Doctoral thesis submitted to
the Faculty of Behavioural and Cultural Studies
Heidelberg University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy (Dr. phil.)
in Psychology

Title of the publication-based thesis
Understanding Dynamic Systems and Non-linearity

presented by
Helen Fischer

year of submission
2015

Dean: Prof. Dr. Klaus Fiedler
Advisor: Prof. Dr. Joachim Funke

Acknowledgements

I would like to thank my supervisor Prof. Dr. Joachim Funke for his kind support throughout these last years, and for giving me the opportunity and freedom to follow my ideas.

I spent so many hours in the office writing this thesis, and I am thankful for having it shared with Daniel and Doro who made coming to the office feel like coming to our place.

My thanks go to my friends for patiently walking with me all this way, Judith, Jenny, Verena, and Nathalie and to Anne, Myriam and Johanna who feel like family to me.

I want to thank my parents for believing in me and supporting me on the way. I am deeply thankful for Julian and Alex and his endless support in everything I do.

Table of Contents

List of publications for the publication-based thesis

Manuscript 1: Fischer, H., Degen, C., & Funke, J. (2015). Improving stock-flow reasoning with verbal formats. *Simulation & Gaming*. doi:10.1177/1046878114565058

Manuscript 2: Fischer, H., & Gonzalez, C. (2015). Making sense of dynamic systems: How our understanding of stocks and flows depends on a global perspective. *Cognitive Science*. doi:10.1111/cogs.12239

Manuscript 3: Fischer, H. & Holt, D. (2015). *When high working memory capacity and using more training information is and is not beneficial for predicting non-linear processes*. Manuscript submitted for publication.

Introduction

How do we predict the development of atmospheric CO_2 concentration over the next decades given CO_2 emissions and absorptions? How do we predict the development of our bank accounts, or the development of a start-up company? Understanding dynamic systems and predicting their often non-linear developments is of critical importance in many areas of life and has, during the last years, gained increasing interest though the emerging research domain of the psychological dimensions of climate change (Swim et al., 2011). The climate is a dynamic system that poses challenges to understanding (Amelung & Funke, 2013; Moxnes & Saysel, 2009). Correct understanding of the climate system may be a key factor, however, for accurate reasoning and decision-making about climate change (Newell, McDonald, Brewer & Hayes, 2014; Reynolds, Bostrom, Read & Morgan, 2010), and the accuracy of people's mental models of the environment is a good predictor of their degree of concern, voting intentions, and willingness to take action (Bord, O'Connor & Fisher, 2000; Bostrom et al., 2012; Guy, Kashima, Walker & O'Neill, 2013). The present research therefore tries to shed light on how and how accurately people understand dynamic systems and predict non-linear developments. Albeit previous research delivered a rather pessimistic picture of people's understanding of the behavior of dynamic systems (for a review, see Sterman, 2011) and their predictions of non-linearity (Wagenaar & Sagaria, 1975; DeLosh, Busemeyer & McDaniel, 1997; McDaniel, Dimperio, Griego & Busemeyer, 2009), I will argue and present evidence that some of the previously found difficulties can be explained by critical aspects of the research methods used in these studies, and some can be alleviated.

A common type of dynamic system that is of critical importance to climate change is stock-flow (SF) systems. SF systems contain an inflow of an entity flowing from the environment into the system, an outflow of an entity flowing out of the system into the environment, and a stock, the current level of an entity within the system. That is, a stock is an entity that accumulates or depletes over time (Jay Forrester referred to them as *levels*), flows are the rate of change in a stock (*rates*; Forrester, 1968). As such, stocks and flows are not only basic building blocks of many dynamic systems such as of births and deaths shaping a nation's population growth (or decline), a federal government running a budget deficit by spending more than it taxes, or the number of available beds in a hospital changing over time with the rate of people delivered to hospital and their recovery rate; with atmospheric greenhouse gas (GHG) concentration and the greenhouse effect, also probable causes of climate change (Crowley, 2000) adhere to an SF structure: GHG accumulate in the atmosphere as a function of GHG emissions and natural absorptions, and the energy within the climate system accumulates as a function of incoming solar radiation and reflected solar radiation. (So that the higher atmospheric GHG concentration, the more solar energy is retained within the climate system, and the more global temperature is affected.) That is, on a fundamental level, the greenhouse effect and the mechanism of how it affects global temperature follows an SF structure.

Non-linear developments in turn are an inherent property to many natural systems such as the atmosphere (Palmer, 1993). For example, it was found that surface temperature may respond nonlinearly to changes in ozone (Thompson & Solomon, 2002), ecosystems may react nonlinearly to even small changes in climate conditions (Burkett et al., 2005), and crop yields may be nonlinearly affected by changes in temperature (Schlenker & Roberts, 2008). Being able to predict non-linear developments is therefore vital for anticipating these changes, a prerequisite for adaptation planning (Burkett et al., 2005). Taken together, SF systems and nonlinear dynamics are key elements of a basic understanding of climate change, its prerequisites, and its consequences.

Two major difficulties understanding SF systems and predicting nonlinearity were made out, however. First, people have severe *difficulties inferring structural relations,* both of SF systems and of cue-criterion relations describing the over-time change of non-linear processes. It was repeatedly found that when presented with an SF system, the majority of people fail at inducing its correct system structure (e.g., Cronin, Gonzalez & Sterman 2009; Cronin & Gonzalez, 2007). This *SF failure* was found, for example, when participants were told that CO_2 emissions are currently twice as high as absorptions, and then tended to believe that if CO_2 emissions decrease by 30%, atmospheric CO_2 concentration will decrease as well. When assessing understanding of non-linear processes in the function-learning paradigm, participants are given numerical cue and criterion values in a continuous environment and asked to predict future criterion values. Similarly to a lack of rule-induction in SF tasks, the majority of extrapolations was found to be in line with the assumption that participants do not abstract a rule describing the structural cue-criterion relations (DeLosh, Busemeyer & McDaniel, 1997). Rather, most people seem to adopt simple exemplar-based prediction strategies that use only a shred of all the available training information (Kwantes, Neal & Kalish, 2012; McDaniel, Dimperio, Griego & Busemeyer, 2009).

Second, many real-life situations involve linearity, and proportional thinking and the linear model can be used to describe a range of phenomena and various dynamic systems

(Dawes & Corrigan, 1974; Funke, 1993). There is ample evidence on the "illusion of linearity" (Van Dooren et al., 2003), however, showing that students and adults tend to *over-apply linear and proportional thinking to non-linear cases* (DeBock, Van Dooren, Janssens & Verschaffel, 2002; Van Dooren, De Bock, Janssens, & Verschaffel, 2008). Most important for the present research, this over-application of linearity also seems to be the case when predicting the future development of SF systems (Guy, Kashima,Walker, & O'Neill, 2013; Moxnes & Saysel, 2009; Sterman, 2008), and when learning to predict non-linear processes (DeLosh, Busemeyer & McDaniel, 1997; Kwantes, Neal & Kalish, 2012; McDaniel, Dimperio, Griego & Busemeyer, 2009).

Specifically, in the case of SF systems, people tend to apply *correlational thinking*, apparently believing that the output of a dynamic system should be linearly correlated to its input such as believing that when the inflow is increasing, the stock should necessarily be increasing as well (Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Cronin, Gonzalez & Sterman, 2009; Gonzalez & Wong, 2012). Concerning the case of function-learning, prediction models that assume linear extrapolation based on the most similar training exemplars were found to describe human extrapolation well for different—also non-linear—cue-criterion relations (DeLosh, Busemeyer & McDaniel, 1997; Kalish, Lewandowsky & Kruschke, 2004). Interestingly, and demonstrating an over-application of linear thinking in the climate context, it was reasoned that—despite nonlinear relationships being the rule rather than the exception in biological systems—natural resource managers seem to be under the impression of linear ecological responses to climate change (Burkett et al., 2005). Thus, possibly directly related to the first difficulty, lack of induction of the system structure, people seem to apply simplifying, linear prediction heuristics when asked to predict future states of non-linear dynamics.

It is a main goal of the present research to try to contribute to our understanding of people's previously found difficulties with SF systems and non-linearity by investigating the factors that may determine whether people guide their predictions by an induced rule or by linearity heuristics. There will be three major lines of reasoning.

First, in order to convincingly establish people's apparent difficulties to infer the SF system structure, the SF failure, it is necessary to find means that allow for a valid assessment of people's mental models of the SF structure. I will argue, however, that the original task format used in previous research (e.g., Sterman, 2008, 2011), might not be seen a valid test because it contained error-prone aspects that might have lead to a systematic overestimation of correlation-heuristic use, and a systematic underestimation of people’s true understanding of SF systems. That is, in Manuscript 1, I will argue that the previously found SF failure can partially be explained by a format effect.

Second, when participants need to answer questions about higher-level system elements from lower-level system elements such as the development of a stock from given flows, participants need to relate the given lower-level elements instead of observing them in isolation. This is because on higher, macro levels, systems may possess properties that do not inhere in the elements, that are irreducible to isolated system elements (Wilensky & Resnick, 1999). This holds in even the most basic systems such as the properties of the stock of fish in a fishpond arising through the *interplay* of extraction and reproduction, so that it may posses the property of being, say, decreasing even though extraction and reproduction rate both are increasing. For understanding dynamic systems—correctly inferring their system structure—more concrete processing of isolated system elements should thus be detrimental, whereas more abstract processing of patterns of relations between elements should be beneficial. Manuscript 2 will propose a basic structural and cognitive correspondence between processing of elements versus structures in dynamic systems and hierarchical figures (global element made up of local elements): Analogously to how the behavior of a system arises through the functional relations between its elements, the global element in hierarchical figures arises through spatial relations between its elements. If this basic correspondence holds, people who tend to focus on the constituent elements of hierarchical figures might tend to fail at inducing the overall behavior of a dynamic system, while people who tend to focus on the global configuration might tend to succeed at inducing the overall system behavior.

Third, individual ability to memorize and update information about the to-be predicted process (working memory capacity) together with the structure of the process may determine which prediction strategy participants use, and it may determine success or failure of that strategy. In Manuscript 3, I will argue that people are more likely to use cognitively taxing rule-based strategies when they provide a prediction advantage: In accelerating compared to asymptotic processes. Moreover, it was implicitly assumed previously that rule-based predictions lead to higher prediction accuracy compared to linear exemplar-based strategies (McDaniel et al., 2014). However, such a prediction advantage may depend on the structure of the process as well: In asymptotic compared to accelerating environments, integrating a range of training information to infer the system structure may actually lead to overfitting and lower generalizability compared to simple linear exemplar-based strategies.

Improving stock-flow reasoning with verbal formats: Manuscript 1

It was repeatedly found that the majority of people fail at understanding dynamic stock-flow systems (Cronin & Gonzalez, 2007; Gonzalez & Wong, 2012; see Sterman, 2011 for a review). Importantly, this SF failure (Cronin, Gonzalez & Sterman, 2009) was shown to occur even in the basic case of two input variables (inflow and outflow), and one output (stock), and in simple pen-paper based tasks where system control in a simulation was not required (Cronin & Gonzalez, 2007; Gonzalez & Wong, 2012; Sterman, 2002). Given the relevance and ubiquity of SF systems ranging from national debts to climate change, people's difficulties in understanding those systems seems to pose a severe threat. However, given the intuitive claim that complexity increases as the elements relevant to solution increase (e.g., Dörner, 1983; but see Funke, 1984, 2010), and given dynamic system's simple underlying SF structure, people's poor understanding of even the most basic SF systems also seems surprising.

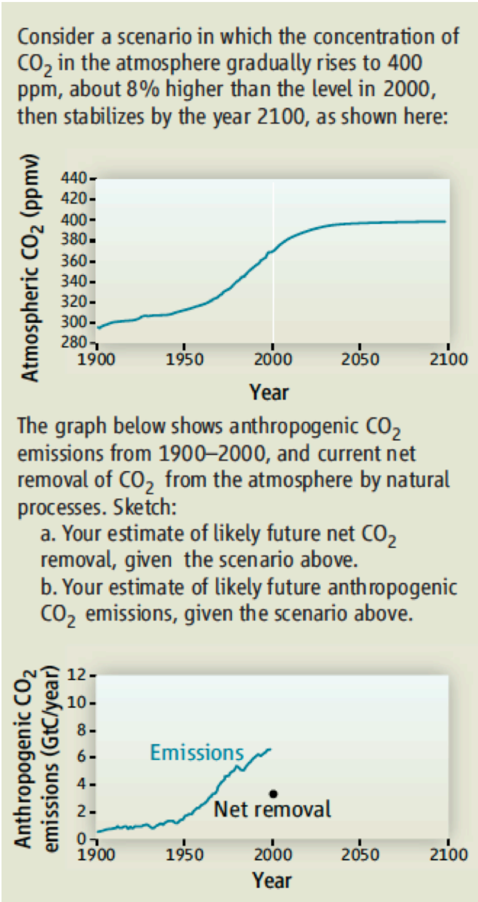
In previous tasks assessing understanding of SF dynamics (Sterman & Sweeney, 2002, 2007), referred to as the original format hereafter, participants were typically presented with a graph depicting a stock such as atmospheric CO_2 concentration stabilizing from the year 2100 onwards, and with a graph depicting previous in-and outflows such as CO_2 emissions and absorptions trajectories (Figure 1). Participants were asked to sketch emission and absorption trajectories such that a stabilizing CO_2 concentration could be achieved. A repeated finding was correlation heuristic use (referred to as *pattern matching* in graphical tasks), sketching in- and outflows that simply parallel (match) the trajectory of the stock, i.e., continuous increase followed by stabilization. SF failure was also demonstrated for multiple choice answer formats, different outcome scenarios (e.g., atmospheric CO_2 concentration stabilizing or dropping to zero), different semantic embeddings including more familiar contexts than CO_2 concentration (Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Sterman & Sweeney, 2002, 2007) and both graphical and textual display (Cronin, Gonzalez & Sterman, 2009). Given this multitude of research that convergently found SF failure, it was reasoned that there is an intrinsic difficulty to understand the structure of dynamic systems: "stock-flow failure is a robust phenomenon that appears to be a function of the mental models constructed and used when encountering a dynamic system" (Cronin et al., 2009, p. 116).

I will present evidence, however, that might call into question both aspects of SF failure, that it is (a) a robust phenomenon (addressed mainly in a previous study to

Manuscript 1: Fischer & Degen, 2012), and that (b) it is a function of erroneous mental models of dynamic systems (addressed mainly in Manuscript 1). Specifically, I will argue that the appearance of SF failure is largely contingent on specifics of the task format used previously: A substantial part of people's difficulties with SF systems can be explained by how in the original format, lower-level, isolated system elements were highly salient, biasing participants to work with these cues rather than to make use of the system structure; and how the original format contained error-prone aspects that caused difficulties in itself, leading to a systematic underestimation of people's true reasoning abilities in dynamic systems.

The first reason to doubt the validity of the original paradigm is that it arguably did not induce the impression that the system structure needs to be abstracted in the first place. Rather, the original task format might have induced the impression that one needs to process the task on a very concrete level. Concrete numbers were highly salient, and as Figure 1 shows, this was also true—or rather: *especially* true—for the "textual" version of the task (Cronin et al., 2009). However, no specific numbers are needed to infer the system structure, as the authors argue themselves: "In all cases, it is possible to answer correctly without knowledge of calculus and without carrying out any calculations" (Cronin et al., 2009, p. 126). The previously found "robustness" of the SF failure might therefore be caused by a systematic induction of erroneous reasoning strategies in all versions of the original tasks.

In a first study (N=170), we (Fischer & Degen, 2012) tested the asserted robustness of the SF failure, and whether the original format induces the impression that the task needs to be solved by calculating instead of by trying to induce the system structure. To test the robustness of the SF failure, we investigated whether solution rates vary as a function of minute surface changes in both the graphical (e.g., Sweeny & Sterman, 2000; Sterman & Sweeney, 2007) and the textual version of the original task (Cronin et al., 2009) while keeping constant the system structure. To test whether participants try to solve the task by calculating, we varied whether the initial stock was given or not while, again, keeping constant the system structure. If participants try to calculate with the given numbers, the initial stock is a relevant starting point for adding and subtracting the respective flows, influencing solution rates; if participants try to infer the system structure, however, the initial stock is an irrelevant information that should not affect solution rates. Furthermore, we hypothesized that the use of the correlation heuristic might have been induced by the way flows and stocks were displayed in the original task format because the inflow followed the same trajectory as the stock (see Fig.1). This parallel, and obviously correct, development in the past might easily induce the impression that the future development of flows and stocks should be parallel as well—that is, matching patterns.



In a department store, people enter and leave over a 30-miunte 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. The original task (Sterman & Sweeny, 2002, 2007; Sterman, 2008; Cronin et al., 2009) in the graphical (CO_2 task; left) and the textual version (department store task; right).

We developed three variations of the graphical display that were administered within participants in different scenarios (e.g., members of a club, customers within a department store), and that will be described with the department store scenario. Equivalently to the original format, participants were first shown a graph depicting the trajectory of the stock, and were then given the following task: “The following Figure depicts the number of people entering and leaving the department store. Please draw how many people must enter and leave the store in order to achieve the above stock”. First, we varied whether the initial stock (IS) was given or not (IS vs. ~IS condition). In the ~IS condition, participants only received the introductory sentences above, in the IS condition, participants received the additional information: “At the beginning, 32 customers are inside the store”. Second, we varied the presentation of in- and outflow. While in the original paradigm, the inflow was depicted as a line and the outflow was depicted as a dot (Figure 1), we created an additional condition (2L vs. ~2L condition) in which both the in– and outflow were presented as lines because a dot might induce the impression of a static quantity rather than that of a flow-variable, and because verbal protocols of a pilot study showed that many participants were indeed puzzled by the different visualization of in-and outflow. Third, we varied whether a pattern matching solution was suggested in the task display or not (PM vs. ~PM condition). In the PM condition, the inflow followed the same trajectory as the to-be completed stock, in the ~PM condition, the inflow followed a different trajectory as the to-be completed stock such as decreasing when the stock was increasing. In the textual display, we varied whether the initial stock was given, or not (with the same additional information as used in the graphical display) such that the ~IS condition was equivalent to the textual version of the original task.

As Figure 2 shows, solution rates varied drastically as a function of variations in the display (Please note that the ~IS, ~2L condition is equivalent to the original task format and may serve as a baseline; Sterman & Sweeney, 2002, 2007). Largely in line with previous results, the solution rate was 16% in the original task; it increased up to 40%, however, when the initial stock was given and both flows were displayed as lines. In the textual display, solution rates doubled from 40% in the original task to 80% when the initial stock was given, suggesting that in the “textual” version, participants are especially prone to solving the task by calculating. In contrast to our expectation, however, the suggestion of pattern matching in the task display did not affect solution rates or use of the pattern matching heuristic.

