Simulation & Gaming
2015, Vol. 46(3-4) 255–269

DOI: 10.1177/1046878114565058
sag.sagepub.com

Improving Stock-Flow Reasoning With Verbal Formats

Helen Fischer[1], Christina Degen[2], and Joachim Funke[1]

Abstract

Background. **Stock-flow (SF) problems** are **ubiquitous** in nature, ranging from the accumulation of water in a tub to the accumulation of CO_2 in the atmosphere. However, research on **SF failure** repeatedly demonstrates that people have severe difficulties understanding even the most basic SF problems.

Purpose. This study tested the **hypothesis** that people's **understanding** of SF problems depends on the **presentation format** used. Specifically, we expect SF failure to decrease when avoiding previously used scientific formats comprising coordinate systems and graphs, and SF problems are presented in verbal formats.

Method. Participants (N = 107) solved a range of different SF problems with experimentally **varied presentation formats** (verbal vs. graphic). We assessed fundamental **understanding** of **graphs and graphical** versus **verbal production** of stocks and in- and outflows.

Results. Solution rates show that (a) **SF failure** is at least partially caused by specifics of the **presentation format** used previously; (b) fundamental **misunderstandings** in the construction of graphs can **explain previous findings**; and (c) the **majority of participants arrived at the correct solution** when SF problems were presented verbally.

Conclusion. The present study indicates that people are able to **solve SF problems** when they are presented in **accessible formats**. This result bears implications

[1]Universitaet Heidelberg, Germany
[2]Uniklinik Heidelberg, Germany

Corresponding Author:
Helen Fischer, Psychologisches Institut, Universitaet Heidelberg, Hauptstr. 47-51, 69117 Heidelberg, Germany.
Email: helen.fischer@psychologie.uni-heidelberg.de

for simulation-based **learning** and **assessment**, and for the **communication** of SF problems.

Keywords
accessible formats, dynamic problems, graphs, inflow, misunderstandings, outflow, presentation format, scientific formats, stock and flow failure, stock-flow problems, understanding, verbal

It is a well-established finding that humans have severe difficulties understanding stock-flow (SF) dynamics, a phenomenon termed *SF failure* (see Sterman, 2011, for a recent review). Any system comprising a stock that accumulates over time and is dependent on given in- and outflows constitutes an SF system. The structure of SF systems is often explained by a bathtub analogy: The water level (stock) in a bathtub increases if the inflow of water through the faucet exceeds the outflow through the drain; the water level drops if the outflow exceeds the inflow. Consequently, our definition is that people have an understanding of the fundamental SF structure if they understand "that the stock rises when the inflow exceeds the outflow, and vice versa" (Cronin, Gonzalez, & Sterman, 2009, p. 9). Understanding SF systems is critical for many areas of life, ranging from everyday phenomena such as the accumulation of money in a bank account or the regulation of body weight, to more abstract scenarios such as the accumulation of CO_2 in the atmosphere. Given the ubiquity of SF dynamics, it seems perplexing and critical that people have great problems with regulating complex SF simulations containing many interrelated variables (Diehl & Sterman, 1995), and even with understanding extremely simplified SF systems (Booth Sweeney & Sterman, 2000; Sterman & Booth Sweeney, 2002, 2007).

In this article, we argue, however, that the ability to solve SF problems is influenced by the way the problems are presented (*presentation format* henceforth), and that previous research used potentially error-inducing formats. Similarly, research has demonstrated that displaying isomorphic problems in different presentation formats can have a dramatic impact on problem-solving performance, such as on the Wason selection task (Cheng & Holyoak, 1985), the Tower of Hanoi (Kotovsky, Hayes, & Simon, 1985), deductive reasoning (O'Brien, Noveck, Davidson, & Fisch, 1990), graphical tasks (Hegarty, Canham, & Fabrikant, 2010; Novick & Catley, 2007), and mathematical problems (Bassok, 2001; Landy & Goldstone, 2007). The aim of the present article is twofold:

1. to separate difficulties caused by the presentation format of the SF task from difficulties caused by the SF system itself and
2. to develop a presentation format that enables more participants to derive correct conclusions in SF systems

In the context of simulations and games, presentation formats that deliver valid assessment of people's understanding of SF systems are especially important, because the simulation or game should not only measure people's understanding, but also *help* them understand. When giving trigger-based feedback, for example, a simulation might assess how well the learner is doing and it might use this assessment for scaffolding (e.g., giving suggestions on possible actions), for delivering background information (e.g., on the state of the system), or for adaptation (e.g., of the difficulty of the simulation). For all of these purposes, a valid assessment of the learner's understanding of the SF system is essential: If the learner's understanding of the SF system is under- or overestimated, then scaffolding cannot be in tune with the learner, background information might be too difficult or unnecessary, and the simulation might demand too much or too little.

