A11 a), and if there was something they particularly needed to keep in mind with the new means of fox control (A11 b).

Moxnes (1998; Moxnes & Saysel, 2004) has demonstrated that people fail to, for example, cut the number of grazing animals below the equilibrium level when pastures have been too much reduced. This is true even for participants who are able to calculate the equilibrium level or who are explicitly informed about the equilibrium level.

In A10, the participants were instead presented with the goal of the original rabbits-and-foxes task, to make the populations reach an equilibrium state. They were then asked to focus on the fox population, and to think of explanations for a situation where more foxes were born than died (A10 a). They were asked to suggest ways to reduce the number of fox births, and if there were more than one way to achieve that (A10 b), and also which would be the preferable alternative with the goal to achieve equilibrium in mind (A10 c).

The births of foxes can be reduced either by reducing the fox population or by reducing the rabbit population (less food for the foxes). If the fox population is reduced when the rabbit population is large enough to sustain a large fox population, the rabbit population will increase dramatically. The preferred solution for approaching an equilibrium situation is therefore the alternative to initially increase the fox population to reduce the rabbit population, and then reduce the fox population the level considered appropriate.

This has proven to be the fundamental stumbling block for people performing the rabbits-and-foxes task. Indirect reasoning of the kind required is nothing that is practiced in the Swedish high school (e.g., Björk & Brolin, 2000a, 2000b), at least not as a part of the mathematics curriculum. People generally experience trouble with performing indirect reasoning (Evans, Clibbens & Rood, 1995; Jensen, 2003). Therefore weak performance was expected in this task.

Questions A9-A11 seems to have been very confusing to the participants. Only 12 out of 20 gave the correct answer to Question A9. Of the remaining eight, seven refrained from answering at all. Of those seven, six belonged to Group Y. To Question A10 b), suggesting fewer rabbits is only a partially correct answer, but anyone making such a suggestion was graded as correct, as did nine of the ten in Group X, but nobody got A10 c) right. Eight in Group X could explain why the foxes grew more numerous (Question A10 a).

One participant, however, who gave the correct answer to A10 b), that one could either reduce the rabbit population or the fox population, and suggested in A10 c) that one should do a bit of both until one obtained even proportions. It is to some extent in the right direction, although not entirely so.

Only three, of the ten in Group Y, gave an answer to A11 a). They were all correct, and one of them was among those not offering answers to A9. Those who did not answer A11 a) offered no answer to A11 b) either, and no one of the remaining three gave a correct answer to A11 b).

The results do not allow any conclusions regarding the participant's grasp of the control aspect. The tasks intended to address this, Questions A9-A11, were clearly misinterpreted by a large number of the participants, leaving the results obtained rather meaningless.

Summary of Results

The results are summarized in Table 1 below.

Table 1. Results. The number of correct answers to the questions is presented (within parentheses the number of participants who were administered the questions). All questions were graded as either correct or incorrect.

Question	**Group**	**Description**	**Performance**	**Percentage**
A1	X	The easy bathtub task	8 (10)	80
A2	Y	The inverted easy bathtub task	5 (10)	50
A3 a)	X+Y	Exponential growth	17 (20)	85
A3 b)	X+Y	- “ -	17 (20)	85
A4	X+Y	Compensate for rabbits killed	11 (20)	55
A5 a)	X	Mathematically balancing rabbits	7 (10)	70
A6 a)	Y	- “ - foxes	1 (10)	10
A5 b)	X	- “ - rabbits	6 (10)	60
A6 b)	Y	- “ - foxes	6 (10)	60
A5 c)	X	Mathematically balancing	3 (10)	30
A6 c)	Y	- “ -	2 (10)	20
A5 d)	X	- “ -	0 (10)	0
A6 d)	Y	- “ -	0 (10)	0
A7	X	Hugo's bath flows	8 (10)	80
A8	Y	Hugo's bath stock	5 (10)	50
A9	X+Y	Fox control - transport	12 (20)	60

A10 a)	X	Fox control - balance	8 (10)	80
A10 b)	X	- “ -	9 (10)	90
A10 c)	X	- “ -	0 (10)	0
A11 a)	Y	Fox control - hunting	3 (10)	30
A11 b)	Y	- “ -	0 (10)	0
A12 a)	X	Hotel – line graph	4 (10)	40
A13 a)	Y	Hotel - table	6 (10)	60
A12 b)	X	Hotel – line graph	8 (10)	80
A13 b)	Y	Hotel - table	10 (10)	100
A12 c)	X	Hotel – line graph	7 (10)	70
A13 c)	Y	Hotel - table	9 (10)	90
A14	X	Rabbit-fox growth	7 (10)	70
A15	Y	Rabbit-fox decline	6 (10)	60
A16	X	Rabbit-fox equilibrium	10 (10)	100
A17	Y	Rabbit- fox “bump”	4 (10)	40
A18	X+Y	The difficult bathtub task	4 (20)	20
A19	X+Y	Distance from speed	20 (20)	100
A20 a)	X+Y	Velocity from acceleration	8 (20)	40
A20 b)	X+Y	Distance from acceleration	2 (20)	10

Both questionnaires were considered to be rather difficult by the participants, with the Y questionnaire as the more difficult of the two. The mean grade for the X questionnaire were 4.7 (s = 1.4), and for the Y questionnaire 5.8 (s = 0.8); $t_{(18)} = 2.14$, $p < .05$.

The participants spent 1 hour and 10 minutes on average on the questionnaires, ranging from half an hour, for the fastest one, to two full hours, the maximum time allowed, for the participant requiring the most time.