Given the sheer range of solution rates caused by minimal changes in the display (such as lines versus dots), SF failure is not a robust phenomenon that appears irrespective of the task format; rather, the occurrence of SF failure seems to be highly variable, and contingent upon how the system is displayed. If SF failure is caused by a fundamental limitation of people’s mental models (Cronin et al., 2009), however, it should be robust to such minor manipulations in the task display. Moreover, the original task display seems to bias people to answer questions about the stock by calculating rather than trying to induce the structural SF principle, suggesting that from low solution rates in the original format, one cannot necessarily conclude that people possess wrong mental models about the dynamic system, let alone an intrinsic inability to understand the SF structure. Rather, low solution rates seem to reflect people’s difficulties to solve the task by calculating. The pattern matching heuristic, however, seems to be a robust strategy that is not simply induced by the way stocks and flows are displayed.

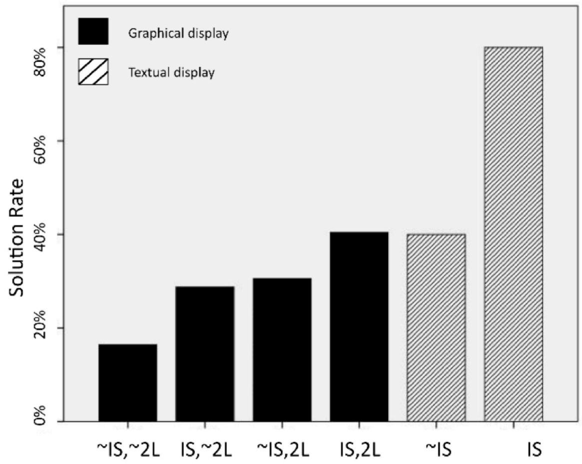


Figure 2. Solution rates for different conditions of both the graphical (left) and textual version (right) of the original SF task format. IS vs. ~IS denotes whether the initial stock is given or not, 2L vs. ~2L denotes whether the in- and outflow is depicted as two lines or not. Please note that the ~IS,~2L condition is equivalent to the original task format (e.g., Sterman & Sweeny, 2002, 2007).

The second reason to doubt the validity of the original task format is that it arguably contained error-prone aspects that are difficult to understand for participants. When assessing people's mental models, it is important to ensure that the construct-relevant variance caused by the different mental models be measured, rather than construct-irrelevant variance caused by the assessment technique itself. However, in spite of variation, different version of the original task had one thing in common: a rather scientific notation including coordinate systems and graphs, percentage values in multiple-choice answer options (e.g., "Gradually rise less than 8%"), and technical terms in both the introduction ("Currently, the net removal of atmospheric CO_2 by natural processes is about half of the anthropogenic CO_2 emissions") and the figure captions (e.g., "GtC/year"). We (Fischer, Degen & Funke, 2015) hypothesized that the original paradigm might have concealed participants' true understanding of SF dynamics, that is, their true mental models of the system and investigated whether the original format assessed not only understanding of the causal system structure, but also of the task format itself. In our first study (Fischer & Degen, 2012), we found that use of the pattern matching heuristic is not simply induced by the task display. Since people generally seem to prefer making use of the causal structure of their environment whenever possible (*rule-bias*; Ashby, Alfonso-Reese, Leola, & Waldron, 1998; Juslin, Olsson & Olsson, 2003), we tested whether correlation heuristic use is reduced if detection of the causal SF structure is made easier. Moreover, leaning on our previous results how the original format seems to bias participants to approach the task by calculating, we tried to find a task format that highlights relations between elements, and tested whether such a format makes it more likely for participants to induce the system structure. In a word, it was the aim of Manuscript 1 to show how the previously robust SF failure can at least partially be explained—and how it can be alleviated.

In Manuscript 1, we investigated whether the previously well-established SF failure can partially be explained by a format effect. Participants (N=107) completed SF tasks structurally similar to those used in previous research. Participants were given a graph depicting in-and outflow of a system, and then completed questions about the system, measuring system understanding (e.g., “If CO_2 emissions relate to CO_2 absorptions as depicted in the Figure above: What happens to atmospheric CO_2 concentration?”). In a within-subjects design, participants submitted their answers first in a multiple choice answer format (e.g., “CO_2 emissions would have to be equal to CO_2 absorptions”), and subsequently in the typical answer format, that is, by graphically sketching their answer into a coordinate system. Participants also completed SF tasks in a completely verbal format that avoided specific numeric information and did not entail coordinate systems at all (e.g., “CO_2 emissions are currently twice as high as CO_2 absorptions. Imagine that emission were reduced by 30%: How would atmospheric CO_2 concentration react?”). Results showed that, first, *for the exact same tasks and participants*, solution rates were significantly reduced when participants needed to submit their answer graphically. An especially telling mistake was that in the typical stabilization task, participants who correctly gave the answer that the inflow would have to equal the outflow, sketched two parallel and clearly separate lines—the typical pattern matching result. Moreover, the majority of participants arrived at the correct solution when the tasks were given in the verbal format. Our results thus suggest that (a) low solution rates found previously (“SF-failure”) can partially be explained by a format effect, namely how participants had to sketch their answer into coordinate systems (b) the typical pattern matching result can partially be explained by a format effect as well, namely how people tend to represent the relation “equal to” as parallel lines; and (c) the majority of participants is able to come to correct conclusions about the behavior of SF systems when presented in verbal formats.

Making sense of dynamic systems: How our understanding of stocks and flows depends on a global perspective: Manuscript 2

The results of Manuscript 1 suggest that the original task format contained error-prone aspects, thereby underestimating people's true understanding of dynamic systems. When these error-prone aspects were avoided, the majority of participants was able to correctly answer questions about SF systems that were structurally identical to the ones used in previous research (see Sterman, 2008, 2011). However, there might be an additional explanation for the increased solution rates: It seems plausible that by our framing of the verbal questions where we explicitly referred to *relations* between system elements (e.g., "CO_2 emissions are currently twice as high as CO_2 absorptions"), we might have changed people's representations of the task from a concrete level of isolated elements to a more abstract level of over-time relations between elements. Similarly, Fischer and Degen (2012) found that solution rates varied drastically as a function of minor changes in surface details of the task format. A possible interpretation is that in the original task format, participants focus on peripheral details that are irrelevant to the system's central feature, its SF structure. Moreover, this study showed that in the original task format, participants might have been biased to calculate, rather than to try to abstract the system structure. It is the aim of Manuscript 2 to specifically investigate the hypothesis that processing information on different levels of abstraction affects system understanding. We (Fischer & Gonzalez, 2015) hypothesized that more concrete processing of isolated system elements is detrimental, whereas abstract processing of overarching patterns is beneficial for system understanding.

What do different levels of abstraction refer to? Burgoon, Henderson and Markman (2013) argue that the term abstraction has been defined and used in a number of different ways in psychological literature. However, the authors also argue there is a common theme in the different usages of abstraction, namely that it refers to a "process of identifying a set of invariant central characteristics" (p.502). Given this definition, more abstract thinking is related to an act of information reduction that entails identifying stable, defining aspects; less abstract thinking in turn would be related to focusing on more variant, peripheral details. As the word *ab-straction* already implies, going up increasing higher level of abstraction implies leaving away, disregarding these increasingly variant properties. When abstraction is studied in paradigms employing visual perception, often hierarchical figures are used that consist of a global configuration made up of local elements. Usually the term global-local processing is used to refer to processing the stimulus at different levels of the stimulus structure. I will thus use the terms abstract as opposed to concrete processing when speaking of information processing on different levels of abstraction in general, and the terms global-local processing mostly when referring to the processing of visual stimuli at different levels of the stimulus structure in particular.

It is possible to attend to stimuli in two fundamentally different ways: one can attend to events by focusing on elements, or by focusing on groups of structurally related elements. In Psychology, this idea received much attention since Navon introduced the distinction between global and local processing in his letter task (Navon, 1977, 2003; for a review, see Kimchi, 1992). In a series of experiments, Navon presented participants with hierarchically organized letters (large, global letters made up of smaller, local letters) and asked participants to name an auditory-presented letter. A converging result was that participants' auditory discriminations were interfered by the global, but not the local letter. Navon concluded that visual perception is organized by global precedence, proceeding from global structures to more and more fine-grained details. More recent research suggests differences in global-local processing between groups (e.g., musicians, Stoesz, Jakobson, Kilgour & Lewicky, 2007; or people from collectivist cultures, Oishi et al., 2014), stable individual preferences (e.g., showing a test-retest stability of r=.79 over 7-10 days, Dale & Arnell, 2013), and differences caused by experimental manipulation (e.g., by having participants repeatedly name global or local letters, Macrae & Lewis, 2002). Most important for the present hypothesis that more abstract processing should be beneficial for understanding dynamic systems, it was repeatedly found that experts process information more abstractly, focusing on stable patterns whereas novices process information more concretely, focusing on specific details that are readily observable (Ferguson-Hessler & DeJong, 1990; Hmelo, Holton, & Kolodner, 2000; Hmelo-Silver & Pfeffer, 2004).

For example, when asked to answer questions about the weather system given weather maps, expert meteorologists compared to novices tend to integrate information more, and tend to focus less on specific elements of the maps (Trafton, Marshall, Mintz & Tricket, 2002). This kind of expert reasoning appears to hold across a range of research and domains. Expert teachers formed more connections among pieces of information when judging photos of class room interactions, and they categorized them more in meaningful problem units than novices (Carter et al., 1988), and expert biologists integrated structural, functional, and behavioral elements when freely drawing and explaining an aquaculture, capturing the global, dynamic interdependencies of the system, while hobbyists focused on perceptually available, static components of the system, providing more local, specific, and focused explanations of the system (Hmelo-Silver & Pfeffer, 2004). Taken together, novices seem to focus more on confined, readily observable and highly specific properties of the system, whereas experts tend to build mental models around integrated, less salient and less specific properties of the system, organizing incoming information more around stable and overarching patterns. In a word, experts compared to novices seem to process systems on a more abstract level.

In Manuscript 2, we test the idea that more abstract processing is beneficial for understanding SF systems, not only because of expert-novice processing differences, but also because—given the definition of abstract processing as defining invariant central characteristics—in the case of a dynamic system, this should entail defining the system structure. “Different” SF systems behave according to their common and invariant property—the SF structure—be the elements CO_2 molecules in the atmosphere or customers in a department store. In one sense, it is thus trivial or even analytic to say that “abstract processing” should be beneficial for system understanding—if abstract processing simply refers to inducing the system structure, and induction of the system structure is then used as a measure for “system understanding”. In order to test the hypothesis that system understanding benefits from abstract processing in a non-trivial way, one should find ways how abstract processing *transfers* from processing of the dynamic system to processing of another task.

We propose a structural and cognitive correspondence between hierarchical figures and dynamic systems: Similarly to how hierarchical figures consist of a global property that does not inhere in isolated elements (but in their spatial relations), the behavior a dynamic system as a whole does not inhere in isolated elements (but in their structural relations). In that sense, then, abstract processing of dynamic systems is to global processing of hierarchical figures as concrete processing of dynamic systems is to local processing of hierarchical figures: Similarly to how one may attend to isolated letters in the Navon task, one may attend to isolated elements of dynamic systems, and similarly to how one may attend to groups of structurally related letters, one may attend to groups of structurally related system elements. In short, a dynamic system may be seen as a structured unit very much in the same sense as a hierarchical figure. If this fundamental claim holds, people who tend to focus on the constituent elements of a hierarchical figure should fail at inducing the system behavior, while people who tend to focus on the global configuration of a hierarchical figure should succeed at inducing the system behavior.

More abstract as opposed to more concrete processing might not only exist as a relatively stable, individual preference for information processing, it might also be possible to treat abstract processing as a cognitive process that can be induced in participants. Cognitive processes, the manipulation, transformation, or reorganization of content (Janiszewski & Wyer, 2014), can differ in accessibility, depending on what processes have been used recently (Gollwitzer, 1999). Currently used operations are more likely to be highly activated, and are therefore more likely to be re-used for processing of novel information (Freitas, Gollwitzer & Trope, 2004; Shen & Wyer, 2008). That is, processes that are activated through previous tasks have a higher probability to be activated in subsequent tasks. Importantly, as opposed to content priming such as semantic priming, goal, affective or behavioral priming, such *procedural priming* effects are supposed to be content-free (for a review, see Janiszewski & Wyer, 2014).

To demonstrate effects of procedural priming, one needs to show that a cognitive process that is used in one task is more likely to be reused in a subsequent task, and this increased likelihood needs to hold even if the second task is unrelated content-wise. An often-replicated example of process priming is structural priming. Structural priming refers to the increased likelihood of using a syntactic structure in a sentence because of its structural parallels to previously encountered sentences. Such repetition occurs in the absence of similarities in content or sound (Bock, 1986; for a review, see Pickering & Ferreira, 2008). That is, structural similarities are transferred from one sentence to the next, while surface constituent similarities are disregarded. In one of the earliest demonstration of structural priming (Bock, 1986), participants repeated prime sentences and then freely described target pictures that were unrelated content-wise. Results showed that participants were more likely to reuse the target syntactic structure of the prime sentence rather than an alternative structure (such as an active rather than a passive voice). In another study, participants listened to auditorily presented sentences with or without a target structure and then described aloud a given picture in one sentence. Results again showed that that the target-primed condition used the target structure more often than the alternative-primed baseline, implying structural persistence across modalities (Bock, Dell, Chang & Onishi, 2007).

This shortly reviewed evidence suggests that it is possible to activate structural aspects of language such that they are more accessible and more likely to be used in subsequent sentences, while disregarding semantic and sound of the sentences. In Manuscript 2, we (Fischer & Gonzalez, 2015) propose the existence of a similar effect for the processing of dynamic systems: That it is possible to procedurally prime participants with a semantically unrelated, previous task such that they are more likely to attend to the structure of the system as opposed to its elements in a subsequent SF task.

In Manuscript 2, we test whether and how global-local processing of hierarchical stimuli transfers to processing of dynamic systems, affecting systems understanding. To measure systems understanding, participants (N=148) completed an often-used SF task, the department store (DS) task (e.g., Cronin et al., 2009). We found that (a) individual differences in processing hierarchical figures (Kimchi-Palmer figures) was related to understanding of the DS task such that people who tended to focus on isolated elements in hierarchical figures were less likely to understand the system's overall behavior than people who tended to focus on the global structure of the figures; (b) individual differences in processing the Kimchi-Palmer figures were related to correlation heuristic use such that local processors tended to use the heuristic more than global processors; (c) the format of the system affects system understanding such that understanding is increased when the answer format highlights relations between elements as opposed to isolated elements; and (d) procedurally priming participants affected system understanding such that participants who repeatedly focused on visual displays in their entirety were more likely to correctly infer the system's behavior compared to participants who focused on specific details of the same displays. These results converge towards global-local processing as a cognitive explanation for when and why SF failure and correlation heuristic use does, and does not occur.

When high working memory capacity and using more training information is and is not beneficial for predicting non-linear processes: Manuscript 3

The previous chapters were mainly concerned with how, in general, interaction between system elements may lead to system properties that do not inhere in the elements (such as the number of people in a department store increasing even though the number of people entering and leaving are both decreasing). What has been largely disregarded, however, is the specific property of non-linearity that is a property of many real-world systems (Wu & David, 2002). In non-linear systems, interaction between system elements produces non-additive effects—effects that have a numerical value that is not equal to the sum or difference of their component parts—so that the output of the system is not directly proportional to its input. Given our strong temptation to apply linear prediction strategies and proportional thinking to non-linear environments (for a review, see Van Dooren, Janssens, De Bock & Verschaffel, 2008), however, predicting non-linear developments is usually less accurate than predicting linear developments (for a review, see Busemeyer, Byun, DeLosh & McDaniel, 1997), and typically leads to a dramatic underestimation of non-linear dynamics known as trend-damping (Harvey & Bolger, 1996; see Armstrong, 2006 for a review).

Inaccuracy of people's predictions is especially pronounced in the case of exponential developments, a fundamental class of highly nonlinear processes in nature. People typically underestimate the rate of exponential change by far, sometimes underestimating the system's output by up to 90% (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979). This typical misconception of exponential time series was found in various modes of data presentation, such as numerical, graphical (Wagenaar & Sagaria, 1975), or directly observable (Ebersbach, Lehner, Resing & Wilkening, 2008), and for different semantic embeddings (Keren, 1983; Kemp, 1984). While it might seem like a rather academic problem to anticipate exponential developments, it is also of high practical importance for accurate reasoning about the prerequisites of climate change and its future development: Three important greenhouse gases—carbon dioxide, nitrous oxide, and methane—have increased exponentially since approximately 1800, the time of the industrial revolution (IPCC, 2014).

An explanation for people's typical underestimation of nonlinear processes was delivered in previous research (DeLosh, Busemeyer & McDaniel, 1997) arguing that the majority of people adhere to exemplar-based, linear predictions irrespective of the function form—that is, even in the case of exponential functions. In the present manuscript, I address the question of how people predict exponential functions, arguing that people do not simply use exemplar-based, linear prediction strategies in general. Rather, people seem to gather some, albeit ill-calibrated, rule knowledge depicting the structural properties of exponential processes.

The issue of how participants extrapolate has long been the subject of debate. Rule-based models assume that during training, participants abstract a global rule describing the ensemble of training information (McDaniel & Busemeyer, 2005). The process of rule-abstraction is comparable to the formation of a regression equation, as participants use the feedback provided to adjust the coefficients of their rule (Kwantes et al., 2012; Bott & Heit, 2004). Linear exemplar-based models such as, most prominently, the hybrid *extrapolation association model* (EXAM; DeLosh et al., 1997) propose exemplar-based learning together with generic linear extrapolation. Participants store single predictor-criterion instances in memory, and when presented with a novel predictor, they retrieve the two nearest-matching training instances and extrapolate linearly through these two stimuli. Linear exemplar-based models were highly successful in previous model tests, accounting better for participants' extrapolations than rule-based models (Kalish, Lewandowsky & Kruschke, 2004).

People might be more likely to use the cognitively more taxing strategy of rule-induction compared to linearity heuristics (Hogarth & Karelaia, 2007), when this provides a prediction advantage: In the case of highly non-linear processes. In previous research, however, the most non-linear function class—the exponential—was only represented by an asymptotic growth function, that is, an increasingly linear processes (see Figure 3; DeLosh et al., 1997). The asymptotic nature of this function could have biased participants to use EXAM, simply because its trajectory can be well-approximated linearly compared to the increasingly *non*linear trajectory of the prototypical case of accelerating exponential growth. It therefore seems plausible that the use of asymptotic stimuli lead to an overgeneralization of the validity of EXAM, and that when accelerating functions are used, more people deviate from generic linear prediction models.

