DOES THE MODIFICATION OF THE REPRESENTATION FORMAT AFFECT STOCK-FLOW-THINKING?

Friederike Brockhaus, Jana Arnold, Marcus Schwarz and Peter Sedlmeier

Chemnitz University of Technology, Germany

Address for correspondence:

Friederike Brockhaus

Department of Psychology

Chemnitz University of Technology

09107 Chemnitz

Germany

Email: friederike.brockhaus@psychologie.tu-chemnitz.de

Phone: ++49-371-53137111

Fax: ++49-371-5318 37111

Abstract

Stock–flow systems are poorly understood even though we are faced with them in many, sometimes very important domains. Several studies found that only few participants solved stock–flow tasks correctly and that modifications of the representation format (e.g. changing the flow graphs) did not improve performance. In the present study, we modified the representation in a more extreme way. Because it seems people have difficulty distinguishing between stocks and flows, we wanted to simplify the transformation of flows into stock. We created teaching pictures where the flows were presented as stocks per time unit. We analyzed if such a representation improved stock–flow performance. Participants had to solve three tasks in different representation formats (tub, bus, and line graph). Regression analyses revealed differences in performance. The modified representation conditions led to a better stock–flow performance (R^2 = .389, partial correlations: .557, $p < .001$ (tub), .498, $p < .001$ (bus)). To test for transfer, all participants had to work on two additional tasks presented as line graphs. The modified representation groups did not solve the additional tasks better than the line-graph group. To summarize, the teaching pictures simplified solving stock–flow tasks but transfer did not occur.

Keywords:

stock–flow, teaching pictures, representation format, problem solving, dynamic systems

Introduction

Stock–flow systems are omnipresent and are very important in our lives. A bank account is a stock-flow system consisting of amount of money, the stock, depending on receipts, the inflow and expenditures, the outflow. Another important stock–flow is the amount of atmospheric CO2 depending on naturally and anthropogenic CO2, and on the absorption of CO2. Even the simplest stock–flow system, consisting of one stock, an inflow, and an outflow, is very difficult to understand. When students were asked to infer the stock pattern from two flow patterns, they usually failed, even if they had a background in natural sciences or economics (Booth Sweeney & Sterman, 2000). In quite a number of studies analyzing stock–flow performance, many participants failed to solve the tasks correctly (e.g., Cronin, Gonzalez, & Sterman, 2009; Ossimitz, 2002). Figure 1 shows one example of a stock–flow task and Figure 2 shows the corresponding solution. Sixty-four percent of the highly educated students in Booth Sweeney and Sterman's (2000) study failed to solve the task correctly.

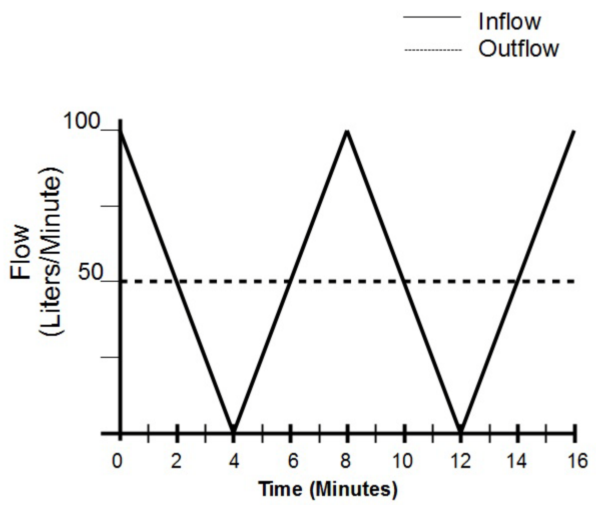

Figure 1: The triangle task (slightly modified from Booth Sweeney & Sterman, 2000).

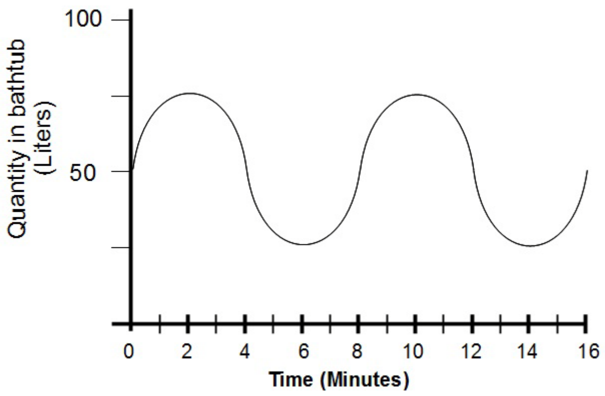

Figure 2: The correct solution corresponding to the triangle task drawn on the solution diagram (slightly modified from Booth Sweeney & Sterman, 2000).

We attempted to simplify the representation of the stock–flow tasks as much as possible to support participants in finding the correct solution. Ossimitz (2002) suggested that the main difficulty in solving stock–flow tasks lies in discerning between stocks and flows. He argued that the flows might be seen as stocks. One systematic error, possibly related to this missing distinction between stocks and flows, is described as the correlation heuristic (e.g., Cronin et al., 2009). If participants follow this heuristic, they (wrongly) expect the stock to follow the same pattern as one of the flows.

There is convincing evidence that our cognitive system is dependent on external representation formats. External representations possess "affordances" that lead to spontaneous or intuitive actions/behavior or judgments/decisions (Norman, 1988). Norman provided the example of the arrangement of light switches in a large room. Even people who often use the light switches turn on the "wrong" lights if the switches are not analogously arranged to the position of the lights. A well-known example of the influence of representation format on problem solving is the Wason Selection Task, which involves logical reasoning (Wason & Shapiro 1971). Participants are given several cards with letters on one side and numbers on the other. One possible rule is "if there is a consonant on one side then there must be an even number on the other side." Four cards are given, for example, "D" "E" "5" "4." Only 10–20% choose the right cards ("D" and "5") to properly test the rule. Yet, if the cards were modified to an everyday context, the results change. The new rule was "only people over the age of 21 are allowed to drink alcohol" and the cards contained on one side people of different ages, for example, an old man and a child, and on the other side drinks, for example, juice and beer. Now the cards with the child and the beer were spontaneously chosen to check if the rule was violated (e.g., Cosmides, 1989). So, the content of the "cover story," a special variation of the representational format of the task, influences the correctness of judgments.

Judgments of probability represent another domain where the representation format influences performance (Gigerenzer, Hertwig, Hoffrage, & Sedlmeier, 2008). The percentage of correct solutions increases if the information concerning the probabilities is presented as "natural frequencies" (e.g., Hoffrage & Gigerenzer, 1998). To summarize, an external representation format can enhance or inhibit correct judgments or actions. Therefore we attempted to change the representation format of the stock–flow tasks in a way that would simplify finding the correct solution.