Thus far, assessment of people's understanding of SF systems has been rather pessimistic: In the dynamic stock and flows task, for example, participants needed to keep the accumulation of a simulated stock such as water or CO_2 within a predefined range by manipulating user in- and outflow rates under the condition of varying environmental inflow and constant environmental outflow (Dutt & Gonzalez, 2007; Gonzalez & Dutt, 2011). It was found that, to achieve the desired stock level, participants used a pattern matching heuristic by simply matching the shape of the flow function (e.g., increasing) to the shape of the environmental inflow function, regardless of the constant environmental outflow. Thus, participants disregarded the fundamental SF structure of the problem.

Stock-flow failure was not only found in simulated environments, but even in basic SF systems that were reduced to the essentials: one inflow, one outflow, and one stock (Sterman & Booth Sweeney, 2002, 2007). In these paper-based tasks, participants were typically first presented with an introduction to the scenario, such as atmospheric CO_2 concentration. They were then presented with a graph depicting atmospheric CO_2 concentration stabilizing from the year 2100 onward and with a graph depicting previous CO_2 emissions and absorptions. Participants were asked to sketch emission and absorption trajectories, so that a stabilizing CO_2 concentration could be achieved. In similar fashion to results from simulations (Dutt & Gonzalez, 2007; Gonzalez & Dutt, 2011), participants typically made use of a pattern matching heuristic, sketching in- and outflows that followed the trajectory of the stock. As a result, drawn emissions typically exceeded absorptions leading to an actual *increase* of atmospheric CO_2 (Sterman & Booth Sweeney, 2002, 2007). SF failure was also demonstrated for multiple-choice answer formats, different outcome scenarios (e.g., atmospheric CO_2 concentration decreasing), and different semantic embeddings (Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Sterman & Booth Sweeney, 2002, 2007). Thus, SF failure has so far been found both in simulations and in a wide range of simplified paper-based tasks.

We argue, however, that it is necessary to distinguish different sources of difficulty that might arise when dealing with SF problems. Specifically, in simulations, participants might lack skills to regulate the system (Mislevy, 2011). If that is the case, participants might know what to do, but they simply cannot do it well or fast enough (i.e., In a department store, people enter and leave over a 30-minute period. In the first minute, 9 people enter and 8 leave. In the second minute, 10 people enter and 5 leave. In the third minute, 9 people enter and 8 leave. In the fourth minute, 14 people enter and 12 leave. In the fifth minute, 9 people enter and 8 leave. In the sixth minute, 9 people enter and 8 leave. In the seventh minute, 8 people enter and 8 leave. In the eighth minute, 7 people enter and 9 leave. In the ninth minute, 4 people enter and 13 leave. In the tenth minute, 7 people enter and 11 leave. In the eleventh minute, 10 people enter and 15 leave. In the twelfth minute, 8 people enter and 12 leave.

1. During which minute did most people enter the store?
2. During which minute did most people leave the store?
3. During which minute were the most people in the store?
4. During which minute were the fewest people in the store?

Figure 1. Textual display of the original presentation format of SF problems.
Source. Cronin, Gonzalez, and Sterman (2009).
Note. SF = stock-flow.

they might possess declarative, but lack procedural knowledge). Moreover, previous paper-based tasks, despite variation, all contained one possibly critical aspect: an overall scientific notation including coordinate systems, graphs, and percentage values. It has been shown in several studies that comprehension of coordinate systems and graphs is error-inducing (Carpenter & Shah, 1998; Gattis & Holyoak, 1996; Shah & Carpenter, 1995) and that participants have difficulties dealing with percentage values (Gigerenzer & Hoffrage, 1995; Hoffrage & Gigerenzer, 1998; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Consequently, people might possess a basic, declarative, understanding of SF systems, but this might have been concealed in previous research in which additional and potentially error-inducing skills and knowledge were needed.