Importantly, in contrast to studies that tested specific model instantiations (DeLosh et al., 1997; McDaniel & Busemeyer, 2005), we aimed at investigating strategy-use by testing basic assumptions that most models within one class (linear exemplar-based versus rule-based) share. This approach allows for conclusions about not only one model, but about several models sharing the same basic assumptions. Prediction accuracy cannot necessarily distinguish between rule-and exemplar-based strategies—rules can be ill-calibrated and still be rules. Predictions' (non-) linearity cannot decisively distinguish either, since exemplar-based models incorporate linear predictions, but when using rules, people typically integrate information linearly as well (for a review, see Brehmer, 1994). Rather, rule-based predictions are *information-intensive* in that they integrate a range of training data describing the overall function form; exemplar-based and hybrid predictions are *information-frugal* in that they rely on the closest-matching training exemplars only. To investigate strategy-use, we compared how participants extrapolate a given exponential function compared to a quadratic function that is identified on the two exemplars most similar to the extrapolation region. Each function pair is thus identical in exemplars most similar to extrapolation values, but differs in rules. Consequently, information-frugal models predict identical extrapolation for each pair, whereas information-intensive models predict differing extrapolations.

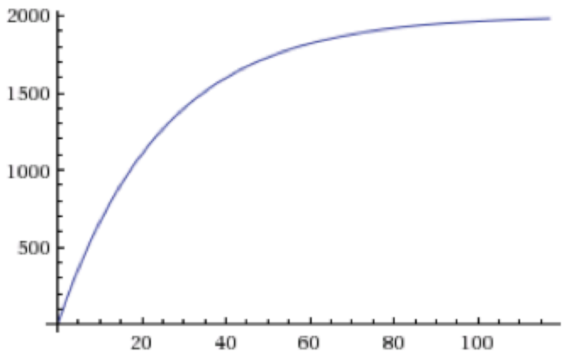


Figure 3. Exponential stimulus function with an increasingly linear trajectory as used in previous research (DeLosh et al., 1997).

While these arguments address how people extrapolate in general, it seems reasonable to speculate that individual differences in working memory capacity (WMC) might influence strategy-choice in function learning. First, a growing body of literature demonstrates the importance of WMC for accuracy and speed of learning in a structurally similar task: categorization of qualitative stimuli (Craig & Lewandowsky, 2012; Lewandowsky et al., 2012). Second, it seems theoretically plausible that WMC should play a key role for the induction of rules given that participants need to memorize previous cue and criterion values, estimate their differences, and update the weights of their derived rule accordingly. These processes arguably encompass storage and transformation—key facets of WM (Oberauer, Süß, Schulze, Wilhelm & Wittmann, 2000). In categorization and multiple-cue judgments, involvement of WMC is known to be high when more cues need to be considered, or more complex rules need to be abstracted (Juslin, Karlsson & Olson, 2008; Mata et al., 2012).

Reasoning by analogy to the conceptually similar case of function-learning, involvement of WMC should be especially pronounced in the present case of non-linear functions that involves integrating a range of cues into potentially highly complex rules (e.g., exponential). Third, a connection between extrapolation strategies and memory in general has long been assumed given that participants need to memorize the previous states of the process in order to forecast its future states (Mackinnon & Wearing, 1991). Despite these compelling reasons, the relationship between WMC and extrapolation is still very unclear and had (at the time of the pretests described below) not been investigated at all.

In two pretests (n=45 each) that will be described together given their strong conceptual relation, we systematically investigated the influence of asymptotic versus accelerating exponential growth and decay functions on strategy-use in a typical function-learning paradigm (e.g., see Kwantes et al., 2012). The first test employed prototypical exponential growth and decay functions (i.e., accelerating increasing and asymptotic decreasing functions), the second test employed inverted exponential functions (i.e., asymptotic increasing and accelerating decreasing). In both tests, function shape (accelerating vs. asymptotic) was varied between, and function direction (increasing vs. decreasing) was varied within participants. In the second tests, we also investigated the influence of WMC (assessed with digit span forward and backwards) on strategy-choice and prediction accuracy in an exploratory manner. We expected participants high in WMC to be able to consider and manipulate more training information and thus to be more likely to abstract a rule describing the function trajectory. Consequently, high compared to low WMC individuals' prediction strategies should (a) deviate more from exemplar-based, linear predictions, and (b) be more accurate.

Participants were told that they were to learn how a stock has developed over time using feedback as a guide, and that they were to predict the further development of that stock. The specific nature of the stock was not presented to avoid confounding with domain-specific background knowledge. During the training blocks of each process, successive time values were presented as numbers, and participants entered their prediction of the stock value as a number into an input box (see Kwantes et al., 2012). Participants were then given feedback in the form of the correct stock at that time (again presented as a number; functions were not shown graphically in order for participants to have to rely on WM). During the testing block, participants predicted the stock for new time values without feedback.

We found promising connections between WMC and strategy-choice. WMC was connected to the absolute prediction difference within function pairs for the asymptotic increasing function, $r(43) = .28$, $p = .03$, but not for the decreasing function, $r(43) = -.10$, $p = .26$. As opposed to the majority of participants, for participates highest in WMC (> 75 percentile), predictions of the asymptotic increasing function ($M = 2350$, $SD = 75$) differed from predictions of its matching quadratic function ($M = 2100$, $SD = 27$), $F(1,171) = 10.25$, $p = .002$, partial $\eta^2 = .06$. At least for the increasing function, WMC thus seems to be connected to strategy-choice: high compared to low WMC participants predicted different trajectories for an exponential as opposed to its matching quadratic function, suggesting that they integrated more training information into their prediction strategies than assumed by information-frugal models.

Interestingly, however, the connection between WMC and prediction accuracy was completely at odds with our expectations: Albeit using more information-intensive strategies, high compared to low WMC individuals' predictions were *less* accurate for the asymptotic growth function (again no significant relationship for the decreasing function). As Figure 4 shows, high WMC individuals systematically overshot when prediction the asymptotic growth function, thereby apparently over-applying its steep training trajectory to its (flat) extrapolation region. Low WMC individuals' simple exemplar-based strategies, in contrast, did not significantly deviate from the correct function.

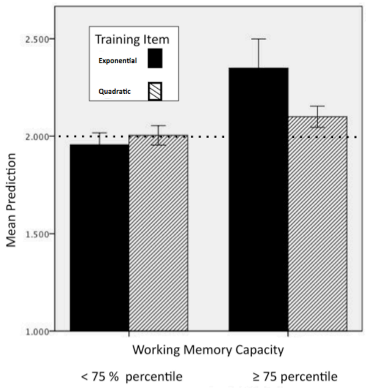


Figure 4. Results from a pretest to Manuscript 3: Mean predictions for low (<75% percentile) compared to high (>75% percentile) WMC individuals after training with an asymptotically increasing compared to its matching quadratic function. Dotted line represents the correct function value. Error bars represent 95% confidence intervals.

Given the promising and interesting findings from the pretests on how WMC relates to prediction strategy-use and accuracy, for the main study, we systematically investigated this relationship for the whole set of exponential functions of the basic form

$$f(x) = \left(a_{sign} \cdot a\right) \cdot \; e^{\left(b_{sign} \cdot b\right) \cdot (x+c))} + d$$

where a_{sign}, $b_{sign} \in \{1, -1\}$ and a, b, d, e $\in \mathbb{R}$ and x $\in \{0 \ldots 100\}$. Variable x is the input, a_{sign} determines whether the function is *increasing* or *decreasing* and b_{sign} determines whether the function is exponentially *accelerating* or *asymptoting* (i.e., decelerating) towards a constant. Given results from the pretests, we expected high WMC individuals to be more likely to deviate from simple exemplar-based accounts, employing information-intensive prediction strategies, which should pay off in highly non-linear (accelerating) functions, but might also lead to systematic overshoot in increasingly linear, (asymptotic) functions. In a word, it was the aim of Manuscript 3 to systematically assess prediction strategy-use and accuracy and how they relate to WMC in environments that differ in their degree of non-linearity.

In Manuscript 3, we systematically investigated whether the shape of the to-be predicted process influences prediction strategy-use, and how WMC (assessed with digit span forward and backward and letter-number sequences) relates to prediction strategy-use and accuracy for the complete set of basic exponential function types (accelerating, asymptoting, increasing, decreasing). In a function-learning experiment (N=296), we found that (a) predictions of approx. half of participants differed within function pairs, suggesting a basic understanding of exponential versus quadratic processes; (b) high compared to low WMC individuals were more likely to deviate from exemplar-based, linear prediction strategies, and tended to integrate more training information into their predictions than assumed by these models; and this tendency was related to (c) higher training calibration in all processes, and to higher prediction accuracy in quadratic and accelerating processes; but also to (d) worse prediction accuracy in the asymptotic processes compared to low WMC individuals' exemplar-based, linear prediction strategies. These results suggest that exemplar-based, linear predictions are not a generic strategy as was found previously; rather, many participants seem to be able to capture the fundamental properties of even exponential functions. And given that higher calibration to training data together with lower generalizability to novel stimuli is a typical result for algorithms that overfit the training data compared to simpler algorithms, these results suggest that high WMC individuals were prone to overfitting: In asymptotic processes, their more information-intensive strategies were outperformed by the simpler and more robust prediction strategies employed by low WMC individuals.

General Discussion

It was the goal of the present research was to shed light on the factors that shape systems understanding and prediction of non-linear developments. Of special interest was to determine factors that influence whether people predict the behavior of dynamic systems and non-linear processes by making use of the underlying structural regularities, or by employing simple, linear heuristics. There are eight core findings.

Summary of core findings

First, the previously well-established finding that people show an intrinsic difficulty to infer correct mental models of SF systems—SF failure—is not a robust phenomenon. It rather seems to be a highly variable phenomenon that varies with minute changes in surface features of the task display. Second, people's ability to infer the structure of SF systems was underestimated, and correlation heuristic use was overestimated in previous research because the original task format contained error-prone aspects (construction of graphs into coordinate systems). Third, if a verbal task format is used that highlights relations between system elements, use of the correlation heuristic is reduced, and SF failure (almost) disappears. Fourth, there is a correspondence between the way people process hierarchical figures and the way they process dynamic systems: people who tend to focus on elements in hierarchical figures tend to use the correlation heuristic more, and tend not to understand the overall system behavior; people who tend to focus on the global structure in hierarchical figures, tend to use the correlation heuristic less, and tend to understand the system behavior. Fifth, abstraction of the system structure is a cognitive process that can be induced with a semantically unrelated task, influencing SF failure in a subsequent task. Sixth, people adopt their prediction strategy to the demands of the environment: In environments that can be well-approximated linearly, more people use information-frugal heuristic strategies than in highly non-linear environments that demand more information-intensive strategies. Seventh, rule-based prediction strategies may not always lead to higher prediction accuracy compared to exemplar-based, linear prediction heuristics as was implicitly assumed previously (McDaniel, Cahill, Robbins & Wiener, 2014). Rather, rule-based predictions are more accurate compared to linear heuristics in accelerating environments, but are outperformed by linear heuristics in asymptotic environments. And eighth, high WMC individuals tend to capture better the underlying structural regularities of non-linear environments, leading to better prediction accuracy in complex, accelerating environments; in simple, asymptotic environments, however, this tendency leads to overfitting to training data, and to inferior prediction accuracy compared to low WMC individuals' computationally simpler exemplar-based strategies.

To assess people's understanding of the SF structure in the original task format, participants needed to submit their answers by sketching graphs into coordinate systems. As Fischer and Degen (2012) and Manuscript 1 (Fischer, Degen & Funke, 2015) showed, however, this answer format caused specific problems in itself, leading to a considerable underestimation of people's true understanding of the SF structure. While most participants were able to correctly answer questions about the system behavior when submitting their answers in a multiple-choice answer format, solution rates were reduced by approximately 40% when submitting this same answer by sketching the respective graph. An especially interesting finding was that even the correlation heuristic can partially by explained by the original answer format. After giving the (correct!) verbal answer that in-and outflow need to be identical, 22% of our participants sketched two parallel and clearly separate (>0.5cm) lines. This answer pattern coincides with the typical finding of linear extrapolation from previous developments of the flows (pattern matching), suggesting that for a part of participants, the repeated finding of pattern matching solutions in previous research cannot be interpreted as a sign of wrong mental models about the system structure. In other words, it's not only that "poor understanding of accumulation leads to serious error on reasoning about climate change" (Sterman, 2008, p.532), but that, at least for some, poor understanding of the construction of graphs leads to serious error in the original SF task format: Both the SF failure in general, and the pattern matching heuristic in particular can in parts be explained by a format effect.

It was found previously that the majority of participants does not acquire an understanding of the structural relations of non-linear developments, that is, they do not abstract rules describing the function form (DeLosh, Busemeyer & McDaniel, 1997; McDaniel, Dimperio, Griego & Busemeyer, 2009). Given that rules can be ill-calibrated and still be rules, in Manuscript 3, we did not use prediction accuracy as a measure for rule-induction (McDaniel et al., 2014), but compared predictions of an exponential function to that of a quadratic twin function identified at the training exemplars most similar to extrapolation values. In line with the typical misperception of exponentiality (Wagenaar & Sagaria, 1975; Wagenaar & Timmers, 1979), participants deviated from the correct exponentially accelerating functions; approximately half of participants did, however, differentiate between exponential functions compared to their quadratic twins—a result that is clearly at odds with the assumptions of linear exemplar-based accounts. Moreover, these participants deviated from simple linear predictions in the direction of the correct function, implying that they must have acquired a basic understanding of exponentially accelerating as opposed to quadratic functions. In so far as the abstraction of global features of the to-be-predicted function is a sign of rule-induction, we argue that participants do acquire rule knowledge (albeit ill-calibrated) about exponential as opposed to quadratic functions. In other words, not acquiring a correct understanding is not tantamount to acquiring no understanding.

Concerning the involvement of WMC on extrapolation in non-linear environments, our finding that high WMC fosters accuracy of predictions in quadratic and accelerating environments is largely in line with previous results stressing the importance of WMC for mathematical skill in general (Alloway & Passolunghi, 2011; DeStefano & LeFevre, 2004), and categorization performance in particular (Craig & Lewandowsky, 2012; Hoffmann, von Helversen & Rieskamp, 2014; Lewandowsky, Yang, Newell & Kalish, 2012; McDaniel, Cahill, Robbins & Wiener, 2014). To our knowledge, the present results are the first to show, however, that the beneficial influence of high WMC is only part of the story: When a simple prediction strategy was available (asymptotic environments), high WMC individuals' tendency to derive more complex rules was outperformed by low WMC individuals' tendency to rely on exemplar-based linearity heuristics. Analogously to how in multiple-cue judgment, it was argued that heuristics succeed when their structure matches the structure of the environment (Hogarth & Karelaia, 2007), in the present case of function-learning, linearity heuristics thus achieved optimal prediction accuracy when they matched the shape of the process in the extrapolation region.

What factors influence rule-induction versus heuristics use?

Investigating the determinants of whether people induce the structural relations underlying SF systems and non-linear processes, we tested factors in the task display and environment, and factors within the individual. Although it might seem more intuitive at first sight that describing and processing dynamic systems abstractly should hinder understanding, the present results on both individual differences and task format effects suggest the opposite: abstraction seems to foster dynamic systems understanding.

Fischer, Degen and Funke (2015) found that when presented in a verbal format, the majority of participants was able to correctly answer questions about determinants of system behavior, suggesting that they must have acquired a correct mental model about its structural relations. Error-prone aspects of the task format artificially reduced solution rates by approximately 40%, and can therefore not fully account for the improvement in solution rates from the original paradigm (16% in Fischer & Degen, 2012, which is largely in line with Sterman, 2011) to the verbal tasks, however (approx. 85% in Manuscript 1). We therefore investigated an additional factor: That by referring to *relations* between elements in the verbal tasks, we changed participants processing of the given system from concrete, local processing of isolated system elements to a more abstract, global level of the system structure.

A later experiment specifically designed to test that hypothesis largely confirmed our assumption on both the task, and the individual level (Manuscript 2). Concerning the former, we not only replicated the previous findings that SF failure and correlation heuristic use appears when isolated elements are highlighted in the task, and that SF failure and correlation heuristic use largely disappear when referring to relations between elements; we were also able to induce processing of elements as opposed to gestalts with a semantically unrelated task (maps task). Similarly, on the individual level, we found that those who tend to focus on overall gestalts as opposed to constituent elements in a hierarchical figures task, tend to use the correlation heuristic less, and tend to come to correct conclusions about the overall system behavior. In sum, the present results converge on the conclusion that that reasoning on a higher level of abstraction reduces heuristic use, and fosters understanding of the structural relations of dynamic systems.

Previous attempts to increase understanding of dynamic systems in general and the climate system in particular focused on increasing analytical thinking, and motivation to deliberate (Cronin et al., 2009; Newell, 2012). For example, in a series of five experiments trying to increase people's understanding of basic dynamic systems presented in pen-and-paper form (similar to those used in Manuscript 2), Newell (2012) hypothesized that overcoming heuristics required more deliberative thinking and consequently encouraged more detailed processing in participants. However, increased deliberation failed to increase understanding of the system, and also failed to reduce correlation heuristic use for the majority of participants. As the present results suggest, the opposite strategy appears more promising: People's understanding of dynamic system cannot be increased by more detailed and analytical processing. On the contrary, it can be increased by more global processing. Instead of teaching calculus, it seems more promising to teach strategies of abstraction and pattern recognition; instead of encouraging more detailed processing, it seems more promising to encourage more abstract processing. That is, it seems more promising to teach people to focus *less* on the details of a dynamic system, not more.

In line with previous results (McDaniel et al., 2014), we showed that WMC is associated with integrating more training information into prediction strategies. High compared to low WMC individuals were more likely to differentiate between function twins, that is, to extrapolate using rules that describe the overall function form. While the present research and arguments focused on how the simultaneous storage and transformation facet of WM should allow for retrieval and updating of information relevant for rule-induction, it was more specifically argued that WM helps building relations between elements in order to establish structures (Oberauer, Süß, Wilhlem, Wittman, 2003). The WM facet of *coordination of elements into structures* is thought to provide access to information elements to abstract novel relations and structure, or, as the authors put it: “WM provides access to varying elements by placing them in a common coordinate system.” (p. 170). Albeit speculative in nature, our result suggest that the downside of such access might be that the higher one’s WMC, the more elements are placed into the cognitive coordinate system, so that the resulting function gets more and more complex (e.g. in the form of polynomials of increasingly higher degrees) – and potentially too complex to generalize well.