Previous attempts to change the representation of stock–flow tasks were not successful, or only slightly so (Brockhaus & Sedlmeier, unpublished document; Cronin et al., 2009). If a central problem in solving stock–flow tasks really lies in participants' difficulty differentiating between stocks and flows, it makes sense to make the transition between the two more salient. In the present experiment, we tried to transform the flow into a "temporary stock." This can be seen as a sum of all entities in one time unit, such as all the water in 1 minute that flows into a tub. Thus, the flow can be seen as a stock for one time unit. A first step in simplifying the stock–flow representation in this way is to use bar graphs. In a bar graph, all entities in one time unit are summarized. Zelazny (1985) emphasized that it is better to use bar diagrams than line graphs when entities of discrete frequencies are shown. But research showed that representing the flows as bar graphs only slightly improved stock–flow task performance (Cronin et al., 2009). Therefore we chose a pictorial representation format to transform the flows into small stocks.

Pictograms are an often used representation format. Otto Neurath published a book back in 1936 about a picture language (called ISOTYPE) for teaching, especially for teaching statistics. He highlighted that the picture language would be especially appealing to people who dislike numbers and graphs. He criticized the use of line graphs for frequencies of discrete entities because the lines misled the reader. For example, if people entering and leaving a hotel on several days were depicted, the lines look as if for each hour, minute, or even second a number of people entering or leaving the hotel were shown. Neurath's picture language consisted of simple pictures without unnecessary detail. This is consistent with advice from multimedia learning. For instance, Domagk (2008) reported many positive functions of pictures in the field of multimedia learning. She claimed that adding pictures to learning material can support the learner; for example, pictures can direct the attention to relevant information. She referred to Dwyer (1975), who emphasized that to reach optimal learning success, a picture should not include too many details and it should not be too simple. In a previous study, we used additional schematic pictures of a stock–flow scenario, such as a picture of a bathtub, to support stock–flow performance (Brockhaus & Sedlmeier, unpublished document). In the present study, we went further and not only added a schematic picture of the scenario but also "transformed" the flows and the stock into such a pictorial representation. By using this kind of teaching picture, the time flow was broken into small units, so that participants would be encouraged to analyze the flows in small steps. The new flow representation was discrete, regardless of whether the entity possessed a discrete quality (such as people) or a continuous quality (such as water). We constructed the pictures so that they corresponded to Booth Sweeney and Sterman's (2000) classic line graphs as closely as possible.

In addition to the influence of the representation format, we analyzed if characteristics of the participants influenced the stock–flow task performance. We were especially interested in the impact of participants' sex and educational background. Many studies found an effect in favor of men (e.g., Booth Sweeney & Sterman, 2000; Brockhaus & Sedlmeier, unpublished document; Ossimitz, 2002; Schwarz & Sedlmeier, unpublished document; Sedlmeier, Brockhaus, & Schwarz, in press). And we compared participants studying natural sciences with participants studying social sciences. We had previously found systematic differences in favor of students of natural sciences (Brockhaus & Sedlmeier, unpublished document). Also studies revealed correlations between grades in mathematics and stock–flow task performance (Schwarz & Sedlmeier, unpublished document): The better the grades, the better the performance.

In further analyses we looked at transfer. If the new representation was helpful to solve stock–flow tasks, it was important to determine if transfer to the classic (line graph) representation occurred. Transfer is often difficult to achieve (for an overview see Reed, 1999). Also in the domain of probability tasks, transfer has not easily been found (Sedlmeier, 1998; Sedlmeier, 2000). Therefore, all participants had to perform stock–flow tasks with line graph representations to test transfer performance.

The central question addressed in the present study was whether a modification of stock–flow tasks that treats flows as "temporary stocks" is able to noticeably improve task performance. In addition, we were interested in the moderating role of sex and level of formal education. Finally, we wanted to find out whether after the successful solution of the task in the new format, participants could transfer this knowledge to solving tasks represented in the conventional format.

Method

Participants

Ninety students (mean age: 22.6 years, *SD* = 4.2, range: 19–46 years, 36 males and 54 females) studying at the Technical University Chemnitz participated in the experiment. Participants were paid or participated for course credit. Twenty-nine students studied a subject in or near the natural sciences.[1]

Materials

We used a paper-and-pencil test. Our tasks had already been used in earlier studies. We used the square wave task and the triangle task, both with 12 instead of 16 time units, both tasks modified from Booth Sweeney & Sterman (2000), and one task with discontinuous flows, similar to the discontinuous task in Cronin et al. (2009) with 12 instead of 30 time units. Figure 3, Figure 4, and Figure 5 depict the tasks (as they were shown in the baseline condition) we used in our experiment. As task context we used water in a tub and passengers entering and leaving a bus.

[1] Subjects participants studied were divided into two groups. One group consisted of subjects in or near the natural sciences, including mathematics. The other group consisted of subjects in the social sciences or subjects including less mathematical courses. Examples of subjects in or near the natural sciences: electrical engineering, engineering economics, (financial) mathematics, mechatronics, physics. Examples of subjects in or near the social sciences: economics, educational sciences, German philology, sociology, political science, psychology. Division was based on the "faculties" (schools/colleges) of the Technical University Chemnitz. The examples include only subjects that were studied by more than two participants.

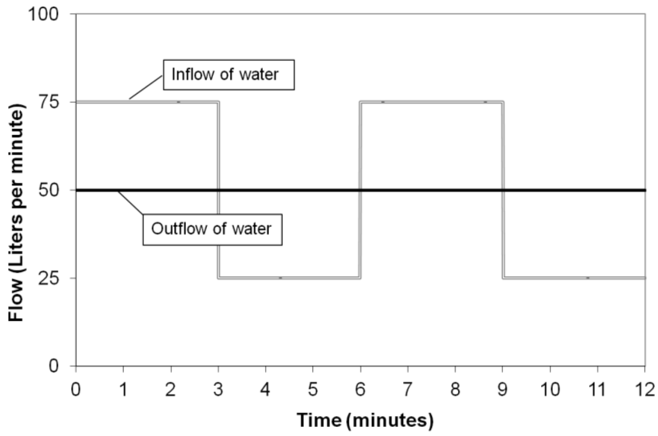

Figure 3: The square wave task in the baseline condition (modified from Booth Sweeney & Sterman, 2000).