Cronin et al. (2009) specifically investigated whether SF failure is a mere artifact of using coordinate systems by presenting participants with alternative formats (line graphs, bar charts, texts, and tables; see Figure 1 for the textual display). Participants needed to solve the so-called department store problem, describing the number of people entering and leaving a department store over a period of time. To control for comprehension of the presentation format, participants were asked at what time most people entered or left the department store. Because the majority of participants were able to answer these control questions correctly, but still showed SF failure, the Cronin et al. concluded that SF failure is not an artifact of the presentation format, but rather a fundamental error in human reasoning.

However, this conclusion might be premature for three reasons. *First*, the control questions could be answered correctly by using simple salience heuristics picking the highest or lowest number given. Thus, only a superficial understanding of coordinate systems was necessary. An arguably deeper understanding, however, is necessary to be able to answer the SF problems. *Second*, the control questions tested interpretation of graphs and not the construction thereof. However, construction of graphs was a prerequisite for solving the SF problems correctly. *Third*, in all data displays—even the textual—specific numerical information was salient. We argue that this salience of quantitative information encourages participants to focus on and work with the given numbers, rather than making an effort to detect the underlying SF structure. It is conceivable that a qualitative presentation format might encourage and enable participants to detect the underlying SF structure.

The experiment presented in this article investigated whether different SF problems measure *construct-relevant* aspects of the problem (understanding of *SF structure*) versus *construct-irrelevant* aspects of the problem (understanding of the *presentation format*) using two different tasks: Interpretation and Production tasks and Verbal tasks.

1. *Interpretation and Production tasks (I/P tasks)*: I/P tasks examined whether participants are equally able to interpret and produce graphs, and whether they are equally able to submit their answers verbally and graphically. These distinctions were introduced to investigate whether participants' potential understanding of SF dynamics was concealed in previous research: If participants are able to answer SF questions correctly when submitting their answers verbally, but then make errors constructing the corresponding line graph, the original presentation format could not be seen as a valid assessment of participants' understanding of SF systems.
2. *Verbal tasks*: Verbal tasks did not rely on coordinate systems or graphs for either problem description or answer format by using multiple-choice answers. Verbal tasks also contained little or no numerical information. Hence, verbal tasks tested whether SF failure could be reduced or even eliminated when an understanding of coordinate systems is not required, no graphical reference is given, and when participants are encouraged to detect the qualitative gist of the problem structure.

We hypothesize the following:

Hypothesis 1 (H1): Even those participants who correctly solve a given SF problem verbally may not be able to construct the corresponding line graph into a coordinate system. Thus, solution rates for one and the same problem will be lower when a graphical answer is required (Question 4 in the I/P tasks) than when a verbal answer is required (Question 3 in the I/P tasks).

Hypothesis 2 (H2): SF failure will be significantly reduced in a verbal and multiple-choice presentation format that comprises no coordinate systems or graphs and little or no quantitative information (verbal tasks).

Method

Participants

A total of *N* = 107 participants (65% females) between 23 and 75 years of age took part in the experiment. Mean age was 48.4 years (*SD* = 16.9). All participants gave written informed consent and were debriefed on the purpose and results of the study. The sample consisted of students from the University of Heidelberg and people from the general population. Participants received course credit or 5€ for participation.