Previous explanations for systems (mis-)understanding and how they relate to abstract processing

In previous research, a multitude of reasons was discussed for people’s difficulties with dynamic systems. It was argued, first, that the behavior of dynamic systems is difficult to understand due to high working memory demand (Hmelo-Silver & Azevedo, 2006; Hmelo-Silver, Marathe & Liu, 2007). WM demand is supposed to be high because dynamic systems contain several interacting events, whose behavior participants need to simulate. Second, it was argued that many people fail at understanding the workings of dynamic systems because they tend to focus on what is directly visible (Hmelo, Holton, & Kolodner, 2000), even though many relevant structures and processes of dynamic systems are not directly observable (Ferrari & Chi, 1998). Third, people were found to have a preference for single and linear causality (Jacobson, 2001). Such a preference for simple and direct causality is critical since many dynamic systems possess interconnected, nonlinear, causality such as feedback loops. Fourth and lastly, the properties of emergent higher-level parts are difficult to understand because cause-effect relationships are not obvious; emergent properties were therefore even called ontologically distinct (Hmelo-Silver, Marathe & Liu, 2007).

Albeit highly different in nature, the previously found reasons for people’s difficulties in understanding dynamic systems are largely in line with the present result of a need for abstraction. WM demand is high when a great number of interrelated variables need to be examined. That is, WM demand is high if people focus on isolated lower-level elements of the system. If people focus on higher-level elements of the system that by definition consist of fewer elements, WM demand should be lower. Summarizing the relevant structures of the system in more abstract terms should thus be economic in terms of cognitive resources. Second, the finding that people tend to focus on what is directly observable might well be connected to people focusing on lower-level aspects of the system since it was argued that lower-level properties tend to be those properties that are more readily observable (for a review, see Burgoon et al., 2013). Instructing people in methods of abstraction could thus alleviate problems with a focus on what is readily observable. Third and fourth, people's preference for single causality and their difficulties with emergent properties can be understood as difficulties arising from reasoning over isolated elements, as opposed to interactions between elements. As Manuscript 2 shows, if people adopt a more abstract view on systems by reasoning over relations between elements, then difficulties with emerging properties can be avoided. In sum, previously found explanations for people's difficulties with dynamic systems may seem diverge; they may, however, be summarized under the idea of a need for abstraction.

Application to climate change: Barriers to understanding and potential solutions

In search for an explanation of the public's wavering opinion on climate change (Dunlap, 2013; Lewandowsky, Oberauer & Gignac, 2013; Pidgeon, 2012) given an increasing scientific consensus on the existence of climate change, its human influence, and its risks (Solomon, 2007; IPCC, 2014), it was argued that easily accessible but isolated cues might influence people's opinion of climate change (Egan & Mullin, 2012; Hamilton & Stampone, 2013;). For example, people's belief in the existence of climate change is known to increase when the day of the study is perceived as warmer than usual (Li, Johnson & Zaval, 2011), or simply when participants are seated in a room with dead indoor plants (Gueguen, 2012). Similarly, the APA task force on psychology and global climate change summarized the barriers they had identified to understanding climate change: "Climate change is a trend in averages and extremes of temperature, precipitation, and other parameters that are embedded in a lot of variability, making it very difficult to identify from personal experience. People often falsely attribute unique events to climate change and also fail to detect changes in climate" (Swim et al., 2011, p.33). Specifically, the report details four major problems:

1. Achieving appropriate understanding of climate change is difficult.

2. Climate change is difficult to experience.
3. Climate change risks are perceived as uncertain and as being in the future.
4. The costs of mitigation are certain and immediate.

As the present results suggest, relying on isolated, "unique" cues may not only influence people’s opinion of climate change; it may also influence people’s understanding of its SF structure. Moreover, by stating that climate change is an aggregate, mirrored in trends and averages rather than single incidents, the report implicitly taps on the distinction between isolated lower-and aggregated higher-level system elements, and the confusion thereof, as a major challenge. Taken together, the present results may offer potential solutions to some of the problems summarized in the task force report, specifically to (1) and (2).

Concerning the first challenge, lack of understanding of the climate, Manuscript 1 shows how understanding of the climate as a dynamic system can be enhanced: By reducing overly scientific barriers to understanding when communicating to (or, rather: with) the general public. Thus, one suggestion on how to deal with the first challenge can be:

Increase understanding of the climate and conditions for climate change by highlighting its underlying SF structure in simple words: “We are living in a global bath-tub!”

This can be achieved by two different means: First, by verbally referring attention to the SF structure, and specifically, how climate change is determined through the relation between GHG emissions and absorptions. And second, if it is necessary to display information in graphical form, by considerably simplifying the respective graphs. For example, latest IPCC reports, including the summary for policymakers (Field et al., 2014; IPCC, 2014), make ample use of highly scientific graphs, mostly employing coordinate systems, scientific jargon, and percentage values. As Fischer & Degen (2012) showed, this may direct people's attention towards these highly salient elements, and bias them to try to calculate instead of trying to infer the (arguably relevant) system structure. As furthermore shown by Manuscript 1, scientific notations can pose a barrier to understanding in themselves, significantly reducing people’s ability to correctly infer the overall system behavior. Our results suggest to use qualitative rather than quantitative means to present the climate system. Whenever possible, we therefore propose to present key elements of the climate system in words.

The second challenge—that the climate as such cannot be experienced—mirrors that the climate is an abstract phenomenon, resting on a higher level of abstraction than those elements of the climate system that can be directly observed such as local temperature, precipitation, or wind. Manuscript 2 showed that if attention is directed towards relations between lower-level system elements, people tend to understand the overall system behavior. Thus, one suggestion on how to deal with the second challenge can be:

Increase processing of the overall behavior of the climate system by inducing abstract as opposed to concrete processing of the system.

Manuscript 2 offers a solution of how this can be accomplished: The climate system should be described in such a way that people's direction is guided towards overarching patterns and groups of elements, rather than isolated, singular elements. This is an aspect similar, albeit more refined than verbal description: A description of the climate system should highlight how specific elements relate to each other and how they jointly develop over a period of time. For example, one of the most prominent and widely received effects of climate change has been the melting of glaciers and ice sheets. It is important to direct attention not only to this isolated effect, but also to effects that arise though its *connections* to other system elements such as that the melting of sea ice leads to a reduction in surface reflectivity (albedo) and thereby to a greater absorption of solar radiation. More solar radiation will accelerate warming, thus increasing the melting of ice (IPCC, 2014).

Future research

The present results suggest that processing on a higher level of abstraction might foster systems understanding. Further research could further elucidate this connection between the degree of abstraction of information processing and dynamic systems understanding. It was argued that words are represented on a more abstract level than pictures because pictures are more concrete, specific, detailed, and contain observable features, whereas words contain an inherent abstraction, referring to concepts (in German, by the way, the formation of concepts is sometimes literally referred to as *Begriff*sbildung), and to a broader range of entities than a given picture (Amit, Algom, Trope & Liberman, 2008; Amit, Algom & Trope, 2009). By describing dynamic systems verbally without explicitly referring to relations between system elements, one could disentangle the potential influence of a salient system structure versus salient system elements and a potential influence of using words versus pictures.

Results of Manuscript 2 (Fischer & Gonzalez, 2015) suggest that it is possible to induce in participants the tendency to process information on a more or less abstract level, influencing systems understanding. Albeit demonstrating the success of such process priming, its exact nature remained unclear, however. Janiszewski and Wyer (2014) develop a model of process priming as part of a spreading activation network (Anderson, 1983). Importantly, the network contains not only semantic nodes, but also process nodes. Each process node has a level of activation that depends on the frequency and regency of previous activation, and determines its likelihood of being re-activated. In this way, a previously activated process can be more accessible in a subsequent task, even if semantically unrelated (as was the case in Fischer & Gonzalez, 2015). Of course, these theoretical explanations remain speculative. Future research is needed to determine whether process priming as employed in the present research with the maps task indeed fostered more abstract processing of dynamic systems through increased activation. Given that the effect of process priming seemed short-lived, it seems important to include such process measures during completion of the system understanding task (and not thereafter). As more abstract categorization results in fewer categories, one possible approach might be to assess the numbers of categories participants use when solving the SF task, for example by having participants sort a set of systems elements into as many categories as they wish (see Lee & Ariely, 2006; Liberman, Sagristano & Trope, 2002).

Perhaps more importantly on a theoretical level, the processes that are induced with the maps task might not be cognitively equivalent to those measured with the Kimchi-Palmer figures, even though tasks both were related to system understanding in Manuscript 2. Albeit in many studies, the maps task (asking participants to focus on the shape of the map as a whole versus the specific location of its capital) was used to induce, and the Kimchi-Palmer figures were used to measure abstract as opposed to concrete processing (see Burgoon et al., 2013, for a review), there might be important differences. First, as was argued above, hierarchical figures entail global properties that do not inhere in the elements, but are a function of the spatial relations of the elements (*relational* properties; Kimchi, 1992). The shape of a map, however, is not global in the sense of a relational property, at least not as a relational property of the location of its capital (relation to what?). Future research should thus try to disentangle the unique and incremental benefit of processing relational properties as opposed to simply broad properties for understanding dynamic systems.

Based on the present results, it seems plausible to speculate that, after increasingly abstract processing of dynamic systems, participants hold fundamentally different representations of the task in memory. Similarly, fuzzy-trace theory holds that people may hold different kinds of representations in memory on fuzzy-to-verbatim continua (Brainerd & Reyna, 1990; Reyna, 2012): superficial verbatim representations such as exact numbers and meaning-based gist representations, the “substance” of information. With respect to SF systems, literal use of information might result in lowest level, categorical representations (e.g., “the inflow is five”), whereas increasingly abstract processing might result in ordinal (e.g., “in minute 5, the inflow is smaller than the outflow”) and increasingly higher-order representations (e.g., “overall, the inflow is bigger than the outflow”).

Concerning prediction of non-linear developments, Manuscript 3 showed that high WMC can be detrimental in the case of asymptotic functions that are well-predicted using simple exemplar-based strategies. Given the majority of studies showing that high WMC fosters categorization performance (e.g., Lewandowsky, 2011b, Lewandowsky et al., 2012), this is an interesting finding. We argued that high WMC individuals might actually be overfitting the training data, resulting in lower transfer accuracy compared to the simpler, information-frugal strategies employed by low WMC individuals. A benefit of simpler strategies could not be found in previous work as it employed functions that needed rule-based strategies to achieve sufficient prediction accuracy (McDaniel et al., 2014). In order to investigate whether high WMC individuals’ tendency to overfit found in the present data is a more generic phenomenon that possibly even generalizes to other and different prediction tasks, one might assess learning versus transfer performance with tasks that can, and others that cannot be solved using simple exemplar-based strategies.

While in function-learning experiments, understanding of structural cue-criterion relationships seems to hinge on over-time computation of rule parameters—and hence on WMC—radically different cognitive processes seem to be necessary for SF tasks: The present results speak for SF tasks as showing properties of an insight problem. For example, the result that a focus on Gestalts is beneficial for solution is a key feature of insight problems as opposed to analytic problems. As such, rearrangement of the problem parts is a decisive cognitive process, rendering an evident solution (“Aha! The inflow is bigger than the outflow, that is all I need to know”). If such rearranging is performed, no over-time calculations are necessary and therefore virtually no WMC resources would be needed (in fact, there is growing evidence that high WMC capacity might even impair the detection of simple solutions in insight problems; Wiley & Jarosz, 2012). Albeit predicting the output of SF systems and predicting the output of continuous developments may seem related from the outset, both tasks might differ radically in the cognitive processes and resources required.

It might prove fruitful to make use of the rationale of global-local processing in the function-learning paradigm to study people's extrapolations of continuous processes such as atmospheric GHG concentration. It was found that people’s extrapolations of such processes are typically conservative, underestimating the rate of change in the data (Lewandowsky, 2011a), the well-established trend-damping phenomenon (Armstrong, 2006). Figures displaying over-time GHG concentration are hierarchical in that they consist of a global trend (for example, an overall increase) made up of local measuring points (showing, for example, a local or recent decrease). To understand the key message of the graph, one must understand all the points on the graph not as individual points, but as collections that define a cognitive structure (Halford, Baker, McCredden & Bain, 2005). In so far as global processing measures a tendency to perceive overarching patterns as opposed to isolated elements, global compared to local processing should result in radically different predictions, and in so far as global processing may be enhanced by the way the elements are displayed (e.g., the relative size and number of local elements, Kimchi & Palmer, 1982; Kimchi, 1992), one could possibly reduce typical misperceptions of these trends, an aspect that might be highly relevant for their communication in, first and foremost, IPCC reports.

Conclusion

Stock-flow structures and non-linear developments are inherent to many real-worlds systems and large-scale problems, ranging from national debts to global climate change. It therefore seems dramatic that previous research repeatedly found SF failure, people's seeming inability to understand SF structures. The present results suggest, however, that SF failure is not a generic and robust phenomenon, and identified factors that can explain when and why SF failure does—and does not—occur: SF failure can in parts be explained by a format effect, and in parts by whether people process the system on a concrete level of disparate elements or an abstract level of structural relations. While abstract processing seems to be beneficial for stock-flow understanding, this is not necessarily the case for predicting continuous processes: Inducing structural cue-criterion relations enhances prediction accuracy for environments with medium-to high difficulty structure in the extrapolation region, but is outperformed by simple exemplar-based strategies in environments that can be well-approximated linearly.

References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137.

Amelung, D., & Funke, J. (2013). Dealing with the uncertainties of climate engineering: Warnings from a psychological complex problem solving perspective. *Technology in Society*, *35*(1), 32–40.

Amit, E., Algom, D., Trope, Y., & Liberman, N. (2008). Thou shalt not make unto thee any graven image": The distance-dependence of representation. In K.D. Markam, W.M.Plein & J.A.Suhr (Eds.), *The handbook of imagination and mental simulation* (pp. 53–68). New York: Psychology Press.

Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General*, *138*(3), 400-415.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Armstrong, J. S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting*, *22*(3), 583–598.

Ashby, F. G., Alfonso-Reese, L. A., Leola, A. & Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458.

Bord, R. J., O'Connor, R. E., & Fisher, A. (2000). In what sense does the public need to understand global climate change? *Public Understanding of Science*, *9*(3), 205–218.

Bostrom, A., O'Connor, R. E., Böhm, G., Hanss, D., Bodi, O., Ekström, F., … Salensminde, I. (2012). Causal thinking and support for climate change policies: International survey findings. *Global Environmental Change*, *22*(1), 210–222.

Bott, L., & Heit, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 38–50.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*(1), 3–47.

Brehmer, B. (1994). The psychology of linear judgement models. *Acta Psychologica*, *87*(2), 137–154.

Burgoon, E. M., Henderson, M. D., & Markman, A. B. (2013). There Are Many Ways to See the Forest for the Trees A Tour Guide for Abstraction. *Perspectives on Psychological Science*, *8*(5), 501–520.

Burkett, V. R., Wilcox, D. A., Stottlemyer, R., Barrow, W., Fagre, D., Baron, J., … Doyle, T. (2005). Nonlinear dynamics in ecosystem response to climatic change: case studies and policy implications. *Ecological Complexity*, *2*(4), 357–394.

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In *Concepts and Categories* (pp. 405–437).

Carter, K., Cushing, K., Sabers, D., Stein, P., & Berliner, D. (1988). Expert-novice differences in perceiving and processing visual classroom information. *Journal of Teacher Education*, *39*(3), 25–31.

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology*, *65*(3), 439–464.

Cronin, M. A., & Gonzalez, C. (2007). Understanding the building blocks of dynamic systems. *System Dynamics Review*, *23*(1), 1–17.

Cronin, M. A., 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*(1), 116–130.

Crowley, T. J. (2000). Causes of climate change over the past 1000 years. *Science*, *289*(5477), 270–277.

Dale, G., & Arnell, K. M. (2013). Investigating the stability of and relationships among global/local processing measures. *Attention, Perception, & Psychophysics*, *75*(3), 394–406.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95.

De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics*, *50*(3), 311–334.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968–986.

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386.

Dörner, D. (1983). Heuristics and cognition in complex systems. *Methods of Heuristics*, 89–108.

Dunlap, R. E. (2013). Climate change skepticism and denial: An introduction. *American Behavioral Scientist*, *57*(6), 691–698.

Ebersbach, M., Lehner, M., Resing, W. C., & Wilkening, F. (2008). Forecasting exponential growth and exponential decline: Similarities and differences. *Acta Psychologica*, *127*(2), 247–257.

Egan, P. J., & Mullin, M. (2012). Turning personal experience into political attitudes: the effect of local weather on Americans' perceptions about global warming. *The Journal of Politics*, *74*(03), 796–809.

Ferguson-Hessler, M. G., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction*, *7*(1), 41–54.

Ferrari, M., & Chi, M. T. (1998). The nature of naive explanations of natural selection. *International Journal of Science Education*, *20*(10), 1231–1256.

Field, C. B., Barros, V. R., Mastrandrea, M. D., Mach, K. J., Abdrabo, M.-K., Adger, N., … Yohe, W.G. (2014). Summary for policymakers. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part a: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 1–32.

Fischer, H., & Degen, C. (2012). Stock-flow failure can be explained by the task format. In Husemann, E. & Lane, D (Eds.). *Proceedings of The 30th International Conference of the System Dynamics Society*, University of St.Gallen, 2012.

Fischer, H., Degen, C., & Funke, J. (2015). Improving stock-flow reasoning with verbal formats. *Simulation & Gaming*. doi:10.1177/1046878114565058

Fischer, H., & Gonzalez, C. (2015). Making sense of dynamic systems: How our understanding of stocks and flows depends on a global perspective. *Cognitive Science*. doi:10.1111/cogs.12239

Forrester, J. W. (1968). *Principles of Systems*. Waltham, MA: Pegasus Communications.

Freitas, A. L., Gollwitzer, P., & Trope, Y. (2004). The influence of abstract and concrete mindsets on anticipating and guiding others' self-regulatory efforts. *Journal of Experimental Social Psychology*, *40*(6), 739–752.

Funke, J. (1984). Diagnose der westdeutschen Problemlöseforschung in Form einiger Thesen. Mit einigen Anmerkungen von D. Dörner. *Sprache & Kognition*, *3*, 159–172.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. *Advances in Psychology*, *101*, 313–330.

Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cognitive Processing*, *11*(2), 133–142.

Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *American Psychologist*, *54*(7), 493-503.

Gonzalez, C., & Wong, H. (2012). Understanding stocks and flows through analogy. *System Dynamics Review*, *28*(1), 3–27.

Guéguen, N. (2012). Dead indoor plants strengthen belief in global warming. *Journal of Environmental Psychology*, *32*(2), 173–177.

Guy, S., Kashima, Y., Walker, I., & O'Neill, S. (2013). Comparing the atmosphere to a bathtub: effectiveness of analogy for reasoning about accumulation. *Climatic Change*, *121*(4), 579–594.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Psychological Science*, *16*(1), 70–76.

Hamilton, L. C., & Stampone, M. D. (2013). Blowin'in the wind: Short-term weather and belief in anthropogenic climate change. *Weather, Climate, and Society*, *5*(2), 112–119.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137.

Hmelo, C. E., Holton, D. L., & Kolodner, J. L. (2000). Designing to learn about complex systems. *The Journal of the Learning Sciences*, *9*(3), 247–298.

Hmelo-Silver, C. E., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *The Journal of the Learning Sciences*, *15*(1), 53–61.

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307–331.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, 2242–2261.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*(3), 733-758.

Howe, P. D., Markowitz, E. M., Lee, T. M., Ko, C.-Y., & Leiserowitz, A. (2013). Global perceptions of local temperature change. *Nature Climate Change*, *3*(4), 352–356.

Intergovernmental Panel on Climate Change. (2014). *Climate Change 2013: The physical science basis: Working group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, MA: Cambridge University Press.

Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, *6*(3), 41–49.

Janiszewski, C., & Wyer Jr, R. S. (2014). Content and process priming: A review. *Journal of Consumer Psychology*, *24*(1), 96–118.

Juslin, P., Karlsson, L., & Olsson, H. (2008). Information integration in multiple cue judgment: A division of labor hypothesis. *Cognition*, *106*(1), 259–298.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*(1), 133–156.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: knowledge partitioning and function learning. *Psychological Review*, *111*(4), 1072–1099.

Kemp, S. (1984). Perception of changes in the cost of living. *Journal of Economic Psychology*, *5*(4), 313–323.

Keren, G. (1983). Cultural differences in the misperception of exponential growth. *Perception & Psychophysics*, *34*(3), 289–293.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychological Bulletin*, *112*(1), 24-38.

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(4), 521-535.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 811–836.

Kwantes, P. J., Neal, A., & Kalish, M. (2012). Item order matters in a function learning task. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 90–97. http://doi.org/10.1037/a0026639

Lee, L., & Ariely, D. (2006). Shopping goals, goal concreteness, and conditional promotions. *Journal of Consumer Research*, *33*(1), 60–70.

Lewandowsky, S. (2011a). Popular Consensus Climate Change Is Set to Continue. *Psychological Science*, *22*(4), 460–463. http://doi.org/10.1177/0956797611402515

Lewandowsky, S. (2011b). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 720–738.

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore,(climate) science is a hoax an anatomy of the motivated rejection of science. *Psychological Science*, *24*(5), 622–633.

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 881–904.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, *38*(6), 523–534.

Li, Y., Johnson, E. J., & Zaval, L. (2011). Local warming daily temperature change influences belief in global warming. *Psychological Science*, *22*(4), 454–459.

MacKinnon, A. J., & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica*, *76*(2), 177–191.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, *13*(2), 194–196.

Mata, R., Pachur, T., Von Helversen, B., Hertwig, R., Rieskamp, J., & Schooler, L. (2012). Ecological rationality: a framework for understanding and aiding the aging decision maker. *Frontiers in Neuroscience*, *6*(19). http://doi.org/10.3389/fnins.2012.00019

McDaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review*, *12*(1), 24–42.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, *143*(2), 668–693.

McDaniel, M. A., Dimperio, E., Griego, J. A., & Busemeyer, J. R. (2009). Predicting transfer performance: a comparison of competing function learning models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 173-195.

Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: information policies. *Climatic Change*, *93*(1-2), 15–37.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

Navon, D. (2003). What does a compound letter tell the psychologist's mind? *Acta Psychologica*, *114*(3), 273–309.

Newell, B. R., McDonald, R. I., Brewer, M., & Hayes, B. K. (2014). The Psychology of Environmental Decisions. *Annual Review of Environment and Resources*, *39*, 443–467.

Newell, B. R. (2012). Why don't people understand accumulation and does it matter for understanding (and doing something about) climate change)? [Powerpoint slides]. Retrieved from http://www.ecu.edu.au/__data/assets/pdf_file/0007/437974/Ben-Newell-Seminars-March-2013.pdf

Oberauer, K., Süss, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193. http://doi.org/10.1016/S0160-2896(02)00115-0

Oishi, S., Jaswal, V. K., Lillard, A. S., Mizokawa, A., Hitokoto, H., & Tsutsui, Y. (2014). Cultural variations in global versus local processing: A developmental perspective. *Developmental Psychology*, *50*(12), 2654–2665.

Palmer, T. N. (1993). A nonlinear dynamical perspective on climate change. *Weather*, *48*(10), 314–326.

Phillips, J. D. (2003). Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography*, *27*(1), 1–23. http://doi.org/10.1191/0309133303pp340ra

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, *134*(3), 427.

Pidgeon, N. (2012). Public understanding of, and attitudes to, climate change: UK and international perspectives and policy. *Climate Policy*, 85–106.

Platzer, C., & Bröder, A. (2012). Most people do not ignore salient invalid cues in memory-based decisions. *Psychonomic Bulletin & Review*, *19*(4), 654–661.

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in Fuzzy-Trace Theory. *Judgment and Decision Making*, *7*(3), 332–359.

Reynolds, T. W., Bostrom, A., Read, D., & Morgan, M. G. (2010). Now what do people know about global climate change? Survey studies of educated laypeople. *Risk Analysis*, *30*(10), 1520–1538.

Riffkin, R. (2014). Climate change not a top worry in us. *Gallup* (March, 2014). Retrieved from http://www.climateaccess.org/sites/default/files/Gallup_Worry%20March%202014.pdf

Schlenker, W., & Roberts, M. J. (2008). *Estimating the impact of climate change on crop yields: The importance of nonlinear temperature effects* (No. w13799). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13799

Shen, H., & Wyer, R. S. (2008). Procedural priming and consumer judgments: Effects on the impact of positively and negatively valenced information. *Journal of Consumer Research*, *34*(5), 727–737.

Solomon, S. (Ed.). (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge, MA: Cambridge University Press.

Sterman, J. D. (2008). Risk communication on climate: mental models and mass balance. *Science*, *322*(5901), 532–533.

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, *108*(4), 811–826.

Sterman, J. D., & Sweeney, L. B. (2002). Cloudy skies: assessing public understanding of global warming. *System Dynamics Review*, *18*(2), 207–240.

Sterman, J. D., & Sweeney, L. B. (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change*, *80*(3-4), 213–238.

Stoesz, B. M., Jakobson, L. S., Kilgour, A. R., & Lewycky, S. T. (2007). Local processing advantage in musicians: Evidence from disembedding and constructional tasks. *Music Perception: An Interdisciplinary Journal*, *25*(2), 153–165.

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

Swim, J. K., Stern, P. C., Doherty, T. J., Clayton, S., Reser, J. P., Weber, E. U., … Howard, G. S. (2011). Psychology's contributions to understanding and addressing global climate change. *American Psychologist*, *66*(4), 241–250.

Thompson, D. W., & Solomon, S. (2002). Interpretation of recent Southern Hemisphere climate change. *Science*, *296*(5569), 895–899.

Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). *Extracting explicit and implict information from complex visualizations*. Berlin-Göttingen-Heidelberg: Springer.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: Expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics*, *53*(2), 113–138.

Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2008). The linear imperative: An inventory and conceptual analysis of students' overuse of linearity. *Journal for Research in Mathematics Education*, 311–342.

Von Helversen, B., & Rieskamp, J. (2009). Models of quantitative estimations: rule-based and exemplar-based processes compared. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 867–889.

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416–422.

Wagenaar, W. A., & Timmers, H. (1979). The pond-and-duckweed problem; Three experiments on the misperception of exponential growth. *Acta Psychologica*, *43*(3), 239–251.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–19.

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, *21*(4), 258–262.

Williams, B., & Hummelbrunner, R. (2010). *Systems concepts in action: a practitioner's toolkit*. Stanford, CA: Stanford University Press.

Wu, J., & David, J. L. (2002). A spatially explicit hierarchical approach to modeling complex ecological systems: theory and applications. *Ecological Modelling*, *153*(1–2), 7–26. http://doi.org/10.1016/S0304-3800(01)00499-9

Declaration in accordance to § 8 (1) b) and § 8 (1) c) of the doctoral degree regulation of Heidelberg

University, Faculty of Behavioural and Cultural Studies

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg
Doctoral Committee of the Faculty of Behavioural and Cultural Studies, of Heidelberg University

Erklärung gemäß § 8 Abs. 1 Buchst. b) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften
Declaration in accordance to § 8 (1) b) and § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.
I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 Abs. 1 Buchst. c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe.
I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname
First name Family name ____________________Helen Fischer______________________

Datum, Unterschrift
Date, Signature ____

Appendix A1

Manuscript 1: Improving stock-flow reasoning with verbal formats

Simulation & Gaming
1–15

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

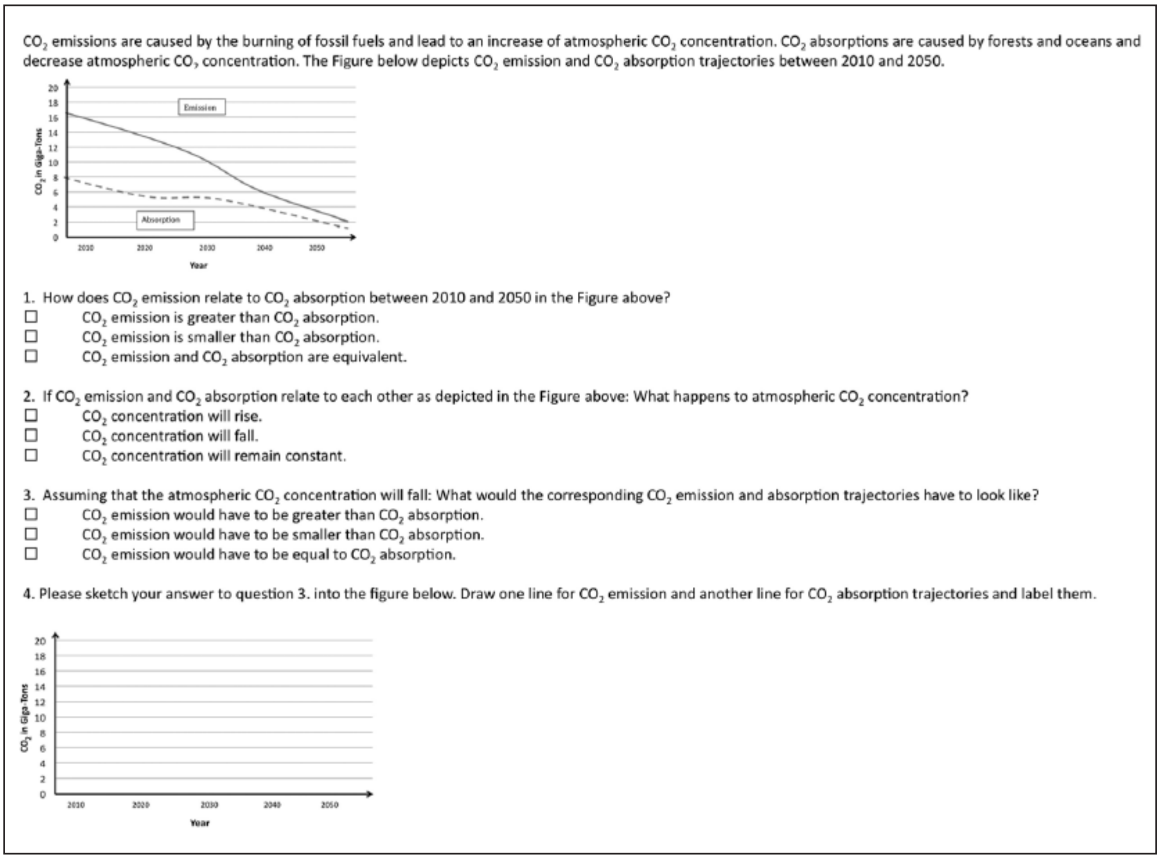
A total of $N = 107$ participants (65% females) between 23 and 75 years of age took part in the experiment. Mean age was 48.4 years ($SD = 16.9$). All participants gave written informed consent and were debriefed on the purpose and results of the study. The sample consisted of students from the University 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

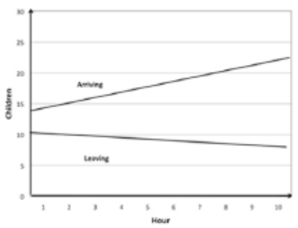
Concerning the communication of SF problems such as the accumulation of debts, or the accumulation of CO_2 in the atmosphere, we suggest that display formats used in media reports such as reports by the Intergovernmental Panel on Climate Change (IPCC) could be rendered more accessible by reducing the 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

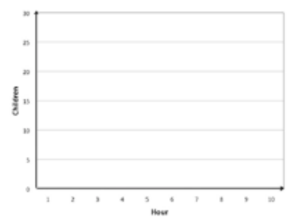
The present experiment demonstrated that people are better able to deal with SF systems if the problems are presented in a purely verbal format. This result suggests that both simulation-based learning and communication of SF problems could be rendered more effective by giving more weight to verbal information. 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.

Appendix A2

Manuscript 2: Making sense of dynamic systems: How our understanding of stocks and flows depends on a global perspective

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2015) 1–17

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12239

Making Sense of Dynamic Systems: How Our Understanding of Stocks and Flows Depends on a Global Perspective

Helen Fischer,[a] Cleotilde Gonzalez[b]

[a]*Department of Psychology, Heidelberg University*
[b]*Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University*

Received 28 July 2014; received in revised form 22 December 2014; accepted 29 December 2014

Abstract

Stocks and flows (SF) are building blocks of dynamic systems: Stocks change through inflows and outflows, such as our bank balance changing with withdrawals and deposits, or atmospheric CO_2 with absorptions and emissions. However, people make systematic errors when trying to infer the behavior of dynamic systems, termed SF failure, whose cognitive explanations are yet unknown. We argue that SF failure appears when people focus on specific system elements (local processing), rather than on the system structure and gestalt (global processing). Using a standard SF task (n = 148), SF failure decreased by (a) a global as opposed to local task format; (b) individual global as opposed to local processing styles; and (c) global as opposed to local perceptual priming. These results converge toward local processing as an explanation for SF failure. We discuss theoretical and practical implications on the connections between the scope of attention and understanding of dynamic systems.

Keywords: Dynamic systems; Global–local processing; Structure versus surface elements; Stock-flow failure; Dynamic decision making

1. Introduction

Many decisions that we make in our daily lives involve keeping a dynamic system under control. We aim at keeping our weight at a healthy stage by consuming the right amount of calories and exercising, or our bank accounts in a positive balance while controlling our expenses according to our incomes. Such stocks and flows (SF) structures

Correspondence should be sent to Helen Fischer, Department of Psychology, University of Heidelberg, Hauptstrasse 47-51, 69117 Heidelberg, Germany. E-mail: helen.fischer@psychologie.uni-heidelberg.de

comprise a stock (i.e., an accumulation) that is influenced by decisions made to increase (inflow) or to decrease the stock (outflow). As such, SF structures are the most basic building blocks of dynamic systems and they also are the source of dynamic complexity due to the over-time accumulation of flows into stocks (Cronin & Gonzalez, 2007; Sterman, 2000).

Research has shown that humans perform poorly in dynamic system tasks, even after extended amounts of practice, performance incentives, unlimited time, and full information (Diehl & Sterman, 1995; Fu & Gonzalez, 2006; Gonzalez, 2005; Martin, Gonzalez, & Lebiere, 2004; Paich & Sterman, 1993; Sterman, 1989, 1994). Furthermore, poor performance has also been shown in extremely simplified dynamic systems reduced to their fundamental elements—one stock, one inflow, and one outflow (e.g., Cronin & Gonzalez, 2007; Gonzalez & Wong, 2012; Sterman, 2002). These results led researchers to suggest a general difficulty in understanding dynamic systems, termed *Stock-Flow failure* (SF failure; Cronin, Gonzalez, & Sterman, 2009). How can we explain SF failure, given the ubiquity of dynamic systems in our environment?

A common mistake in judging SF systems, termed *correlation heuristic*, is the tendency to judge the stock as behaving similarly to its flows (Booth Sweeney & Sterman, 2000; Cronin & Gonzalez, 2007; Cronin et al., 2009; Gonzalez & Wong, 2012). An important example is used in recent studies showing the failure to understand the relationship between the CO_2 stock in the atmosphere and the inflow via anthropogenic CO_2 emissions and the outflow via natural CO_2 absorption (Dutt & Gonzalez, 2012; Guy, Kashima, Walker, & O'Neill, 2013; Moxnes & Saysel, 2009; Sterman, 2008; Sterman & Booth Sweeney, 2007). When participants were given a constant trend of CO_2 absorptions and a decreasing, but higher trend of emissions, they judged that the atmospheric CO_2 concentration would also decrease. This is clearly erroneous, because as long as emissions are higher than absorptions, the CO_2 stock will increase. This result together with examples from many other contexts suggests that following a correlation heuristic is a common mistake (Booth Sweeney & Sterman, 2000; Brunstein, Gonzalez, & Kanter, 2010; Cronin et al., 2009; Gonzalez & Wong, 2012). Although it is an important finding, the correlation heuristic remains a re-description of the typical behavior rather than a cognitive explanation for it.

In the next sections, we develop and test the following hypotheses:

1. The SF failure is related to the tendency to concentrate on the details of a system (local processing) rather than on the gestalt (global processing).
2. It is possible to procedurally prime participants to think globally versus locally, influencing the SF failure.
3. The task format may induce local or global views of a system, influencing the SF failure.

1.1. Global–local processing and understanding of dynamic systems

We suggest that the SF failure may be due to a tendency to concentrate on the details of a structure rather than on the gestalt; the local rather than the global processing initially investigated by Navon (1977). Navon presented participants with large (global) letters constituted by small (local) letters and asked participants whether a target letter matching either the global or the local letter was present on the screen. He found a global dominance effect, showing that participants' decisions were generally faster when the target matched the global letters than when it matched the local letters. Analogous to Navon's letters, dynamic systems can be seen as hierarchical: They consist of a set of constituent elements and an underlying relational structure. Just like the global character (e.g., H) in the Navon task cannot be inferred by looking at its local characters in isolation (e.g., Es or Ls), the behavior of a dynamic system cannot be inferred from looking at its constituent elements in isolation. An abstraction process is, therefore, needed from local-level representations of elements to global-level representations of structure to make inferences about the behavior of the *system* from information about its *elements*.