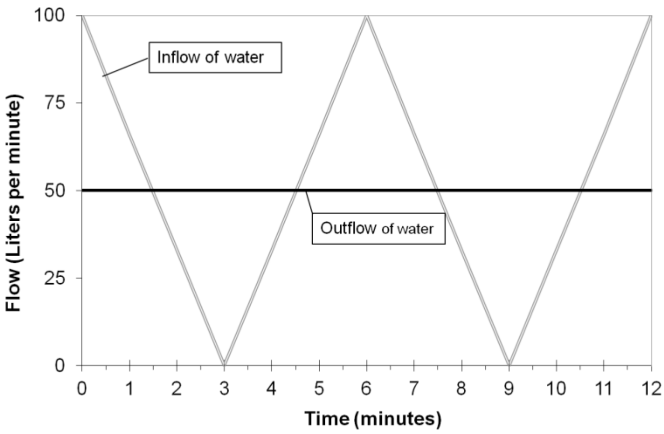

Figure 4: The triangle task in the baseline condition (modified from Booth Sweeney & Sterman, 2000).

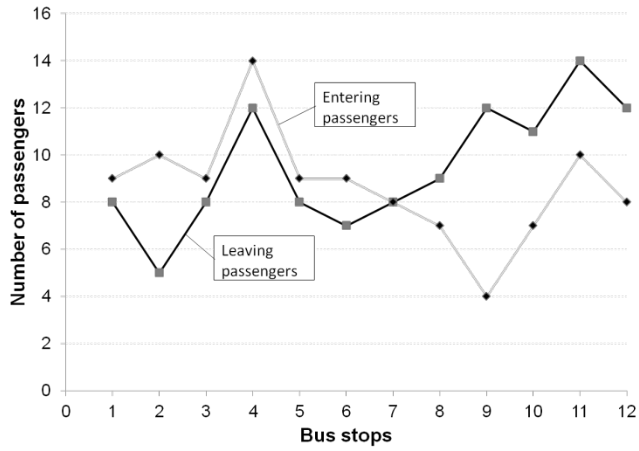

Figure 5: The discontinuous task in the baseline condition (modified from Cronin et al., 2009).

To test if transfer occurred, we added a continuous task taken from Cronin et al. (2009; Figure 6) and a discontinuous task from Ossimitz (2001; Figure 7), again using water and passengers as contexts.

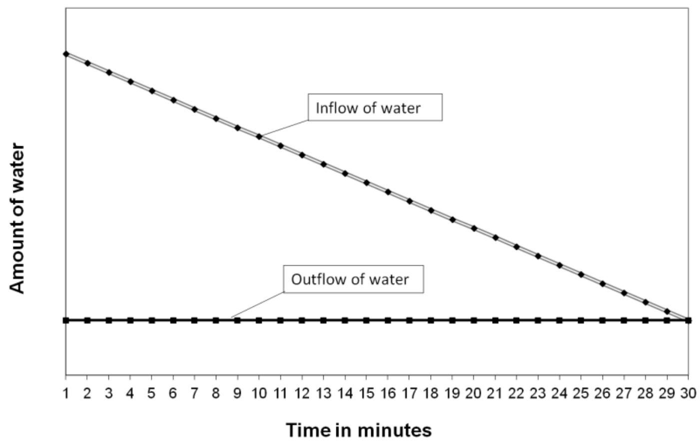

Figure 6: The continuous transfer task (modified from Cronin et al., 2009).

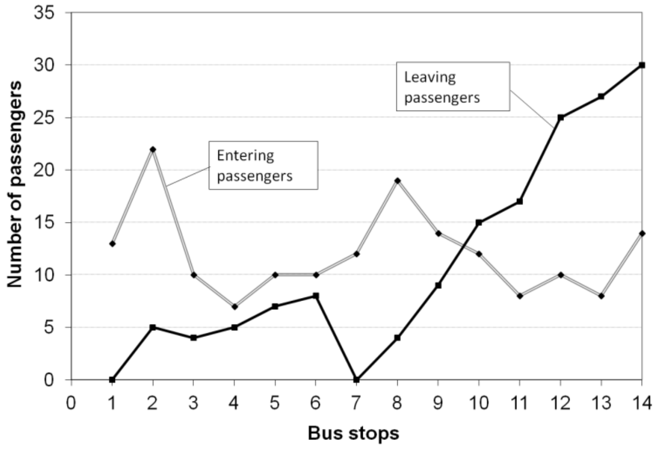

Figure 7: The discontinuous transfer task (modified from Ossimitz, 2001).

We used three different representational formats. In each condition, the solution diagram corresponded to the flow graph. In the baseline condition the flows were shown as line graphs (Figure 3, Figure 4, Figure 5). The cover story in the two continuous tasks was water and in the discontinuous task it was passengers. The solution sheet included an empty diagram on which to draw a solution graph (see Figure 8 for an example of a continuous task).

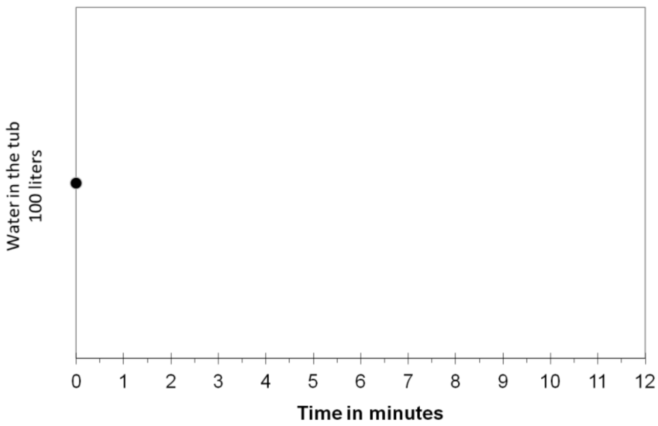

Figure 8: Solution diagram for a continuous task in the baseline condition.

The representation of the task was changed in two ways. In the tub condition, the flows were depicted as small water stocks in tubs that flowed into or out of the big water tub (Figure 9). The solution could be drawn on a sheet that displayed similar empty tubs, where the level of water for each minute could be marked (Figure 10).

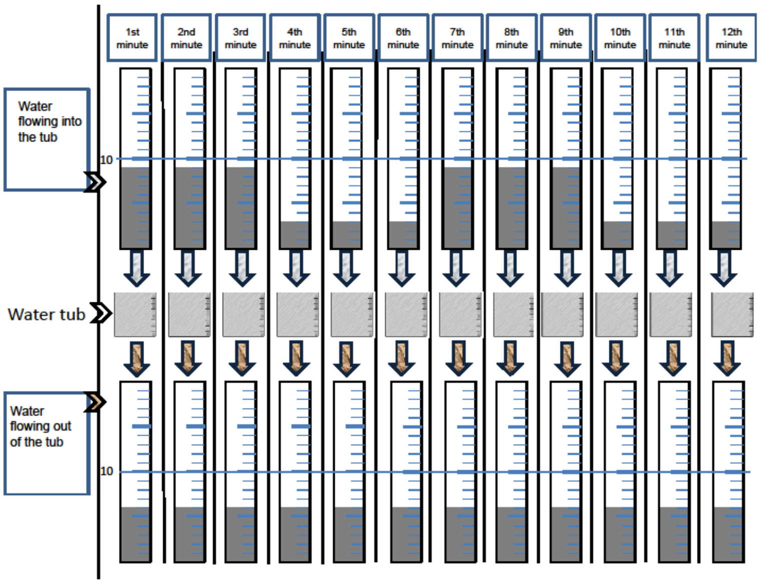

Figure 9: The tub representation depicting the flows of the square wave task.