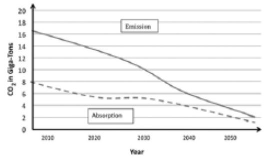
Materials

1. *I/P tasks*: I/P tasks were administered in two scenarios (atmospheric CO_2 concentration, number of children on a playground). Each scenario comprised four questions that we illustrate using the CO_2 scenario (see Figure 2). Participants first received a short introduction to the problem describing the relationship between CO_2 emissions, absorptions, and atmospheric CO_2 concentration. Participants were then presented with a coordinate system depicting in- and outflows and four subtasks exploring fundamental understanding of the graphs (Question 1), verbal production of the resulting stock (Question 2), verbal production of necessary inflows and outflows given a decreasing stock (Question 3), and the graphical production of the answer to Question 3 into a coordinate system (Question 4). Note that Question 3 (verbal production of in- and outflows) was an easy question to test whether participants who are able to produce a correct verbal answer necessarily produce a correct graphical answer. Questions in the playground scenario were identical, except that in Question 3, participants were asked to achieve a stabilizing stock (see Appendix A for the complete playground scenario).
2. *Verbal tasks*: Verbal tasks comprised a verbal description of the problem and a multiple-choice answer format, and were administered in three different scenarios (money in a piggy bank, water in a bathtub, atmospheric CO_2 concentration). Participants first received a short introduction to the problem. In the bathtub (piggy bank) problem, participants were then asked to name the correct strategy in order to achieve a stabilizing (rising) stock, that is, to give a qualitative estimation of flows. In the CO_2 problem, participants needed to determine how the stock reacts if emissions were reduced by 30%, that is, to give a qualitative estimation of the stock. For illustration, in the bathtub scenario, participants were given the following instructions and problem (see Appendix B for the piggy bank and CO_2 scenario):

You have a bathtub. Water runs into this bathtub through the tap. Meanwhile, water runs out of the bathtub through the drain because it does not seal properly. Imagine, ten minutes ago, you started letting water run into the bathtub and you are now satisfied with the water level. What do you need to do in order to keep the current water level constant?

a. Open the water tap a little further.
b. Leave the tap as it is.
c. Close the water tap a little.

CO_2 emissions are caused by the burning of fossil fuels and lead to an increase of atmospheric CO_2 concentration. CO_2 absorptions are caused by forests and oceans and decrease atmospheric CO_2 concentration. The Figure below depicts CO_2 emission and CO_2 absorption trajectories between 2010 and 2050.



1. How does CO_2 emission relate to CO_2 absorption between 2010 and 2050 in the Figure above?
- ☐ CO_2 emission is greater than CO_2 absorption.
- ☐ CO_2 emission is smaller than CO_2 absorption.
- ☐ CO_2 emission and CO_2 absorption are equivalent.

2. If CO_2 emission and CO_2 absorption relate to each other as depicted in the Figure above: What happens to atmospheric CO_2 concentration?
- ☐ CO_2 concentration will rise.
- ☐ CO_2 concentration will fall.
- ☐ CO_2 concentration will remain constant.

3. Assuming that the atmospheric CO_2 concentration will fall: What would the corresponding CO_2 emission and absorption trajectories have to look like?
- ☐ CO_2 emission would have to be greater than CO_2 absorption.
- ☐ CO_2 emission would have to be smaller than CO_2 absorption.
- ☐ CO_2 emission would have to be equal to CO_2 absorption.

4. Please sketch your answer to question 3. into the figure below. Draw one line for CO_2 emission and another line for CO_2 absorption trajectories and label them.

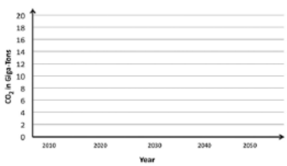


Figure 2. Example of the I/P task: Atmospheric CO_2 scenario (translated).
Note. Participants are presented with emissions and absorptions trajectories. The following subtasks test participants' understanding of graphs (Question 1), verbal production of the resulting stock (Question 2), verbal production of necessary inflows and outflows given a decreasing stock (Question 3), and graphical production of the answer to Question 3 into a coordinate system (Question 4). I/P = interpretation and production.

Thus, the bathtub scenario was a verbal translation of the original presentation format (Sterman & Booth Sweeney, 2002, 2007) comprising one inflow and one outflow and a to-be-stabilized stock.

Procedure

Each participant completed both I/P tasks (playground, CO_2) and one randomly assigned verbal task (bathtub, piggy bank, CO_2). Presentation order was randomized.