We suggest a basic structural and cognitive correspondence between dynamic systems and hierarchical figures consisting of local elements and a global configuration (see Fig. 1). Hierarchical figures such as the Kimchi–Palmer figures (Kimchi & Palmer, 1982) consist of global configurations made up of local elements. The local elements in SF systems (i.e., levels of inflow or outflow at particular points in time) may correspond to the local elements in hierarchical figures, while the global structure of the SF system (i.e., the relationships of inflows and outflows over time) may correspond to the global elements in hierarchical figures. Thus, the system's behavior can only be inferred from the interrelations between the elements and the system structure. For example, to predict atmospheric CO_2 concentration, one needs to see how CO_2 emissions and absorptions relate to each other over a period of time (i.e., structure) rather than seeing the exact levels of emissions or absorption at a specific point in time (i.e., constituent elements).

How does the processing of hierarchical stimuli apply to understanding the behavior of dynamic systems? People show inherent differences in their global and local processing

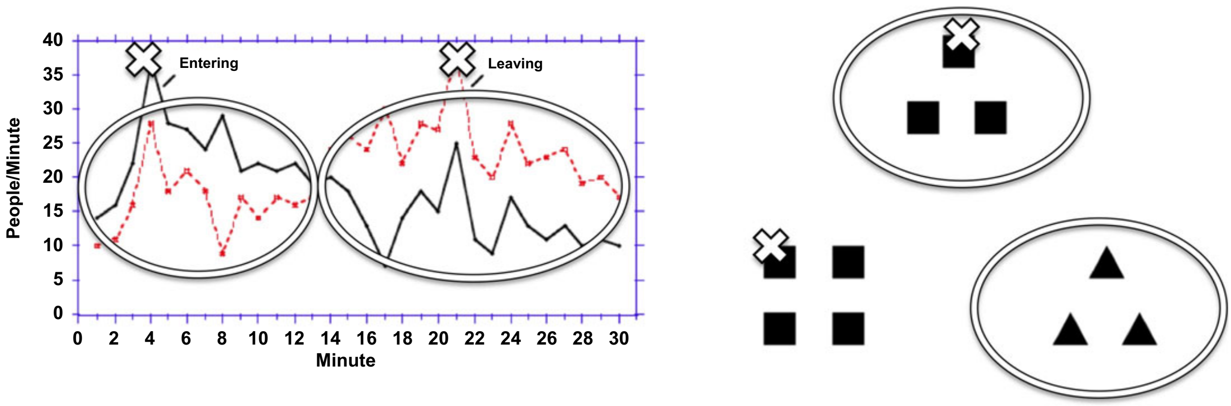


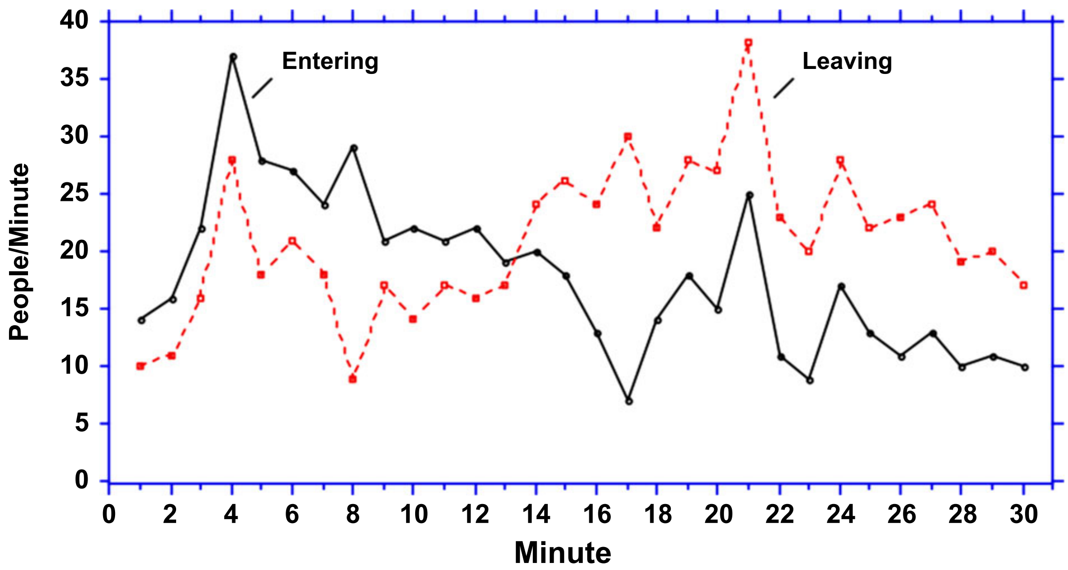
Fig. 1. Comparison of the proposed structural and cognitive parallels between hierarchical figures and dynamic systems. A tendency to focus on local elements (crosses) should be detrimental; a global focus on gestalts (circles) should be beneficial for inferring the overall system behavior.

styles (content-free ways of perceiving the environment; Tulving & Schacter, 1990). Thus, although research shows that many individuals may exhibit global dominance and a preference for global information processing (Kimchi, 1992; Navon, 1977), recent studies directly examine global–local processing styles as individual differences and suggest that global–local processing styles are inherently a strong characteristic of human information processing (Dale & Arnell, 2014). This research also shows that individual preferences for global or local information may influence performance in other tasks (Dale & Arnell, 2014). We expect that people who tend to focus on elements in hierarchical figures also tend to focus on elements in dynamic systems, whereas people who tend to focus on the global structure in hierarchical figures also tend to focus on the global structure of the system, thereby influencing the SF failure.

Furthermore, we expect that it would be possible to procedurally prime the perception of the gestalt of visual displays. Global procedural priming may enhance the activation of abstract objects in memory and induce creativity and broader thinking even in unrelated (content-wise) tasks (Friedman, Fishbach, Förster, & Werth, 2003). Research has shown that using a global perceptual priming manipulation (e.g., instructions to look at broad segments of a map) resulted in better creativity and novelty in a subsequent task, in contrast to local priming (e.g., instructions to look at narrow segments of the same map; Friedman et al., 2003). Most important, the priming and test tasks did not overlap in content, concluding that the effect of priming on the subsequent task was due to a correspondence in process rather than in content. This result agrees with recent findings in which structural similarity was crucial for reducing SF failure in contrast to surface similarity (Gonzalez & Wong, 2012). Thus, we expect that global–local perceptual priming with an unrelated visual task should influence the SF failure. Participants who are instructed to look at specific elements of a visual display should be more likely to focus on the system details, whereas participants who are instructed to look at the entire gestalt of a visual display should focus on the system's gestalt, that is, its structure. Consequently, SF failure will occur more after local than after global perceptual priming.

A related hypothesis is that the task format may highlight the local or global views of a system, influencing the SF failure. Fig. 2 displays a common task used in a number of studies, the department store (DS) task (Cronin & Gonzalez, 2007; Cronin et al., 2009; Sterman, 2000). Using this task, many researchers have studied whether people can infer the behavior of the system as a whole (number of people inside a store), given inflows and outflows in a graph (number of people entering and leaving). Questions 1 and 2 test whether participants can read the flows correctly. Questions 3 and 4 (called "SF questions" henceforth) test whether participants can infer the behavior of the stock. Solution rates for the SF questions are commonly below 50% (Brunstein, Gonzalez, & Kanter, 2010; Cronin et al., 2009; Gonzalez & Wong, 2012), irrespective of alternative graphical representations (Cronin & Gonzalez, 2007; Cronin et al., 2009); irrespective of the domain and the participants' experience (Brunstein et al., 2010); irrespective of the patterns of flows (Cronin et al., 2009; Gonzalez & Wong, 2012); and even when motivation and learning are induced (Cronin et al., 2009). Using correlational thinking, participants

The graph below shows the number of people *entering* and *leaving* a department store over a 30-minute period.



Please answer the following questions.

Check the box if the answer cannot be determined from the information provided.

1. During which minute did the most people enter the store?

Minute ________ ❑ Can't be determined

2. During which minute did the most people leave the store?

Minute ________ ❑ Can't be determined

3. During which minute were the most people in the store?

Minute ________ ❑ Can't be determined

4. During which minute were the fewest people in the store?

Minute ________ ❑ Can't be determined

Fig. 2. Department store (DS) task as used in Sterman (2000) and in some of the experiments in Cronin et al. (2009).

tended to judge that the stock is highest at the point of the highest inflow or the point of the greatest difference between inflow and outflow (Cronin & Gonzalez, 2007; Cronin et al., 2009).

We suggest, however, that the temptation to think locally and use a correlation heuristic may have been induced by the question format. These questions ask about *one specific* minute during which the stock was highest or lowest, instead of highlighting the gestalt of the trends and the relation between the flows. We suggest that if questions are designed to emphasize the global gist of the trends, and to highlight the relative salience of the relations between the flows, attention should shift toward how the parts relate to the whole (Kimchi, 1992; Macrae & Lewis, 2002), decreasing the SF failure.

2. Methods

2.1. Participants

A total of 148 participants (80 female, 67 male, 1 unknown) with a mean age of 34.9 years (*SD* = 12, range = 18–64) took part via the Internet. Participants were recruited over Mechanical Turk (MTurk), an online participant recruitment site. Participants were compensated after completion of the study through their MTurk account with a flat \$0.75 Participants were restricted to US IP addresses and had completed at least high school; 33% had a 4-year college degree in a range of different fields, the largest groups being Business (10%), Psychology (7%), and English (3%).

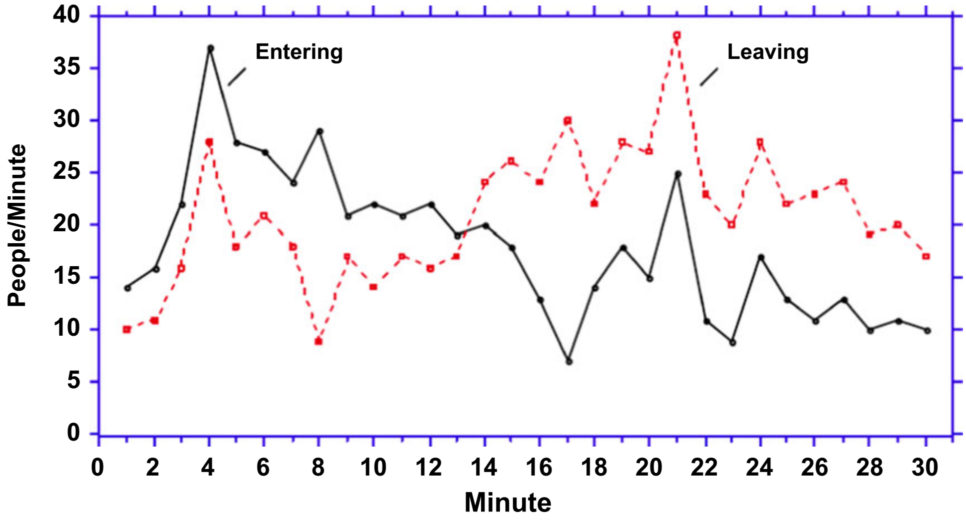
2.2. Materials

The main dependent variable in all hypotheses was the SF accuracy in the original DS task.

The question format was manipulated in the DS task. The original, *local* DS task format (Fig. 2) was compared to a modified, *global* format (Fig. 3). Please note that in both formats, calculations are unnecessary: One needs to only understand that the number of people inside the store rises as long as the number of people entering is greater than the number of people leaving. Thus, one can directly infer from the graph that the most people are inside at minute 13 (see Cronin et al., 2009). Both formats used the exact same introduction and graph. We removed the option to mark “can’t be determined” from both formats, as in past research this is often the second most common mistake after the correlation heuristic (Cronin et al., 2009). Instead, 7-point Likert scales assessed subjective confidence in each answer: How confident are you in your answer? 0 = *Not confident at all* and 7 = *very confident*.

To measure individual global–local processing styles, we used the Kimchi–Palmer-Figures task (Kimchi & Palmer, 1982) that consists of triangles and squares made up of smaller triangles and squares. For each of 16 trials, participants indicated whether a target figure

The graph below shows the number of people *entering* and *leaving* a department store over a 30-minute period.



Please answer the following questions.

1. How are the people entering related to the people leaving the store between time periods 1 to 14?
    - More people entering than leaving
    - More people leaving than entering
    - Same amount of people entering and leaving

2. How are the people entering related to the people leaving the store between time periods 14 to 30?
    - More people entering than leaving
    - More people leaving than entering
    - Same amount of people entering and leaving

3. How would you best describe the accumulation of the number of people in the store between time periods 1 to 14?
    - Increasing
    - Decreasing
    - Stable

4. How would you best describe the accumulation of the number of people in the store between time periods 14 to 30?
    - Increasing
    - Decreasing
    - Stable

Fig. 3. Department store task with questions in the global format.

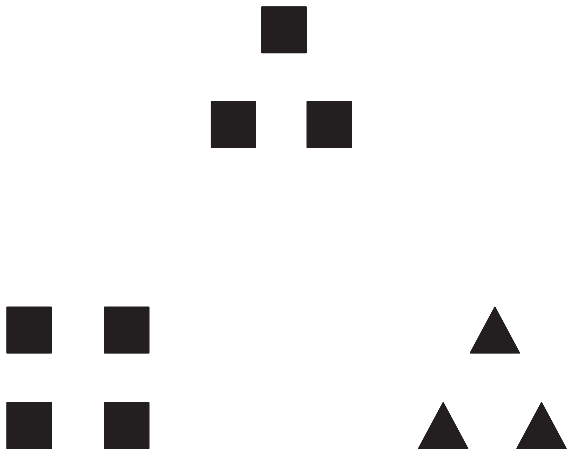


Fig. 4. Sample trial from the Kimchi–Palmer-Figures task. The target figure is displayed above; the bottom-left figure represents the local match, and the bottom-right figure represents the global match.

(e.g., a global triangle made of local squares) was more similar to a sample figure that matched its global or its local form (Fig. 4). Display of the figures was counterbalanced with respect to the global (local) match appearing on the left (right). Mean ratings were conducted for each participant, ranging from 0 (completely local processing style) to 1 (completely global processing style).

To procedurally prime participants (global vs. local vs. control), a maps task similar to Friedman et al. (2003) was used.[1] A crucial difference is that the instruction used by Friedman et al. was designed to manipulate attention scope, whereas the instruction in the present study was designed to manipulate a focus on details versus on gestalt of the display. For each of seven trials, a state map was presented on the screen (Fig. 5). The focus of attention was varied via different instructions: The global group was instructed to look at the respective state in its entirety and to describe the overall shape of the whole state. The local group was instructed to attend to the respective capital only and to describe the exact location of that specific city. The control group was instructed to think about an item that characterizes the state and to name that item. The control instructions were chosen to not influence pre-existing processing styles. For all three conditions, the respective descriptions (overall shape vs. specific location vs. item) were given while the map was still presented on the screen. Upon pressing enter, the next map was displayed.

2.3. Procedure

Participants were told that they were going to take part in two tasks, one about visual perception and one about decision making. They were told that the study would take approx. 10 minutes and that it needed to be completed in one sitting. Participants first completed one of three randomly assigned between-subjects perceptual priming treatments with the maps task. Second, participants answered the within-subjects DS task

Fig. 5. Sample map used for procedural priming. Participants attended to the map in its entirety (global perceptual priming), to its respective capital (local perceptual priming), or to neither: They thought about an item that characterizes the respective state (control group).

in the original (local) and the modified (global) formats. To control for potential order effects, global-formatted and local-formatted DS task were presented in random order. Third, individual global-local processing style was measured with the Kimchi–Palmer-Figures task.

3. Results

Control analyses showed that first, solution rates in the original DS task did not differ when the global-formatted questions were answered first (M = .19, SD = .32) compared to when the local-formatted questions were answered first (M = .18, SD = .30), $t(146)$ = .24, p = .81.). Second, mean perceptual processing styles did not differ between global (M = .63, SD = .44) or local perceptual priming (M = .68, SD = .42), or the control group (M = .67, SD = .43), $F(2, 146)$ = .21, p = .81. To increase readability, the following analyses test the specific hypotheses independently.

3.1. How does a global question format influence SF failure?

Table 1 shows the proportion of responses in the global and local question formats in the control group (n = 43). The accuracy of questions 1 and 2 was not different for the local and global question format; thus, the new global format does not offer an advantage to interpreting the flows. As expected, however, people's ability to infer the stock was influenced dramatically by the question format. In the local format, only 19% and 16% of participants were able to infer the system's overall behavior, while in the global format, 65% and 84% of participants (SF questions 3 and 4, respectively) were able to infer the correct system behavior.

Table 1
Percentage of correct responses in the local and global questions in the control group (n = 43)

	Question Format		
DS Question	Original (local) %	Modified (global) %	McNemar
Q1	95	91	n.s.
Q2	88	91	n.s.
SF Q3	19	65	40.12, $p < .001$
SF Q4	16	84	78.22, $p < .001$

Moreover, in the global question format, confidence ratings were connected to SF accuracy, $r(147) = .39$, $p < .001$, whereas in the local question format, confidence ratings were not, $r(147) = .001$, $p = .49$. This dissociation indicates that the global question format helped participants gain insight into the structure of the task, leading to subjective confidence ratings that were diagnostic of actual SF accuracy.

3.2. Does global–local procedural priming influence SF failure?

Aggregated over all participants, priming did not seem to influence SF accuracies $F(2, 144) = 0.1$, $p = .90$. For those participants who answered the original DS task first (i.e., immediately after the priming manipulation, $n = 71$), priming did influence SF accuracy (Table 2). Specifically, after global priming ($M = .27$, $SD = .38$), participants achieved higher SF accuracies compared to after local priming ($M = .12$, $SD = .22$), $t(40.7) = 1.7$, $p = .048$, $d = .5$. There were no significant differences between the global or local priming group and the control group, both $p > .05$. That is, in line with our hypothesis, global perceptual priming significantly increased SF accuracies in the DS task compared to local perceptual priming. Unexpectedly, however, global priming, albeit in the expected direction, did not significantly improve SF accuracies compared to the control group.