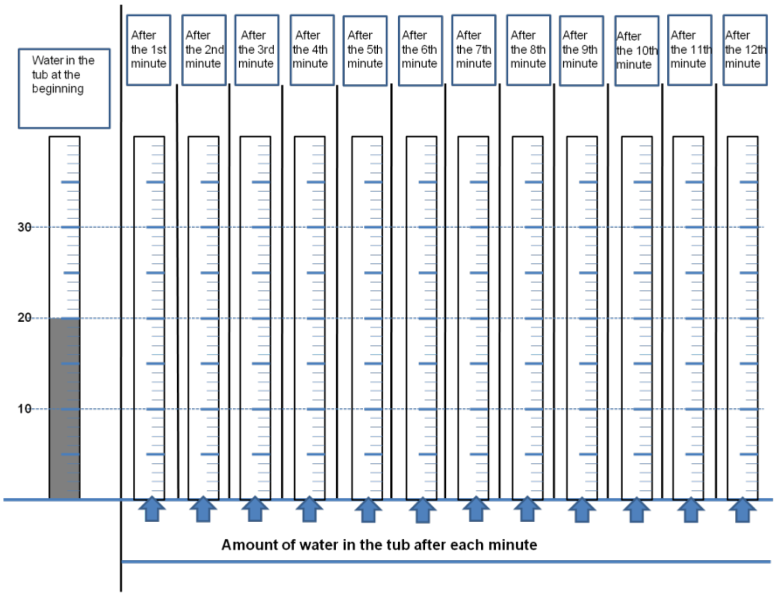

Figure 10: The tub representation solution sheet.

In the bus condition the flows were shown as small people who enter or leave a bus (Figure 11). To draw the solution, participants received a sheet with a grid of boxes. The boxes were shaded or ticked to represent the number of people on the bus (Figure 12).

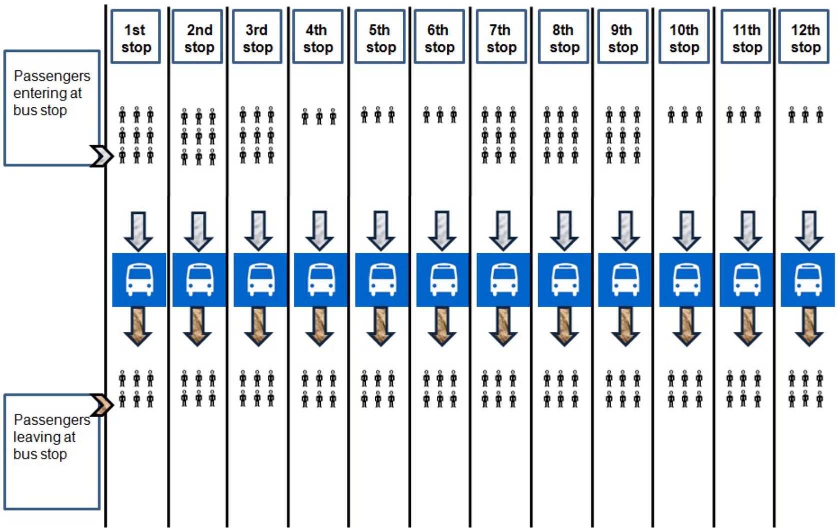

Figure 11: The bus representation depicting the flows of the square wave task.

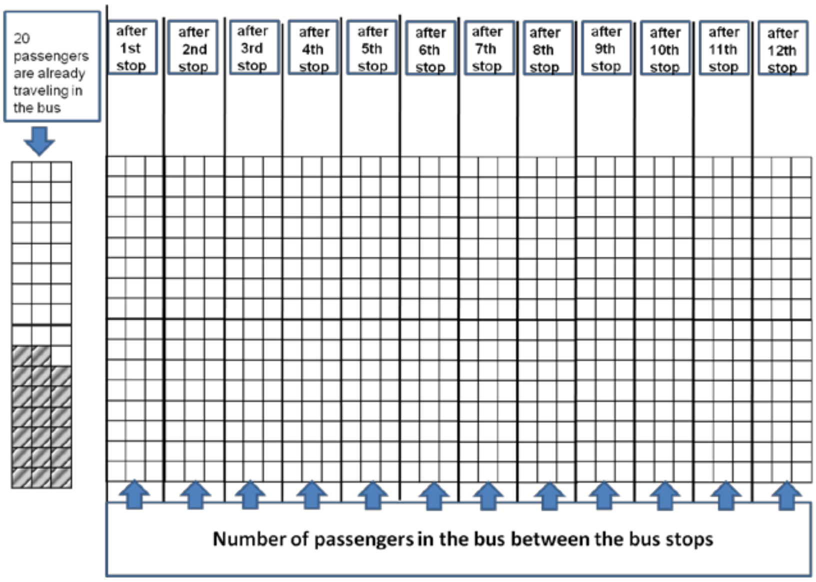

Figure 12: The bus representation solution sheet.

The two transfer tasks, one with discontinuous flows and people entering or leaving a bus and the other with continuous flows and a water tub, were presented as line graphs in all conditions. For the continuous task the solution had to be drawn on a solution diagram; for the discontinuous task, a few questions about the flows and one question concerning the stock had to be answered. The participants had to answer at what time the inflow was highest, at what time it was lowest, at what time the outflow was highest and lowest, and finally at what time the stock was highest. Moreover, each participant had to fill in a questionnaire about demographical data, such as age, sex, subject of study, and if they had participated in a previous stock–flow study.

Measures

The dependent variable was the correctness of the solution. Two points were given if the solution was completely correct, meaning if the changes in the stock over time were correctly judged. If a participant did not find the correct solution pattern but was correct concerning the stock's increase or decrease, 1 point was given. If not even the stock's increase or decrease was correctly judged, 0 points were given. Remember the triangle task and its solution (Figure 1 and Figure 2): Participants who drew this pattern got 2 points. Participants who drew a solution pattern as shown in Figure 13 got 1 point because the stock's decreases and increases were drawn at the correct times. But the slopes in the different time units were not drawn correctly, so they did not get 2 points. If a participant drew a solution pattern as shown in Figure 14, 0 points were given, because not even decreases and increases were judged correctly.