It appears as if the questionnaires are not too time consuming, so a few more questions could be added. The participants perceived the questions to be difficult, although not unreasonably so. A closer look at how the questions are divided into the two different questionnaires seems to be needed, because the Y questionnaire was experienced as significantly more difficult than the X questionnaire.

Questions Remaining

Concerning tasks

For reading and drawing graphs the stock from text task (Question A8) needs to be changed so the result fits within the graph area. Otherwise I see no need for anything more.

Tasks addressing the understanding of a zero crossing of the derivative (the peaks and troughs of a function) need to be added, and so do tasks testing more specifically the understanding of the relation of between a function and its derivative, and its integral or primitive function(s). Particularly, tasks are needed where the derivative is non-constant, when it is increasing or decreasing (constantly or at an increasing or decreasing rate).

There is also a need for tasks that address the ability to perform algebraic and graphical integration and that separates between these two abilities.

The task testing the ability to construct algebraic expressions, and combine them into a system of equations (Questions A5 and A6) needs to be supplemented with some tests of the ability to combine equations, to combine them to form systems of equations, and to solve single equations as well as systems of equations.

To test for the understanding of a continuously evolving process, adding examples from other domains than the ecological would perhaps be desirable.

The tasks addressing the control aspect are certainly in need of improvement. It is a difficult aspect to address with paper-and-pencil tests, but it should at least be possible to do better than this study. It would, however, probably be better approached by simple simulation tasks.

Concerning participant groups

When a reasonably well-working set of tasks has been devised, it needs to be administered to other participant groups as well, such as undergraduate, as well as graduate, students in mathematics, engineering, and system dynamics to learn about what circumstances that are beneficial for acquiring the different knowledge elements. Students may participate in system dynamics classes, and learn to build systems dynamics models, for example, and still be lacking in the understanding of basic bathtub tasks (see, for example, Biber & Kasperidus, 2004), and there might be elements that traditional educations are quite successful at teaching.

Concerning education

In the present study, hypotheses about the participants' prior knowledge were based solely on studies on high school mathematics textbooks. It would, in all certainty, be beneficial to interview teachers about how they teach the identified knowledge elements, successful strategies they have identified, and what their students tend to find particularly demanding.

Conclusions

We do not know what kind of knowledge that is required to understand dynamic systems and system dynamics that people learn or do not learn in schools.

We do not know what questions to ask and we do not fully understand the answers we receive (or fail to receive).

More research effort is needed addressing these questions, or how else are we to improve education or information about dynamic systems?

Acknowledgements

The Swedish Armed Forces Research and Development Program supported this research. I thank Dr. Berndt Brehmer for his valuable comments on this paper. I also thank Johanna Brehmer for recruiting the participants and administering the questionnaires.

References

Biber, P., & Kasperidus, H. D. (2004). Integrated modeling approaches and system dynamics in education related to sustainable resource management, forestry, and land use management. *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England.

Björk, L-E., & Brolin, D. (2000a). *Matematik 3000. Kurs A och B lärobok, Naturvetenskap och teknik [Mathematics 3000. Course A and B textbook, Science and technology]*. Stockholm: Natur och Kultur.

Björk, L-E., & Brolin, D. (2000b). *Matematik 3000. Kurs C och D lärobok, Naturvetenskap och teknik [Mathematics 3000. Course A and B textbook, Science and technology]*. Stockholm: Natur och Kultur.

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

Crossman, E. R. F. W., &Cooke, J. E. (1974). Manual control of a small-response system. In E. Edwards, & F. P. Lees (eds.), *The human operator in process control, pp. 51-66*. London: Taylor and Francis.

Evans, J. St. B. T., Clibbens, J., & Rood, B. (1995). Bias in conditional inference: implications for mental models and mental logic. In W. Schaeken & G. De Vooght (eds.), D*eductive reasoning and strategies* (pp. 1-22). Mahwah, NJ: Erlbaum

Fisher, D. M. (2003). Student performance on the bathtub and cask flow. *Proceedings of the 21st International Conference of the System Dynamics Society*. New York City, USA.

Kainz, D., & Ossimitz, G. (2002). Can students learn stock-flow thinking? An empirical investigation. Proceedings of the 20th International Conference of the System Dynamics Society. Palermo, Italy.

Kapmaier, F. (2004). Findings from four years of bathtub dynamics at higher education institutions in Stuttgart. *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England.

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

Jensen, E. (2003). (Mis)understanding and learning from feedback relations in a simple dynamic system. Doctoral dissertation. *Örebro Studies in Psychology, 3*. Örebro University, Sweden.

Jensen, E. (2005). Learning and transfer from a simple dynamic system. *Scandinavian Journal of Psychology, 26*, 119-131.

Lyneis, J. M., & Lyneis, D. A. (2003). Bathtub dynamics at WPI. *Proceedings of the 21st International Conference of the System Dynamics Society*. New York City, USA.

Moray, N. (1987). Intelligent aids, mental models, and the theory of machines. *International Journal of Man-Machine Studies, 27*, 619-629.

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

Moxnes, E., & Saysel, A. K. (2004). Misperception of global climate change: information policies. *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England.

Ossimitz, G. (2000). *Entwicklung systemischen Denkens. Theoretische Konzepte und empirische Untersuchungen [Developing systems thinking. Theoretical concepts and empirical investigations]*. München: Profil Verlag.

Ossimitz, G. (2002). Stock-flow thinking and reading stock-flow related graphs: an empirical investigation in dynamic thinking abilities. Proceedings of the 20th International Conference of the System Dynamics Society. Palermo, Italy.

Zaraza, R. (2003). Bathtub dynamics in Portland at SYMFEST. Proceedings of the 21st International Conference of the System Dynamics Society. New York City, USA.