3.3. Is global–local processing style related to SF failure?

Internal reliability of the global–local processing scale (Kimchi–Palmer-Figures task) in the present study was high ($\alpha = .98$). To test the influence of processing style irrespective of the influence of procedural priming, participants in the control group were classified as *global processors* ($n = 20$) or *local processors* ($n = 23$) based on the median of the distribution of global–local scores ($M = .66$, $SD = .43$, *median* = .93). In line with our reasoning, SF accuracies were higher for global ($M = .24$, $SD = .36$) than local processors ($M = .08$, $SD = .26$) in the original local question format, $t(40.9) = -1.67$, $p = .05$, Cohen's $d = .52$. Interestingly, in the global question format, global processors ($M = .80$, $SD = .35$) achieved only marginally higher solution rates than local processors, ($M = .67$, $SD = .42$), $t(41) = -1.13$, $p = .13$, Cohen's $d = .34$.

Moreover, for global processors, SF accuracies improved by a factor of only 3.3 from the local to the global format, whereas for local processors, SF accuracies improved by a

Table 2
Mean accuracy scores of the SF questions in the original DS task as a function of perceptual priming condition (local vs. control vs. global)

Priming Condition			*t*-tests		
Local (n = 24) *M*(*SD*)	Control (n = 21) *M*(*SD*)	Global (n = 26) *M*(*SD*)	Local–global	Local–control	Global–control
.12 (.22)	.17 (.32)	.27 (.38)	$t(40.7) = 1.7$, $p = .048$, $d = .53$	$t(43) = -.50$, $p = .31$	$t(45) = .97$, $p = .17$

factor of 8.4 (leading to a smaller difference between global and local processors in the global task). These results indicate that, first, global processors are able to take advantage of their abilities in the local question format while, second, local processors benefitted disproportionately from the global format.

3.4. Is global–local processing connected to correlational thinking?

Analogously to previous research (Cronin & Gonzalez, 2007; Cronin et al., 2009), answers in the local format were coded as correlation heuristic use when participants answered minute 4 or 8 in Question 3 (maximum inflow or maximum netflow) or minute 17 or 22 in Question 4 (maximum outflow or minimum netflow). Also, similar to past studies where participants were asked to draw the stock trend (e.g., Dutt & Gonzalez, 2012), answers were coded as correlational thinking when imitating the trend of the flows (answer “Decreasing” in Question 3 to imitate the inflow and “Stable” in Question 4 to imitating the outflow). Thus, if participants use the correlation heuristic they would answer erroneously in both SF Question 3 and 4. There was a significant reduction in correlation heuristic use from the local to the global format for both SF Question 3, McNemar = 12.04, $p < .001$, and SF Question 4, McNemar = 12.5, $p < .001$ (Table 3).

Finally, to investigate the connection between global–local processing styles and the correlation heuristic irrespective of procedural priming, again the control group was used. In line with our reasoning, global–local processing was negatively connected to correlational thinking in the original DS task, $r(42) = -.27$, $p = .03$. In the global-formatted task, global–local processing was not significantly, albeit also negatively, connected to correlational thinking $r(42) = -.17$, $p = .13$.

4. Discussion

This research provides cognitive explanations for a robust and consistent error in processing dynamic systems, the SF failure. First, we find that the question format may induce local or global processing, and thus influence the SF failure. Second, the SF failure is related to individual local rather than global processing styles, and the more local participants’ processing style, the more they tend to use the correlation heuristic. Third,

Table 3
Most common answers of the control group (n = 43) to the SF questions 3 and 4 in the local versus global DS task

Answer	SF Question 3 (%)	SF Question 4 (%)
Local format		
Max entering t = 4	*12*	0
Max leaving t = 21	5	*9*
Max net inflow t = 8	*47*	2
Max net outflow t = 17	7	*44*
Most in store t = 13	**19**	7
Fewest in store t = 30	0	**16**
Global format		
Decreasing	*16*	**84**
Increasing	**65**	2
Stable	18	*14*

Note. Bold values specify the correct answers; italic values specify correlation heuristic use.

by procedurally priming participants to process information globally rather than locally, they are able to decrease the SF failure.

SF failure in the local question format of the DS task was dramatically higher than in the global question format. In the local question format, only a minority of participants was able to infer the stock's behavior. These proportions are much in agreement, although lower than in past research (Cronin et al., 2009; Sterman, 2008). Most of the past research in the DS task was conducted in universities with a high level of mathematics education; the lower proportions may be due to the more general population used in the current study. When interrelations between system elements were highlighted in the global format (Fig. 3), instead of specifics of system elements in the original format (Fig. 2), a large proportion of participants were able to correctly infer the overall system behavior. This result suggests that the previously found SF failure may at least in part be attributed to the way questions about the system were asked, or specifically how the local question format can direct participants' attention to isolated system elements rather than the system structure.

An alternative explanation for the better understanding of the SF system in the modified format is that the global questions provided more or more relevant information than the local questions by referring to "time periods 1 to 14 (14 to 30)" or by referring to the relation between the flows. However, first, previous research found that the order of the questions about the flows (Q1 & Q2) versus the stock (Q3 & Q4) does not have an effect on SF accuracy (Cronin et al., 2009), and moreover, this additional information exerted no effect on the questions regarding the flows, only on the questions regarding the stock. Therefore, it is more likely that the difference in accuracy is due to a fundamental need for global structural information in inferring the behavior of the system as a whole. This conclusion is supported by the results regarding the global–local processing styles and global–local perceptual priming.

Global processors had higher SF accuracies in the original DS task, compared to local processors. That is, the individual's processing style was directly related to the SF failure. For the global question format, the global processors were only slightly better than the local processors, and local processors benefitted disproportionately from the global question format. These results suggest that first, as expected, global processing styles are beneficial for inferring the overall behavior of the dynamic system. Second, global processing styles are less beneficial for the understanding of transparent systems with a salient system structure; they are especially beneficial for the understanding of intransparent systems that do not easily reveal their structure.

Similarly, in the original format highlighting isolated system elements, participants not only used the correlation heuristic more than in the global format highlighting the system structure—they also tended to use the correlation heuristic more, the more local (or less global) their processing style. That is, the more people see the elements of dynamic systems as structurally related, the less they tend to believe that the output of the system should simply be linearly correlated with its isolated input (such as erroneously believing that the moment of biggest stock should coincide with the moment of biggest inflow). In the global format highlighting the system structure, however, no significant relationship between individual processing styles and correlation heuristic use existed, implying, again, that participants hardly profit from global processing styles in the case of structurally transparent systems.

It is important to note that correlational thinking may be a product of our mind to create an economic solution to dynamic systems. Correlational thinking can be a successful strategy in simple, linear systems (such as water boiling faster when we turn up the heat; assuming all else is constant); this kind of thinking fails, however, in more complex, dynamic systems (such as expecting that cutting the deficit would directly cut the debt; Sterman, 2008). Thus, in most dynamic contexts, the only simple and correct solution would be to focus on the system's gestalt: its SF structure.

Perceptually priming participants to look at the gestalt of a visual display (global priming) increased their ability to infer the system's behavior in an unrelated SF task compared to priming participants to look at the details of the same display (local priming). Since the perceptual priming and the subsequent SF task did not overlap in content, we conclude that processes were primed: a global processing of gestalts versus local processing of elements. However, our results also show that such an effect of procedural priming may be short-lived and present only in the task immediately following the perceptual priming task. That is, at least with our procedural priming task, no longer lasting effects on participants' understanding of dynamic systems may be achieved.

The present results bear theoretical implications for the connections between people's scope of attention and their understanding of dynamic systems. It was argued before that broadening or narrowing the scope of attention to external perceptual input is achieved by the same mechanism as broadening or narrowing the scope of attention to internal conceptual representations (Friedman et al., 2003). We showed that directing attention toward gestalts or elements of visual displays can affect whether people's thinking is subsequently directed toward gestalts or elements of systems. That is, we provide first evidence that the link between attention to external and internal stimuli may not only exist for the mere scope of attention (i.e., whether we focus narrowly or broadly) but also the level of attention in hierarchical constructs (i.e., whether we focus on gestalts or elements).

In sum, these findings are in line with our basic assumption of a correspondence between the way people process hierarchical figures and the way they process dynamic systems: Global processing enables one to perceive a system's elements as structurally related and to infer the overall behavior of the system from the behavior of its parts. People who tend to process information locally by focusing on specific details fail to understand the system's behavior, whereas people who tend to process information globally by looking at overarching structures tend to understand the system's behavior.

In order to enhance people's ability to deal with dynamic systems, these results offer a range of solutions. One could enable people to adopt higher order perspectives by teaching strategies of abstraction and pattern recognition. Given our result that perceptual priming with a purely visual task affects people's ability to infer the behavior of a dynamic system, it also seems necessary to ensure that the tasks performed immediately before (or even during) interacting with a dynamic system do not induce a focus on details and elements, but on patterns and structure—even if those tasks are completely unrelated content-wise. Furthermore, we found that highlighting the relations between a system's elements verbally improves people's understanding of the system. It might also prove helpful to highlight relations between the elements visually by grouping the constituent elements of dynamic systems in a way that implies global structure. This way, one could induce attention on the structure of the system instead of its elements. In other words, one could induce looking at what is signified instead of on the signs.

It remains an important open question, in how far the connections between a global–local focus in hierarchical figures and dynamic systems still hold in dynamic, interactive environments since hierarchical figures do not convey information about iterative processes. Although we do expect a global focus on relations between elements to still be beneficial for inferring the system behavior, this is for future research to decide.

Albeit SF accuracy for global processors (M = .24) was three times as high as for local processors (M = .08) in the local DS task, solution rates were still quite low in absolute numbers. It might, therefore, prove fruitful to investigate how even many people with a tendency to process information globally can be led astray when the task format highlights isolated system elements. Using eye tracking, for example, one might reveal how participants' perceptual focus changes as a function of task format and time on task in such a way that pre-existing processing styles change while interacting with a global- versus local-formatted system. By combining such an approach with different and other tasks than the ones we used, one could also investigate how some people might even be able to adapt their processing styles to align with specific task requirements. Assessing reaction times might also be a valuable approach to test for alternative heuristics. For example, in the global-formatted DS task, although correlation heuristic use was significantly reduced, some people might still use an alternative, heuristic approach by associating the respective line on top with an answer option due to semantic similarity, such as "entering" with "increasing" and "leaving" with "decreasing." Our understanding of the cognitive strategies used could thus be increased further by reaction times because a bimodal response time distribution is expected if there is a subset of participants using a heuristic approach to solve even global-formatted systems.

We introduced global–local processing as a fundamental cognitive explanation of how people deal with dynamic complexity and why so many fail with even its most simplistic form: SF systems containing one inflow, one outflow, and one stock. Our results converge on the conclusion that less successful participants approach SF problems in a local manner by focusing on the system elements, whereas successful participants approach them in a global manner by focusing on the system structure, its emerging gestalt. As stated at the beginning, SF systems themselves can be seen as building blocks: as the elements of more complex systems constituted by several SF subsystems. It seems reasonable to speculate that a global perspective should be even more beneficial for the understanding of more complex systems than for the basic system we used. Systems containing many interacting subsystems can hardly be regulated using analytical strategies both because of limited cognitive capacities and because real-life information is mostly fuzzy. Global processing may enable us to imagine surrounding systems in their most economic form and to recognize basic structural regularities in a dynamic world.

Acknowledgments

This research was supported the National Science Foundation Award number 1154012 to Cleotilde Gonzalez. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. We thank Hau-yu Wong, Dynamic Decision Making Laboratory, for her editorial and insightful comments.

Note

1. We performed a pilot study to assess the efficacy of the procedural priming task with online participants (n = 204). The exact same perceptual priming via the maps task was used; to assess understanding of the dynamic system, we used a task that was structurally equivalent to the local DS task with a different context (subscribers and unsubscribers of a magazine). We found that SF reasoning marginally improved after global (M = .09, SD = .25) compared to local priming (M = .04, SD = .15), t = 1.4. p = .08.

References

Booth Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249–286. doi: 10.1002/sdr.198.

Brunstein, A., Gonzalez, C., & Kanter, S. (2010). Effects of domain experience in the stock–flow failure. *System Dynamics Review*, *26*(4), 347–354. doi: 10.1002/sdr.448.

Cronin, M. A., & Gonzalez, C. (2007). Understanding the building blocks of system dynamics. *System Dynamics Review*, *23*(1), 1–17. doi: 10.1002/sdr.356.

Cronin, M. A., 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*(1), 116–130. doi: 10.1016/j.obhdp.2008.03.003.

Dale, G., & Arnell, K. M. (2014). Lost in the forest, stuck in the trees: Dispositional global/local bias is resistant to exposure to high and low spatial frequencies. *PLoS ONE*, *9*(7), e98625. doi: 10.1371/journal.pone.0098625.

Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, *62*(2), 198–215. doi: 10.1006/obhd.1995.1043.

Dutt, V., & Gonzalez, C. (2012). Human control of climate change. *Climatic Change*, *111*(3–4), 497–518. doi: 10.1007/s10584-011-0202-x.

Friedman, R., Fishbach, A., Förster, J., & Werth, L. (2003). Attentional priming effects on creativity. *Creativity Research Journal*, *15*(2-3), 277–286. doi: 10.1080/10400419.2003.9651420

Fu, W., & Gonzalez, C. (2006). Learning to control dynamic systems: Information utilization and future planning. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 244–249). Mahwah, NJ: Lawrence Erlbaum Associates.

Gonzalez, C. (2005). Decision support for real-time dynamic decision making tasks. *Organizational Behavior and Human Decision Processes*, *96*(2), 142–154. doi: 10.1016/j.obhdp.2004.11.002.

Gonzalez, C., & Wong, H. (2012). Understanding stocks and flows through analogy. *System Dynamics Review*, *28*(1), 3–27. doi: 10.1002/sdr.470.

Guy, S., Kashima, Y., Walker, I., & O'Neill, S. (2013). Comparing the atmosphere to a bathtub: Effectiveness of analogy for reasoning about accumulation. *Climatic Change*, *121*(4), 579–594. doi: 10.1007/s10584-013-0949-3.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, *112*(1), 24–38. doi: 10.1037/0033-2909.112.1.24.

Kimchi, R., & Palmer, S. E. (1982). Form and texture in hierarchically constructed patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(4), 521–535. doi: 10.1037/0096-1523.8.4.521.

Macrae, N., & Lewis, H. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science*, *13*(2), 194–196. doi: 10.1111/1467-9280.00436.

Martin, M. K., Gonzalez, C., & Lebiere, C. (2004). Learning to make decisions in dynamic environments: ACT-R plays the beer game. In M. C. Lovett, C. D. Schunn, C. Lebiere, & P. Munro (Eds.), *Proceedings of the Sixth International Conference on Cognitive Modeling*. Vol. *420* (pp. 178–183). Pittsburgh, PA: Lawrence Erlbaum Associates Publishers.

Moxnes, E., & Saysel, A. K. (2009). Misperceptions of global climate change: Information policies. *Climatic Change*, *93*(1–2), 15–37. doi: 10.1007/s10584-008-9465-2.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. doi: 10.1016/0010-0285(77)90012-3.

Paich, M., & Sterman, J. D. (1993). Boom, bust, and failures to learn in experimental markets. *Management Science*, *39*(12), 1439–1458. doi: 10.1287/mnsc.39.12.1439.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, *35*(3), 321–339. doi: 10.1287/mnsc.35.3.321.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, *10*(2–3), 291–330. doi: 10.1002/sdr.4260100214.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. New York: McGraw-Hill.

Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review*, *18*(4), 501–531. doi: 10.1002/sdr.261.

Sterman, J. D. (2008). Risk communication on climate change: Mental models and mass balance. *Science*, *322*(5901), 532–533. doi: 10.1126/science.1162574.

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*(3–4), 213–238. doi: 10.1007/s10584-006-9107-5.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*(4940), 301–306. doi: 10.1126/science.2296719.

Appendix A3

Manuscript 3: When high working memory capacity and using more training information is and is not beneficial for predicting non-linear processes

When high working memory capacity and using more training information is and is not beneficial for predicting non-linear processes

Helen Fischer & Daniel V. Holt

Department of Psychology, Heidelberg University, Germany

.

Helen Fischer (corresponding author), University of Heidelberg, Department of Psychology, Hauptstrasse 47-51, 69117 Heidelberg, Germany. Email: helen.fischer@psychologie.uni-heidelberg.de.

Predicting the development of dynamic processes is vital in many areas of life. Previous findings are inconclusive as to whether higher working memory capacity (WMC) is always associated with using more accurate prediction strategies, or whether higher WMC can also lead to using overly complex prediction strategies that do not improve accuracy. In this study, participants predicted a range of systematically varied non-linear processes based on exponential functions where prediction accuracy could, or could not be enhanced using well-calibrated rules. Results indicate that higher WMC participants seem to rely more on well-calibrated rule-based strategies, leading to more accurate predictions for processes with highly non-linear trajectories in the prediction region. Predictions of lower WMC participants, in contrast, point towards an increased use of simple exemplar-based predictions strategies, which perform just as well as more complex strategies when the prediction region is approximately linear. These results imply that with respect to predicting dynamic processes, working memory capacity limits are not generally a strength or a weakness, but that this depends on the process to be predicted.

Keywords: prediction; working memory capacity; function-learning; rule-based versus exemplar-based; non-linear dynamic processes

Making predictions about the development of dynamic processes is important in many areas of life, be it judging when the food will be done while cooking, monitoring a patient in critical care, or controlling an industrial power plant. However, research as early as Wagenaar and Sagaria (1975) has shown that the accuracy of predictions varies widely, particularly when processes do not follow a simple linear pattern. In this article, we focus on the contribution of working memory capacity (WMC) for explaining this variation in predictions, particularly considering the interaction of WMC and the type of process to be predicted. Intuitively, higher memory capacity should always improve prediction accuracy since people need to consider information about the past of a process in order to forecast its future (cf. Mackinnon & Wearing, 1991). And indeed, research has shown that higher WMC is related to higher prediction accuracy for continuous processes (function-learning) and categorization tasks (Bröder et al., 2010; Lewandowsky et al., 2012; McDaniel, Cahill, Robbins, & Wiener, 2014). One common explanation for this observation is that higher WMC is associated with an improved ability to actively maintain and manipulate past information, which is needed to calibrate cognitive prediction algorithms to the learning data, and to abstract systematic regularities. These regularities, or “rules” can then be used for prediction (McDaniel et al., 2014). In this line of reasoning, WM capacity limits are a weakness that lead to less accurate predictions.