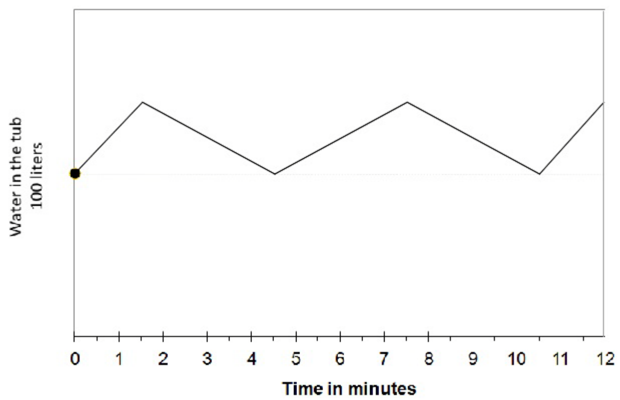

Figure 13: An incorrect solution pattern for the triangle task that was given 1 point.

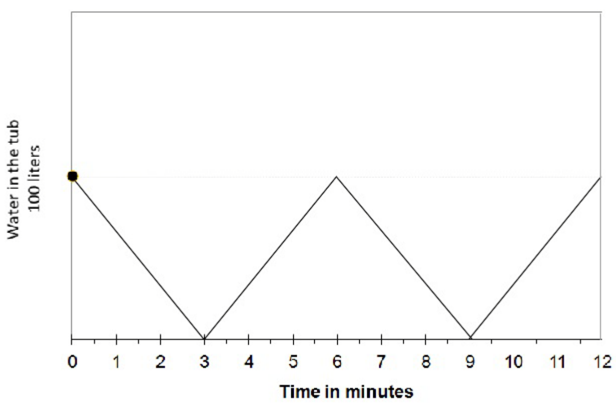

Figure 14: An incorrect solution pattern for the triangle task that was given 0 points.

The continuous transfer task was judged the same way. To solve the discontinuous transfer task, participants had to answer the question "at what time are the most passengers in the bus?" The answer was either correct (1 point) or incorrect (0 points).

For the analyses to determine if transfer occurred, we compared baseline percentages of the transfer tasks to the percentages of participants who solved a first-stage task (i.e., the square wave, triangle, or discontinuous task) and the transfer tasks correctly.

Procedure

Each participant had to solve three first-stage tasks. All were represented in the same way; that is, participants got all three tasks as the tub representation, the bus representation, or the baseline representation. After working on the three first-stage tasks, each participant had to solve the two transfer tasks. In all cases the transfer tasks were presented as line graphs. The order of the three first-stage tasks and the two transfer tasks was varied for each block of tasks separately. Finally, the additional questionnaire had to be filled out.

Statistics

To show explorative data analyses, we used bar graphs and boxplots (the line in the box represents the median, the box contains the inner 50% of the sample, and outliers are depicted as circles). Moreover we conducted regression analyses.

Results

We first tested if participants who had participated in previous stock–flow experiments performed better than the other participants. Fourteen participants had already taken part in such studies. Their performance did not differ from the other participants' performance (r = .079, p = .462). Therefore, we included all participants in our analyses. We also tested for position effects for the different first-stage tasks. No effects occurred, except for the triangle task, where we found a significant difference. Performance on the triangle task in the third position was better than in the first (r = .289, p = .025) and second (r = .310, p = .016) position, but no other difference in performance was significant. Therefore, we analyzed the task performance as if no position effect occurred.

How the modification of representation format influenced performance

Figure 15 shows the percentages of correct solutions for the different first-stage tasks in the three presentation conditions. In all cases n = 30 because each participant worked on all tasks.

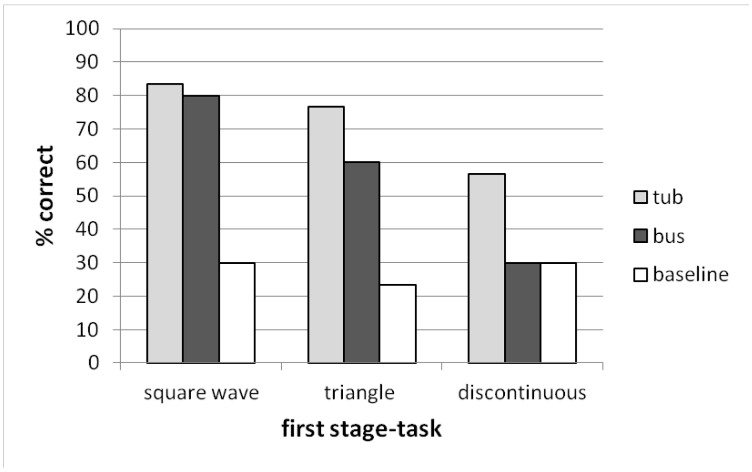

Figure 15: Percentages of correct solutions for the three first-stage tasks and the three conditions (tub, bus, and baseline; n = 30 for each bar).

Figure 16 shows boxplots of the dependent variable called sum of all tasks. It was the sum of points on the three first-stage tasks shown in the different formats. The sum ranged from 0 points (no task was solved correctly) to 6 points (all three tasks were solved correctly). It can be seen that the participants in the tub condition performed best, followed by the participants in the bus condition, who performed slightly worse, with identical median solution rates. The median in the baseline condition was much lower.

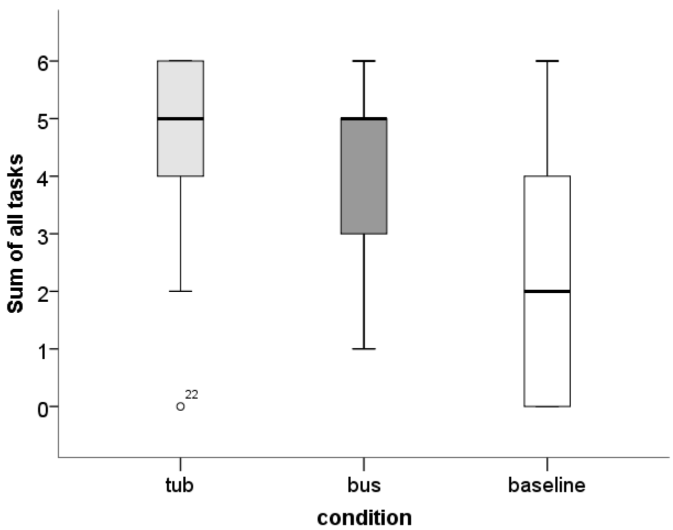

Figure 16: Boxplots for the overall performance on all three first-stage tasks.

We performed a standard multiple regression between correct solutions as the dependent variable and representation format, sex, and subject of study as independent variables. The dependent variable was the sum of points on the three first-stage tasks. The three conditions of different representation formats were coded as two dummy variables. Table 1 displays the unstandardized regression coefficients (*b*) and the standardized regression coefficients (β). The partial correlations are reported as measures of effect size. The overall explained variance was significantly different from zero, $F(4,89) = 13.55$, $p < .001$, with $R^2 = 0.39$; that is, more than a third of the variability in the stock–flow task performance was predicted by representation format, sex, and subject of study.