Results

In the I/P tasks, the majority of our sample (M = 97%) was able to correctly read and interpret the graphs (Question 1, see Figure 2). Also verbal production tasks about both flows (Question 2, M = 83%) and stocks (Question 3, M = 89%) were answered correctly by the majority of participants, producing no significant difference between both tasks, $\chi^2(1, N = 107) = 1.39$, $p = .24$. However, in line with our expectations, translating verbal answers of flows (Question 3) into a graphical presentation (Question 4) was only accomplished by 57% of the sample. A McNemar test yielded a significant difference between solution rates of the verbal and the graphical production tasks (Questions 3 and 4), $\chi^2(1, N = 107) = 8.65$, $p = .003$, indicating that for most participants, answers were easier to provide in a verbal than in a graphical format. Unexpectedly, while no significant differences were found in the CO_2 scenario compared with the playground scenario for Questions 1 to 3 ($p > .05$), a McNemar test yielded a significant difference between solution rates of the two scenarios in the graphical production task (Question 4), $\chi^2(1, N = 107) = 16.80$, $p < .001$: While 79.3% of the participants were able to sketch their answer in the CO_2 scenario, only 35.4% were able to sketch their answers in the playground scenario. That is, participants were more correct drawing the relation "outflow must be smaller than the inflow" than drawing the relation "outflow must equal inflow." We found a typical mistake in sketching the latter: Instead of drawing two identical lines, 22% of participants drew two parallel lines, resulting in different y values for in- and outflows. (Note that lines were only rated as parallel, and not as identical if they were at least 0.2 inch apart.) In summary, we found that when participants needed to submit their answers graphically, solution rates to the SF questions were dramatically lower than when participants submitted their answers verbally.

In line with our hypothesis, the majority of our sample was able to answer SF questions in the verbal tasks, yielding an average correct solution of $M = 86\%$. Specifically, solution rates ranged from 98% and 90% (bathtub and piggy bank task, respectively) to 70% (CO_2 task). Thus, SF failure could be reduced when a presentation format without coordinate systems and graphs and without a focus on quantitative information was used.

Discussion

The present experiment tested whether SF failure can at least partly be explained by the presentation format. Results showed not only that participants have difficulties dealing with the graphical format used in previous research, but also that that SF reasoning improves dramatically in a verbal format.

In line with our hypothesis, I/P tasks revealed that the requirement of the standard task to produce graphs may have decreased solution rates. We found that solutions to one and the same task were reduced by up to 50% when a graphical compared with a verbal answer was required. Thus, submitting answers graphically results in a dramatic underestimation of participants' true SF reasoning abilities.

One task with a stabilizing stock was particularly revealing: In the verbal condition, most participants arrived at the correct solution (inflow equaling outflow); when asked to draw this exact answer into a coordinate system, however, nearly one quarter of our participants sketched two parallel lines. This misconception in the construction of graphs may partially explain the typical mistake in the standard task with stabilizing stock (e.g., Sterman & Booth Sweeney, 2007): Our results suggest that at least some participants may well have the correct verbal representation of the inflow needing to equal the outflow, but then submit a wrong answer by sketching the inflow paralleling the outflow. Thus, the original task presentation format using coordinate systems and graphs seems to underestimate participants' ability to grasp SF problems because an error-inducing layer is added between participants' mental representations and their submitted answers.

Although different cognitive mechanisms might be required for estimating flows from a given stock than vice versa, solution rates between both kinds of I/P tasks did not differ significantly. This result implies at least that the majority of people are able to accomplish both if the tasks are presented verbally. Possibly, this might even hint toward the underlying cognitive mechanisms being rather similar.

When both a focus on quantitative information and the use of coordinate systems and graphs were avoided in the verbal tasks, a majority of participants arrived at the correct solution to different SF problems. This result suggests that participants are able to understand the qualitative gist of SF problems when they are presented verbally.

Even the use of the pattern matching heuristic was significantly reduced in the verbal CO_2 task given that 70% of participants correctly answered that the stock increases, even if CO_2 emissions are reduced. In other contexts, it was repeatedly shown that participants are able to overcome simple heuristics with insight and prefer to make use of the causal structure underlying the problem (Brehmer, 1976; Garcia-Retamero, Wallin, & Dieckmann, 2007; Gonzalez, 2004). Similarly, it was assumed before that participants might *either* use the pattern matching heuristic *or* make use of the problem's causal structure (Cronin et al., 2009). In line with this reasoning, present results suggest that if SF tasks are presented in such a way that participants have problems understanding their causal structure, they make use of the simple pattern matching heuristic. If, however, tasks are presented in such a way that participants can detect their causal structure (verbal tasks), participants are able to arrive at more complex inferences.