Table 1: Multiple regression of representation condition and personal variables on the stock–flow task performance (dummy coded variables: male = 0, natural sciences = 0)

Variable	*b*	β	Partial correlation	*p* value
(Constant)	3.028			<.001
Tub representation	2.527	.605	.557	<.001
Bus representation	2.166	.519	.498	<.001
Sex	-0.595	-.148	-.178	.099
Subject	-0.787	-.189	-.225	.036

The modified representation formats led to better performance than the format in the baseline condition. Both coefficients differed significantly from zero. The modification of the representation format was helpful. Participants in the tub condition performed best, followed by the participants in the bus condition, both showing large effect sizes.

The influence of sex and subject of study

Males performed better than females, and students of subjects in or near the natural sciences outperformed students of other subjects. All the coefficients differed significantly from zero. Both partial correlations were smaller than those for the representation format, and the effect size for sex was smaller than that for subject. We also tested if an interaction of sex and subject occurred. The regression analyses including the variables for representation format, sex, subject, and the interaction of sex and subject indicated that no interaction existed (partial correlation for the interaction: .075, $p = .492$).

Moreover, we analyzed whether there was an influence of sex and subject depending on the representation formats. That is, we analyzed the influence of both variables in the baseline condition, and separately for their influence in the two modified conditions. Participants' sex had the strongest influence on stock–flow task performance in the baseline condition, with male participants performing better (regression analyses only for the baseline condition revealed partial correlation: .597, $p < .001$). When we tested the two modified representation conditions without baseline almost no influence of sex was visible (partial correlation: .070, $p = .596$). The new representation formats seemed to lower, even to eliminate an effect of sex. For subject of study, a regression analyses only for the baseline condition showed a partial correlation of .356 ($p = .058$). For the two modified representation conditions without baseline condition, a regression analyses revealed a partial correlation of .211 ($p = .109$). The new representations seemed to lower an effect of subject, whereas in the baseline condition, it was a very strong effect.

Analyses of transfer

Boxplots show that the performance for the transfer stock–flow tasks was very similar in the different conditions (Figure 17). Actually, the participants in the baseline condition performed better than those in the bus condition.

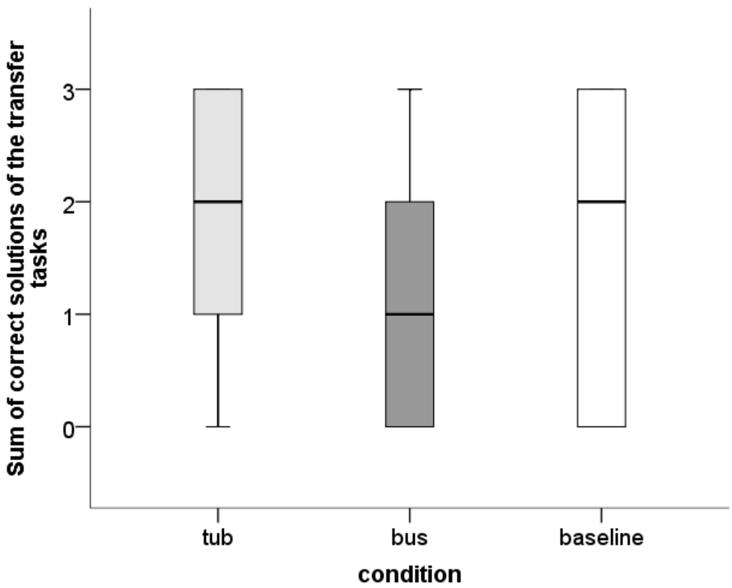

Figure 17: Overall performance for the transfer tasks.

We used the baseline rates of the transfer tasks' solutions as limits to judge if transfer occurred. The baseline rates were 47% for the continuous and 50% for the discontinuous transfer task. Therefore, 50% was our benchmark to test if transfer occurred. We analyzed for each first-stage task in each of the conditions how many participants had solved the respective task correctly. We then calculated the percentage of those participants who also solved the two transfer tasks. This step was conducted separately for the continuous (Table 2) and discontinuous (Table 3) transfer tasks.

Table 2: Percentage of correctly solved continuous transfer tasks (base rate: first-stage-task correct solvers correspond to 100%)

Task	Condition		
	Baseline	Tub	Bus
Square wave	78%	44%	44%
Triangle	100%	43%	41%
Discontinuous	44%	28%	22%

Table 3: Percentage of correctly solved discontinuous transfer tasks (base rate: first-stage-task correct solvers correspond to 100%)

Task	Condition		
	Baseline	Tub	Bus
Square wave	89%	64%	67%
Triangle	100%	61%	59%
Discontinuous	67%	33%	44%

The best performance occurred in the baseline condition independently of the transfer task. In the modified representation conditions, performance differed for the continuous and the discontinuous transfer task. When looking at the solutions of the continuous transfer task, all rates were under the benchmark of 50%. When looking at the discontinuous transfer task, some of the rates were larger than 50%. Only the rates of the discontinuous first stage task in the tub and bus conditions were lower than 50%. So, the results might be interpreted as meaning that a small amount of transfer occurred for the discontinuous transfer task, but only for the correct solvers of the square wave and the triangle task. In the baseline condition, transfer occurred systematically, but only in the continuous first stage tasks. For correct solvers of the discontinuous first-stage task, no transfer occurred.

Because not only the percentages of correct solvers but also the number of participants included in the analyses is important, we include figures that display both details. Figure 18 shows how many of the participants who solved each first-stage task correctly also solved the continuous transfer task correctly. The complete boxes, that is, including the unshaded and the shaded parts, show the number of participants who completely solved the first-stage task. The filled parts of the boxes display how many of them also answered the continuous transfer task correctly. Therefore, the empty parts of the boxes show the percentage of participants who had solved the first task correctly but did not solve the transfer task. For example 25 participants in the tub condition solved the square wave task correctly. Forty-four percent (11 participants) also solved the continuous transfer task and 56% (14 participants) did not solve the continuous transfer task correctly. The total number of participants that solved the transfer task correctly was larger in the modified conditions. It was highest in the tub condition. But the percentage of first-stage correct solvers who also correctly solved the transfer task was highest in the baseline conditions for the square wave and the triangle task (in the latter even 100%).

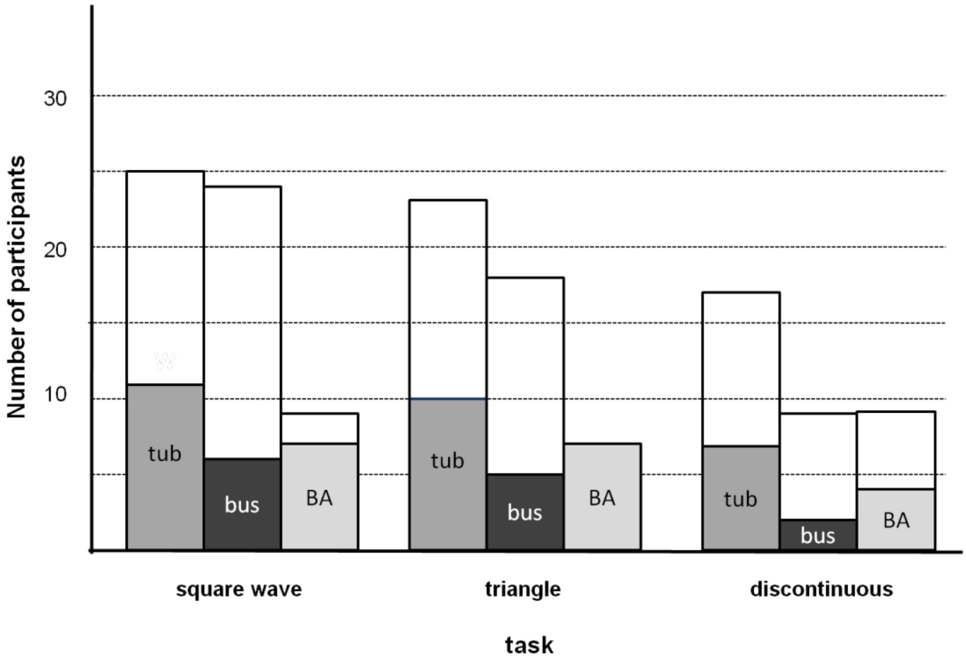

Figure 18: Number of participants in each condition (tub, bus, baseline) who correctly solved the continuous transfer task (shaded areas) out of the number who correctly solved (i.e., earned 2 points) the first-stage task (shaded plus unshaded areas), by task. BA: baseline.

The analogous picture for the discontinuous task shows a slightly different pattern (Figure 19). Still, the percentages of participants who solved both tasks correctly were highest in the baseline condition, but for this task, also the percentages in the tub condition were all over 50%. Again, the total number of correct solvers was highest in the tub condition.

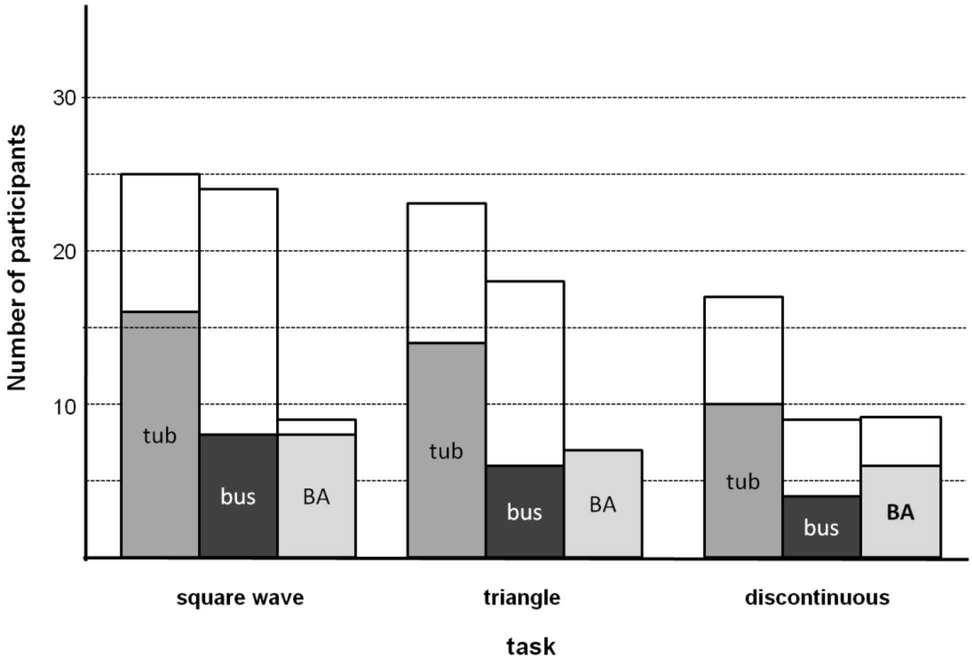

Figure 19: Number of participants in each condition (tub, bus, baseline) who correctly solved the discontinuous transfer task (shaded areas) out of the number who correctly solved (i.e., earned 2 points) the first-stage task (shaded plus unshaded areas), by task. BA: baseline.

Regression analyses revealed that sum of all first-stage tasks used as predictor was not significant when only tub and bus conditions were included in the analyses (partial correlation continuous transfer task: -0.163, p = .227; discontinuous transfer task: .099, p =.465). To summarize, modified representation formats led to better stock–flow task performance. The effect sizes were high compared to effect sizes of previous studies. Unfortunately transfer from the picture format to a line graph format barely occurred.

Discussion

We analyzed the influence of a new kind of representation format on performance in three different stock–flow tasks. We depicted the tasks as teaching pictures where flows were represented as “small stocks” and the time flow was broken into small time units. This modification helped the participants solve the stock–flow tasks. More correct solutions were found in the modified representation conditions. Especially the tub representation led to better performance than the baseline representation.

Our results might be explained by results in problem-solving research. In a review of problem-solving strategies, Gick (1986) described the problem-solving process as follows. The solver first constructs a representation for the problem and then searches for a solution that is then implemented. If it is a successful solution the process is finished. If it is not successful, the solver goes back one or two steps in the process. In the first step (constructing a representation) the problem solver can activate prior knowledge that can lead to schema activation. If a schema is activated, the problem solver does not search for a solution but uses his or her knowledge of solving precedent problems to implement a solution for the present problem. So, prior knowledge can reduce the need to search for a solution by activating a schema. A line graph representation of a stock–flow problem may activate a wrong schema, for example, the schema "summing up the values of the lines to get the final result," and the process of accumulation is not conducted. Modified representations (e.g., teaching pictures) instead might directly activate the correct schema of accumulation. It is also possible that they do not activate a schema at all; the problem solver may search for a solution, and the probability for finding the correct solutions might to be larger in the modified representation format.

We also analyzed some moderator variables. As often found in stock–flow studies, males outperformed females (Booth Sweeney & Sterman, 2000; Brockhaus & Sedlmeier, unpublished document; Ossimitz, 2002; Schwarz & Sedlmeier, unpublished document; Sedlmeier et al., in press) and students of subjects in or near the natural sciences performed better than students of other subjects (e.g., social sciences; see, e.g., Brockhaus & Sedlmeier, unpublished document). Similarly, an effect of grades in mathematics on stock–flow task performance was found, with better grades linked to better performance (Schwarz & Sedlmeier, unpublished document). In the present study, the differences between the sexes disappeared in the teaching picture representations. Moreover the differences in dependence on the study subjects decreased strongly in the teaching picture formats. This seems to be consistent with Neurath's (1936) suggestion that people who dislike graphs with lines and numbers (presumably the female participants and students of social sciences might fall into this category to a larger extent) cope better when they are offered teaching pictures. It supports the assumption that no expertise is necessary to solve stock–flow tasks if they are displayed using teaching pictures.