Limitations and Future Directions

The question may be raised as to whether our SF tasks were too easy, especially because of the exclusive reliance on multiple-choice answer formats with only three answer options. It seems plausible that questions involving the selection of an option are easier to answer than questions requiring the construction of an answer. Nevertheless, average correct solution rates were clearly over 30% guessing rate, implying at least that most participants were able to understand the qualitative gist of SF problems when they were presented verbally. Participants also showed a systematic error in constructing line graphs (parallel lines as representing equal in- and outflows) that was only detected *because* of the easier structure of multiple-choice answers. Moreover, old participants performed equally well as young student samples that are most likely more experienced in answering multiple-choice questions. Thus, we argue that the reason for higher solution rates goes beyond the choice of multiple-choice answer formats. It is up to further research to determine, however, just how far solution rates will change when more and more difficult answer options are provided—present results demonstrate that at least a basic understanding is possible in that participants are able to distinguish between the three basic states of the SF system (increasing, stable, decreasing) and its flows (inflow bigger than, smaller than, or equal to outflow).

Furthermore, one could argue that, albeit structurally equivalent to SF tasks used previously, our verbal tasks gave away the problem structure to participants. However, even though verbal tasks differed in the extent to which the structure was made explicit to the participant in the answer options, solution rates were high even in the most difficult task: Whereas in the piggy bank scenario, the magnitude of the inflow was explicitly related to the magnitude of the outflow, in the bathtub scenario, only the inflow was mentioned in the answer options, and participants needed to establish the relation between in- and outflows on their own. In the CO_2 scenario, this relation even needed to be established for a specific amount of inflow reduction. Consequently, verbal tasks did not simply give away the problem structure, but they better enable participants to detect it.

Therefore, whether people can or cannot detect the SF structure seems to depend on how the problem is presented. This result opens a window to a range of possible research questions on the link between perception and processing of SF problems. Previously, clear links have been shown between perception and higher level cognitive processes. For example, it was shown that global perceptual attention enhances creative thinking (Friedman, Fishbach, Förster, & Werth, 2003). Likewise, global versus local perceptual attention might affect solutions to SF problems as well, because a global focus (e.g., "overall, the inflow is bigger than the outflow") is likely to result in higher solution rates than a local focus (e.g., "in Year 5, the inflow is 7Gt of CO_2"). Thus, future research could deepen our understanding of the links between problem presentation, perceptual attention, and solution strategies. The present experiment demonstrated the *existence* of such a link; future research is needed, however, to demonstrate the exact *nature* of it. Potential insights could then be used to help people understand and deal with more complex SF problems than the ones presented here (e.g., problems containing multiple in- and outflows or nonlinear trajectories).

Implications for Simulation-Based Learning and Assessment

Contrary to previous arguments (Sterman, 2008), present findings suggest that participants have an understanding of the fundamental SF structure, a finding that bears implications for simulation-based assessment and learning. In simulation studies, it was shown that people have great difficulties regulating simulated SF systems (Dutt & Gonzalez, 2007; Gonzalez & Dutt, 2011). Given the present findings, one possible explanation for this phenomenon can be excluded: It is not impossible per se for people to understand basic SF structures. Consequently, two possible explanations remain: First, while the present experiment used basic SF problems, typical simulations, in contrast, contain a range of variables and resulting interactions, putting high demand on cognitive capacity and decreasing human ability to process the system structure

(Halford, Baker, McCredden, & Bain, 2013). Second, even participants who are able to detect the simulated system structure might lack necessary procedural knowledge on how to *deal* with the system (Mislevy, 2011). Consequently, it seems necessary to investigate to what extent people's difficulties regulating dynamic and complex SF systems can be accounted for by (a) a lack of understanding of the system structure due to many interacting variables and (b) a lack of procedural knowledge.

It has been the subject of debate how simulations and games should be designed to be as effective as possible (Morgan, 2000). The present results suggest that the problem presentation could be complemented with verbal descriptions of the respective system. For example, a game on the climate system could help struggling learners by presenting additional, verbal information on how CO_2 is emitted into, and absorbed from the atmosphere. That way, the system structure would be made more accessible and the learner could proceed to more advanced questions, for example on possible actions to regulate the climate system.

Moreover, we suggest that learners' knowledge about the SF system should not only be inferred from their actions, but should additionally be assessed verbally to deliver an assessment of both knowledge and skill while working with a simulation.