We also tested if transfer occurred. We looked at the percentage of participants who solved the first-stage tasks in the modified representation format correctly who also solved the transfer tasks in the classic line graph format correctly. Only in some cases did transfer occur. The numbers of participants who correctly solved both the first-stage tasks and the transfer tasks were much lower in the modified representation conditions than in the baseline conditions, where all or almost all participants correctly solved the first-stage and transfer tasks.

Transfer is in many domains hard to achieve (Reed, 1999). Reed gave examples of different relations between problems where transfer may occur. One example was "isomorphic problems" where two problems have the same isomorphic solution structure but this may not be obvious at first glance. It is very challenging to make obvious that two problems have the same solution structure so that the known schema of the correct solution procedure is activated, as illustrated in Gick's (1986) description of the problem-solving process (as described above). In our study, it seemed that the schema that produced correct solutions in the teaching picture conditions was not activated again in the additional transfer tasks and therefore no systematic transfer occurred. Only in the baseline condition did most of the participants who had solved the first-stage tasks also correctly solve the transfer tasks. This result is not surprising because the representation format is the same and it is very probable that the same solution schema was activated.

The difference in performance between the first-stage baseline tasks and the transfer tasks (about 30% vs. about 50%) may have stemmed from the transfer tasks being easier to solve, because in previous studies, no position effects occurred (Brockhaus & Sedlmeier, unpublished document). In the present experiment, significant differences were found for the triangle task depending on position, but because only two of nine possible differences were significant, we assumed the significance was coincidental.

The difference between the tub and bus conditions is a surprising result. Participants might have had more difficulty solving the tasks in the bus condition because of its unusual arrangement of the entities. Remember that in the bus condition (Figure 11), three people were depicted in one row. But in Western society it is more usual and easier to calculate in sets of five. In the tub condition, this was implemented through larger marks on the tubs for

5, 10, 15, and so on (Figure 9). This might have led to more calculation errors occurring in the bus condition and therefore better performance in the tub condition.

In conclusion, the teaching picture format seems to simplify finding the correct solutions. Especially people who are not so fond of solving formal problems seem to profit from the modified format. The same can be said for people who have less formal training in mathematics and similar domains: Differences dependent on participants' sex and subject of study disappeared. Unfortunately, no systematic transfer occurred. So, further studies should explore how to obtain stable transfer. Another important issue is to optimize the teaching pictures. In our pictures it was necessary to calculate the stock at each time unit to find the correct solution. Performance was worse on the bus representation tasks than on the tub representation tasks. In our opinion the reason lies in the suboptimal way of subsuming entities (three instead of five in a row). To determine if this is correct, future research might analyze what the best presentation is for calculating such tasks. Besides the "three-in-a-row" problem, other issues might be relevant.

Moreover it might be of interest to find a representation that does not encourage calculating the solution, but where a "basic understanding" is reached. For example, Cronin et al. (2009) created tasks where questions concerning the flows and the stock were asked (as in our discontinuous transfer task). These tasks seemed rather focused on a basic stock–flow understanding. It might be interesting to test such a representation that focuses not on calculating but on basic understanding. An example can be found in Neurath's (1936) work on ISOTYPE. Neurath showed a way to depict the rates of births and deaths where one pictogram represented an amount of an entity. For the bus condition, this would mean not one single person but a specified number of people would be represented by one pictogram of a person. For the tub condition, one tub pictogram would represent a specified number of liters. Participants should be able to work out the correct solutions to the questions with and without calculating, but the pictures should encourage participants not to calculate. If such a representation was created, a good option for presenting stock and flow data in everyday life—for instance, in the media or school books—would be found. Our advice at the moment is this: To present stock–flow material, do not use line graphs but show teaching pictures.

References

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

Brockhaus, F., & Sedlmeier, P. (2013). *Is intuition the key to understanding stock and flow systems? The influence of modifying the systems' representation format on the stock flow performance*. Unpublished manuscript, Chemnitz University of Technology. Chemnitz.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

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

Domagk, S. (2008). Bilder. In Niegemann, H., Domagk, S., Hessel, S., Hein, A., Hupfer, M. & Zobel, A. *Kompendium multimediales Lernen.* (pp. 207- 238). Berlin: Springer.

Dwyer, F. M. (1975). Effect of students entering behavior on visualized instruction. *Journal of experimental Education, 43*, 78-83.

Gick, M. L. (1986). Problem-Solving Strategies. *Educational Psychologist, 21*, 99-120.

Gigerenzer, G., Hertwig, R., Hoffrage, U. & Sedlmeier, P. (2008). Cognitive illusions reconsidered. In C. R. Plott, & V. L. Smith (Eds.), *Handbook of experimental economics results. Volume 1*. Amsterdam: North-Holland. (pp. 1018-1034).

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, *73*, 538–540.

Neurath (1936). *International Picture Language*. London: Kegan Paul.

Norman, D. A. (1988). *The psychology of everyday things*. New York, N.Y.: Basic Books.

Ossimitz (2001). *Unterscheidung von Bestands- und Flussgrößen*. Universität Klagenfurt. Retrieved from http://wwwu.uni-klu.ac.at/gossimit/pap/pap.htm

Ossimitz (2002). *Stock-Flow-Thinking and reading stock-flow-related graphs: An Empirical Investigation in Dynamic Thinking Abilities*. Paper presented at the 20th International Conference of the System Dynamics Society, Palermo, Italy.

Reed, S. (1999). *Word Problems: research and curriculum reform*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Schwarz, M., & Sedlmeier, P. (2013). *It is not that easy: animated representation formats and instructions in stock-flow tasks*. Unpublished manuscript: Chemnitz University of Technology.

Sedlmeier, P. (1998). The distribution matters: Two types of sample-size tasks. *Journal of Behavioral Decision Making, 11*, 281-301.

Sedlmeier, P. (2000). How to improve statistical thinking: Choose the task representation wisely and learn by doing. Instructional Science, 28, 227-262.

Sedlmeier, P., Brockhaus, F., & Schwarz, M. (in press). Visual integration with stock-flow models: How far can intuition carry us? In P. Bender, R. Hochmuth, P. Fischer, D. Frischemeier,& T. Wassong Eds.) *Using tools for learning mathematics and statistics*. Berlin: Springer.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23*, 63-71.

Zelazny (1985). *Say it with charts*. Illinois: Dow Jones-Irwin.