It is interesting to speculate how the use of pictorial (not graphical) information might affect understanding of SF systems. For example, to visualize SF systems, one might see actual CO_2 molecules collect in the atmosphere. In contrast to scientific notations such as coordinate systems and graphs, pictorial information should not be intrinsically difficult to understand. According to cognitive load theory, however, processing of information generally uses up cognitive resources that cannot be used for processing of other information. If one piece of information can be fully comprehended on its own, additinal information such as a picture does not aid learning, but uses up cognitive resources nevertheless (Sweller & Chandler, 1994). It is possible that when additional pictorial information is given, the text or simulation is processed less intensively, and learning can even deteriorate (Rasch & Schnotz, 2009; Schnotz & Bannert, 1999). Consequently, additional pictorial information would need to add informational value that the simulation or text alone does not deliver, and it would need to deliver that information in a computationally efficient way (Rasch & Schnotz, 2009).

Implications for Communication of SF Problems

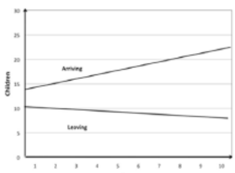
Concerning the communication of SF problems such as the accumulation of debts, or the accumulation of CO_2 in the atmosphere, we suggest that display formats used in media reports such as reports by the Intergovernmental Panel on Climate Change (IPCC) could be rendered more accessible by reducing the amount of quantitative information to a minimum. Thus far, these reports contain a large number of scientific graphs on atmospheric CO_2 (see, for example, the most recent IPCC, 2007, report). Importantly, the way information is presented not only affects the understanding of the problem, but also the quality of subsequent decision making (Covey, 2011). It was argued, for example, that people's misunderstanding of SF structures inherent to climate change could explain their lack of motivation to contribute to climate change mitigation (Sterman, 2008). Consequently, presenting SF problems such as climate change in a verbal format not only enhances people's understanding of the problem, but might, as a result, also affect their ability to decide on a correct solution, or even whether to pursue a solution.

Conclusion

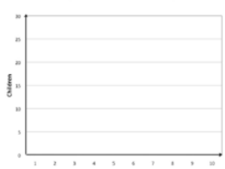
The present experiment demonstrated that people are better able to deal with SF systems if the problems are presented in a purely verbal format. This result suggests that both simulation-based learning and communication of SF problems could be rendered more effective by giving more weight to verbal information. On a more general level, these findings support the idea that people can deal with even highly complex problems if they are presented in accessible formats.

Appendix A

The following Figure depicts the number of children entering and leaving a playground.



1. How does the number of children entering the playground relate to the number of children leaving the playground?
   [] The number of children arriving equals the number of children leaving the playground.
   [] The number of children arriving is higher than the number of children leaving the playground.
   [] The number of children arriving is lower than the number of children leaving the playground.
2. If the number of children arriving at and leaving the playground relate to each other as depicted above: How will the number of children who actually are on the playground develop over time?
   [] The number of children on the playground will rise.
   [] The number of children on the playground will fall.
   [] The number of children on the playground will remain constant.
3. Assuming that the number of children on the playground will remain constant: How would the number of children arriving have to relate to the number of children leaving?
   [] The number of children arriving would have to be greater than the number of children leaving.
   [] The number of children arriving would have to be equal to the number of children leaving.
   [] The number of children arriving would have to be less than the number of children leaving.
4. Please sketch your answer to question 3 into the Figure below. (Several solutions are possible, please sketch only one).



Playground scenario of the I/P tasks (translated).
Note. I/P = interpretation and production.

Appendix B

Piggy Bank and CO_2 Scenario of the Verbal Tasks (Translated)

Piggy bank scenario. Imagine that you have a piggy bank. Each month, you throw money into the piggy bank, and you also take some money out of the piggy bank. Imagine that you want to buy yourself a book worth 20€. You count the money inside your piggy bank and notice that you currently have 10€. What do you need to do to ensure the amount of money will increase to 20€?

a. You have to take less money out of the piggy bank than you throw into it.
b. You have to take more money out of the piggy bank than you throw into it.
c. You have to take out as much money as you throw into the piggy bank.

CO_2 scenario. CO_2 emissions are caused by the burning of fossil fuels and lead to an increase of atmospheric CO_2 concentration. CO_2 absorptions are caused by forests and oceans and decrease atmospheric CO_2 concentration. CO_2 emissions are currently twice as high as CO_2 absorptions. Imagine that emissions were reduced by 30%: How would the atmospheric CO_2 concentration react?

a. Atmospheric CO_2 concentration would increase.
b. Atmospheric CO_2 concentration would decrease.
c. Atmospheric CO_2 concentration would remain constant.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was partly supported by a grant from the German Federal Ministry of Education and Research (BMBF) to Joachim Funke, grant number 01UV1012.

References

Bassok, M. (2001). Semantic alignments in mathematical word problems. In D. Gentner, K. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 401-433). Cambridge, MA: MIT Press.

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

Brehmer, B. (1976). Learning complex rules in probabilistic inference tasks. *Scandinavian Journal of Psychology*, *17*, 309-312.

Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, *4*, 75-100.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.

Covey, J. (2011). The effects of absolute risks, relative risks, frequencies, and probabilities on decision quality. *Journal of Health Communication*, *16*, 788-801.

Cronin, A. M., Gonzalez, C., & Sterman, 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.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, *62*, 198-215.

Dutt, V., & Gonzalez, C. (2007). Slope of inflow impacts dynamic decision making. In J. Sterman, R. Oliva, R. Langer, J. Rowe, & J. Yanni (Eds.), *Proceedings of the 25th International Conference of the System Dynamics Society* (pp.79-84). Boston, MA: System Dynamics Society.

Friedman, R. S., Fishbach, A., Förster, J., & Werth, L. (2003). Attentional priming effects on creativity. *Creativity Research Journal*, *15*, 277-286.

Garcia-Retamero, R., Wallin, A., & Dieckmann, A. (2007). Does causal knowledge help us be faster and more frugal in our decisions? *Memory & Cognition*, *35*, 1399-1409.

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 231-239.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684-704.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, *46*, 449-460.

Gonzalez, C., & Dutt, V. (2011). A generic dynamic control task for behavioral research and education. *Computers in Human Behavior*, *27*, 1904-1914.

Halford, G., Baker, R., McCredden, J., & Bain, J. (2013). How many variables can humans process? *Psychological Science*, *16*, 70-76.

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 37-53.

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

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science*, *290*, 2261-2262.

Intergovernmental Panel on Climate Change. (2007). *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Climate Change: The Physical Science Basis). Cambridge, UK: Cambridge University Press.

Kotovsky, K. K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*, 248-294.

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 720-733.

Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment* (CRESST Report 800). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California.

Morgan, K. (2000). Cross-cultural considerations for simulation-based learning environments. *Simulation & Gaming*, *31*, 491-508.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied*, *13*, 197-223.

O'Brien, D. P., Noveck, I. A., Davidson, G. M., & Fisch, S. M. (1990). Sources of difficulty in deductive reasoning: The THOG task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *42*, 329-351.

Rasch, T., & Schnotz, W. (2009). Interactive and non-interactive pictures in multimedia environments: Effects on learning outcomes and learning efficiency. *Learning and Instruction*, *19*, 411-422.

Schnotz, W., & Bannert, M. (1999). Einflüsse der Visulaisierungsform auf die Konstruktion mentaler Modelle beim Bild- und Textverstehen [Influences of the visualization format on the construction of mental models during picture and text comprehension]. *Zeitschrift für experimentelle Psychologie*, *46*, 216-235.

Shah, P., & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*, 43-61.

Sterman, J. D. (2008). Risk communication on climate: Mental models and mass balance. *Science*, *23*, 532-533.

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, *108*, 811-826.

Sterman, J. D., & Booth Sweeney, L. (2002). Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review*, *18*, 207-240.

Sterman, J. D., & Booth Sweeney, L. (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change*, *80*, 213-238.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185-223.

Author Biographies

Helen Fischer is pursuing a doctoral degree in the psychology graduate program at Heidelberg University, Germany. Her research interests include problem solving in the context of climate change, stock-flow reasoning, and implicit learning.

Contact: helen.fischer@psychologie.uni-heidelberg.de.

Christina Degen has studied clinical and cognitive neuroscience at Maastricht University. She is now a PhD candidate at University Hospital Heidelberg where she is researching problem solving and potential neurobiological and lifestyle factors influencing age-related cognitive decline.

Contact: Christina.Degen@med.uni-heidelberg.de.

Joachim Funke is a professor for experimental and theoretical psychology in the Psychology Department of Heidelberg University, Germany. His research interests are in all fields related to complex problem solving.

Contact: joachim.funke@psychologie.uni-heidelberg.de.