However, numerous findings following the “ecological rationality” approach (e.g., Marewski, Gaissmaier & Gigerenzer, 2010) show that in a surprising number of situations simple, information-frugal heuristics – which rely on minimal information and are not working memory intensive – perform just as well as complex rules (Marewski & Schooler, 2011). Similar arguments have been made for the limited capacity of working memory itself (Cowan, 2010), arguing that rather than being a compromise between processing capacity and metabolic efficiency, a limited WMC may confer genuine advantages. One reason is that because higher WMC endows one with the ability to execute mental algorithms on-line, higher WMC participants are more likely to employ complex strategies, even when simpler solutions are available (for a review, Wiley & Jarosz, 2012). In this line of reasoning, capacity limits are a strength that may lead to more elegant solution strategies. What may unite the conflicting views on the necessity of higher WMC for predictions is that higher WMC may be beneficial for predicting processes that require the abstraction of complex rules, but not for processes that fit simple, information-frugal prediction strategies.

To investigate this question, we used a learning paradigm with different types of exponential processes, where participants where first trained with the beginning of a process and then predicted how the process would continue. Information about the process and participants' predictions was given in numerical form trial-by-trial. No descriptive summary information, for example in the form of graphs, was available, so that participants needed to rely on WM to make predictions. We used variations of exponential processes for three reasons. First, they are practically relevant, as many processes in real life show exponential dynamics: The spreading of diseases, tipping points in ecological systems, and population growth can all have dramatic consequences if not anticipated early on. Second, empirical research has shown that people find them difficult to predict in general (Wagenaar & Sagaria, 1975), but little is known about the effects of individual differences on prediction accuracy. Third, choosing a positive vs. negative exponent allows to manipulate whether the process is asymptotic (increasingly linear), or accelerating (increasingly non-linear) that is, whether the simplest possible prediction strategy (linearity) performs optimal—or not.

Prediction strategies can be summarized in two broad classes: rule-based and exemplar-based. Rule-based models assume that participants abstract a global rule (e.g., exponential versus quadratic) describing the ensemble of information on the to-be predicted process (McDaniel & Busemeyer, 2005). To do so, participants use the feedback provided to update the parameters of their rule in order to calibrate their prediction algorithms to the training data (Koh & Meyer, 1991). Exemplar-based models, in contrast, assume that participants store single exemplars of cue-criterion mappings in memory (Nosofsky, 1988), and that only the most similar training exemplars are retrieved for extrapolation. The *extrapolation association model* (EXAM; DeLosh et al.,1997) and the *population of linear experts* (POLE; Kalish, Lewandowsky & Kruschke, 2004) assume that participants learn associations between x- and y-values (EXAM), or between x-values and a matching linear function, an expert (POLE). Participants extrapolate linearly through the two most similar x- and associated y-values (EXAM), or using the most similar x-value and its expert (POLE). Rule-and exemplar based strategies differ with respect to how much training information they use, and how well they use it. Rule-based strategies are more information-intensive in that a substantial amount of training information is used to induce and calibrate the rule (even if the resulting rule itself is comparatively simple and represents an economic form of information representation). Exemplar-based strategies, in contrast, are information-frugal in that only the most similar training exemplars are used for making a prediction. And only rule-based, but not exemplar-based, strategies imply that participants calibrate their prediction algorithms to the training data.

If exemplar-based strategies use less training information and are less calibrated: How can they *ever* make as accurate predictions as complex rules? In asymptotic processes, even if one abstracted the correct function rule, this would not pay off in terms of prediction accuracy because complex rule-based predictions and simple linear predictions based on the exemplars most similar to extrapolation overlap in these cases (Fig. 1). Moreover, rule-induction tends to be error-prone because several problem steps need to be performed on-line and in WM (Beilock & DeCaro, 2007), whereas the linear predictions required for EXAM or POLE are typically performed with near optimal accuracy (Busemeyer, Byun, Delosh & McDaniel, 1997). Abstracting information-intensive rules hence cannot outperform simple linear strategies for predicting the later parts of asymptotic processes, but may even impair predictions. In increasingly non-linear processes such as the accelerating functions shown in Figure 1, in contrast, rule-based strategies that capture the process' trend should clearly be more accurate because simple linear predictions would lead to a dramatic underestimation of the trajectory.

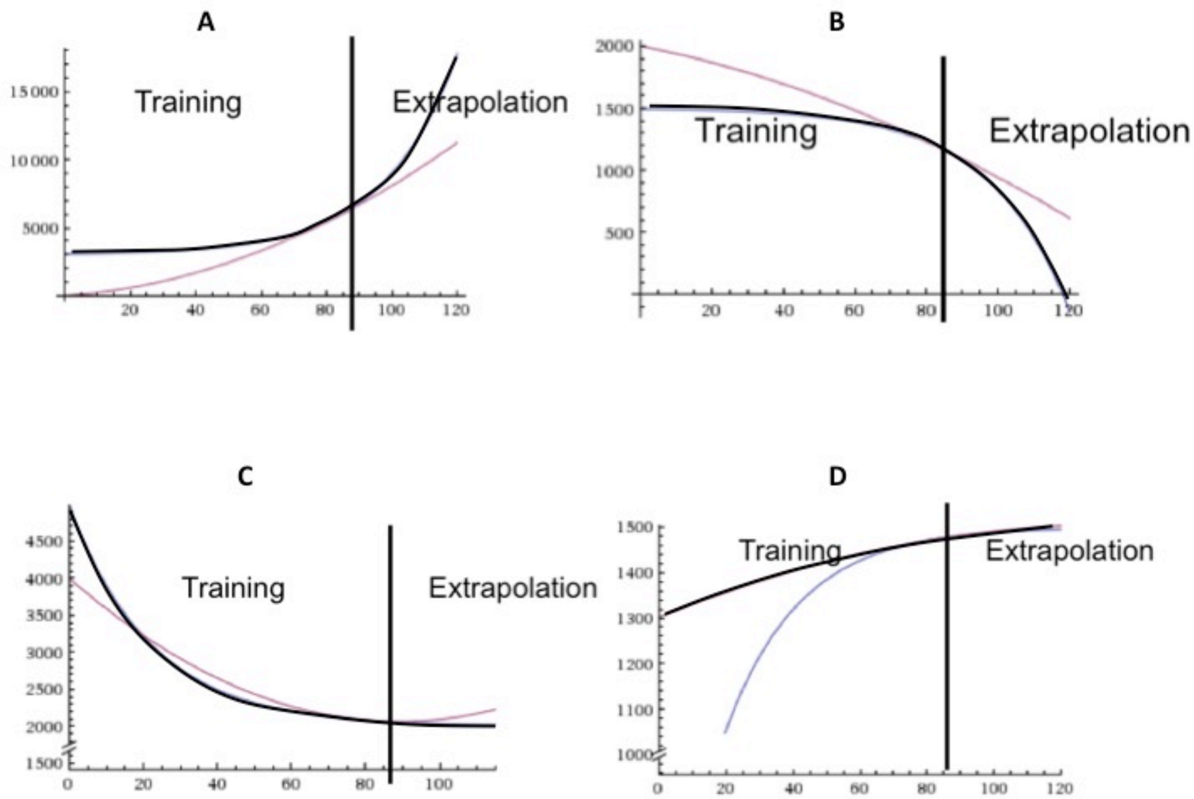


Figure 1. Design of the experiment and the accelerating increasing (A), accelerating decreasing (B), asymptotic decreasing (C) and asymptotic increasing (D) processes. Participants learned about the processes from time points 0 to 85 (training), and predicted the

processes from time points 90 to 115. Participants predicted exponential processes (black), and their respective quadratic twins (grey) that run through the same two training exemplars closest to extrapolation.

It seems plausible that higher WMC should lead to more accurate predictions in processes that benefit from the induction of information-intensive rules, that is, accelerating processes in the present experiment. This is because rule-induction requires one to memorize cue and criterion values, estimate their differences, and update the rule accordingly. These processes arguably encompass storage and transformation of information—key facets of working memory (Oberauer, Süß, Schulze, Wilhelm & Wittmann, 2000). In categorization and multiple-cue judgments tasks, WMC indeed predicted performance in tasks requiring rule-based strategies (Hoffmann, von Helversen & Rieskamp, 2014), and the involvement of WMC is particularly high when more cues need to be considered, or more complex rules need to be abstracted (Juslin, Karlsson & Olson, 2008; Mata et al., 2012). WMC is also associated to rule accuracy in a prototypical rule-induction task—Raven's Progressive Matrices—, particularly for items requiring complex rules (Little, Lewandowsky & Craig, 2014; Wiley et al., 2011). Despite these compelling connections, there is only one study directly assessing how WMC affects continuous predictions. McDaniel, Cahill, Robbins and Wiener (2014) suspected that greater WMC allows participants to actively maintain a range of cue-criterion values, and to concurrently compare them over trials to abstract a functional rule. Higher WMC participants were indeed more likely to use rule- as opposed to exemplar-based strategies, and achieved higher prediction accuracy. In this study, however, only one type of process was used (V-shaped) that was particularly chosen to demand rule-based strategies.

In cases where simple linear prediction models perform optimal—asymptotic processes in the present case—, higher WMC might not pay off in terms of better predictions. Exemplar-based predictions rely on verbatim recall of only the most similar training exemplars, and therefore require hardly any WM resources. Hoffmann, von Helversen and Rieskamp (2014), for example, found that WMC is not related to performance in exemplar-based judgment tasks at all. It is unclear, however, whether higher WMC participants will use simple exemplar-based strategy in the first place, or whether they will try to apply rules even if this strategy doe not pay off. Some studies found higher cognitive capacity to foster adaptive strategy-use (Bröder, 2003; Lewandowsky, Yang Newel & Kalish, 2012), so that one possibility is that in the case of asymptotic processes, higher WMC participants would perform just as if they had lower WMC, and use the simpler exemplar-based strategies. Other studies, however, found higher WMC to be detrimental for performance in tasks where simple strategies perform optimal. In Luchins' classic water jug problem, a complex solution strategy is needed for the first trials, but a simpler strategy becomes available later on. Because WMC is related to one's ability to focus attention, higher WMC participants were less likely to switch to the simple strategy and instead persisted in using the overly complex strategy (Beilock & DeCaro, 2007). Similarly, in a judgment task for which a simple similarity-based strategy performed optimal, participants were more likely to use the optimal strategy when less WM resources were available under cognitive load (Hoffmann, von Helversen & Rieskamp, 2013). In the present study, we therefore expect that higher WMC should pay off when predicting complex, accelerating processes, but higher WMC may not pay off when predicting simple, asymptotic processes.

In line with the findings reported above, we assume that higher WMC fosters the use of well-calibrated rules, which integrate substantial amounts of training information for prediction as opposed to the use of simple exemplar-based extrapolation strategies involving little training information. We designed a prediction experiment based on the exponential functions shown in Fig. 1 to test this assumption and several hypotheses derived from it.

First, we expect participants with higher WMC to show better accuracy during the training phase, as they are more able to induce a rule adequately describing the training information, that is, a rule that is better calibrated to the training data.

Second, if higher WMC participants rely more on rules, their accuracy in training and extrapolation phases should correlate more strongly. The more the induced rules correct, the more this should benefit later predictions, the more the rules are incorrect, this should harm later predictions.

Third, we expect lower WMCs participants to rely more on only the most recent information available for extrapolation, while rule-based predictions associated with higher WMC should integrate more training information. To test this hypothesis we constructed a quadratic “twin” process for each exponential stimulus. For each point in the training phase, exponential processes and their quadratic twins had a different value *except* for the two points just before the extrapolation region, which were identical. Hence, the more participants rely only on the most recent exemplars for extrapolation, the more similar their extrapolation for both twins should be. Rule-based extrapolation, which by definition integrates more training information, should increase the difference in predictions between twins. In contrast to previous research, which equated rule-use with higher accuracy (McDaniel et al., 2014), this design allows to investigate whether prediction strategy and prediction accuracy are differentially affected by WMC.

Fourth and finally, differences in relying on rule-based versus exemplar-based predictions should result in an interaction of WMC and stimulus type with respect to extrapolation accuracy. We expect that the increased use of rule-based prediction improves prediction accuracy for the increasingly non-linear part of accelerating functions. However, as complex rules have no benefit for predicting the increasingly linear part of asymptotic processes, we expect no difference between higher and lower WMC in these conditions

Method

Design. Each participant predicted four processes, two exponential and their respective quadratic twins. Acceleration of the exponential processes (accelerating or asymptotic) was varied between-subjects, direction (increasing or decreasing) within-subjects. To reduce carryover effects, WMC tests were interleaved between prediction tasks.

Participants. In total, 290 students from Heidelberg University participated (201 female, mean age = 23.0, *SD*=2.7). All participants gave written informed consent, and were rewarded with 5 Euro.

Materials. Four exponential and quadratic function pairs were constructed such that the twinned functions were identical at the two points closest to extrapolation (time points 80 and 85) but different at all other points (see Fig.1 and Table 1). Functions were constructed for a time range (x-value) from 0 to 115 increasing in steps of 5, i.e., 24 time steps.

Table 1. Stimulus functions.

Acceleration	Direction	Exponential function	Quadratic twin function
accelerating	increasing	$3000+e^{(0.045*x+4.2)}$	$0.6494*(x + 12.355)^2$ - 99.13)
accelerating	decreasing	$2000+e^{(-0.045*x+8)}$	$0.2445*(x - 89.042)^2 + 2061$
asymptotic	increasing	$1500 - e^{(-0.045*x+7)}$	$-0.0108*(x - 137.6351)^2 + 1506.15$
asymptotic	decreasing	$1500 - e^{(0.045*x+2))}$	$-0.0491*(x + 57.979)^2 + 2165.11$

Working memory tests. WMC was assessed using digit span forward (DSF), digit span backward (DSB) and letter-number-sequences (LNS). We selected these WM measures as they are easy to administer and structural equation modeling shows that they load on the same working memory construct assessed by more complex measures such as operating span, LNS somewhat more so than Digit Span (Shelton et al., 2009). In both span tasks, increasingly longer series of digits were presented which participants repeated in forward (DSF) or reverse (DSB) order using an on-screen keyboard. In the LNS task, increasingly longer series of alternating digits and numbers were presented, and participants repeated the numbers in ascending and the letters in alphabetical order. Span lengths ranged from four to a maximum of nine trials for DSF/LNS, and eight for DSB. Each subtest was finished after failing two (LNS: three) consecutive trials of a given span length. Individual WMC scores are the mean of the z-standardized scores (consisting of the sum of correct trials) for all subtests.

Procedure. Participants were instructed that they were to predict four different processes over time. The nature of the processes was left unspecified. Each process contained 24 trials, 18 learning and 6 extrapolation trials. During each trial, participants were shown the current time point (labeled “time”) approximately in the middle of the computer screen and entered their prediction for that time point as a number into a textbox (labeled “value”) directly below. Time points then increased in steps of five from 0 to 115, and for each time point, participants made one prediction. During time 0 to 25 and 55 to 85, participants then received feedback showing the correct value of the process at that time, displayed below the time point and their prediction. During time 30 to 55 and 90 to 115, no feedback was given. The first block of trials without feedback served to enhance the learning process trough an intermediate testing phase (cf. Kang et al., 2011), the second block constitutes the main extrapolation region (Fig.1). If not noted otherwise, “predictions” refers to the extrapolation phase in the remainder of this text. As previous studies did not control for potential effects of explicitly instructing participants to search for a rule, half of the participants was instructed to use the feedback provided to find a rule describing the processes, while the other half was instructed to simply observe the feedback provided and use their intuition to make predictions.

Results

Outliers. If more than half of the predictions for a single function had z-standardized values > 5 relative to the whole sample, data from the corresponding participants were excluded from the analyses for this function. This resulted in 4% of all data points excluded. The experiment had at 80% power at $\alpha < .05$ to detect effects of d = 0.35 in between-subjects comparisons and dz = 0.26 for within-subjects comparisons (calculation based on cells with lowest counts).

Dependent variables. As indices of accuracy, we used the Mean Absolute Error (MAE) and the Mean Relative Error (MRE), where MRE is the MAE standardized by the correct function value at any given point. For comparisons within function twins we use MAE, otherwise MRE to correct for different scaling. To assess how much participants differentiate in predictions between two process twins as a measure of strategy-use, we simply calculated the mean difference of predictions. The sample was split into higher and lower WMC participants based on the median.

Effects of instruction. Supporting the use of unspecified instructions in previous studies, the dedicated rule-search versus intuitive instructions did not exert an effect on accuracy nor on differentiation, neither for accelerating nor for asymptotic processes, all *p's* >.1. We therefore excluded this factor from further analyses.

Effects of acceleration and direction. We compared relative prediction errors (MRE) as a function of acceleration (accelerating vs. asymptoting) and direction (increasing vs.

decreasing) of the exponential processes to assess their difficulties. We found a main affect of both acceleration $F(1, 222) = 643$, $p <.001$ and direction $F(1, 221) = 269$, $p <.001$, and a significant interaction $F(1, 221) = 263$, $p <.001$ such that while prediction errors were generally higher for accelerating (M=.44 SD=.011) compared to asymptotic functions (M=.013 SD=.013), and generally higher for decreasing (M=.35 SD=.015) compared to increasing processes (M=.10 SD=.005), prediction errors increased especially for the accelerating decreasing process (M=.68 SD=.02). The quadratic functions were in between accelerating and asymptotic functions, with a smaller error than their accelerating twins, $t(154)=-21$, $p <$ .001, but a larger error than their asymptotic twins $t(113)=8.4$, $p<.001$. In sum, the processes had the following difficulties: increasing < decreasing and asymptotic < quadratic < accelerating.

WMC and training accuracy. Testing whether participants calibrate their prediction strategies to the training data as a prototypical sign of rule-induction, we found that higher compared to lower WMC participants' strategies were better calibrated (lower MAEs) in all processes, that is, the accelerating increasing (M=210, SD=118 vs. M=283, SD=170), $t(137.3) = 3.0$, p=.002, accelerating decreasing (M=25, SD=9 vs. M=29, SD=12), $t(123) = 2.0$, p=.003, asymptotic increasing (M=82, SD=63 vs. M=122, SD=82), $t(76.8) = 2.7$, p=.004, and asymptotic decreasing (M=165, SD=136 vs. M=207, SD=144), $t(107) = 1.6$, p=.06.

We then tested to what extent participants use their—more or less—calibrated strategies for predictions. In the accelerating processes, calibration and prediction errors (MAE) were correlated for both lower, $r(66)=.39$, p=.001, and higher WMC participants, $r(71)=.29$, p=.007 (no sig. difference, $z = .56$, $p>.1$). Interestingly, however, this pattern did not hold for the asymptotic processes. Calibration and prediction errors were unrelated for lower WMC participants, $r(55)= -.09$, p=.24, but significantly related for higher WMC participants, $r(60)=.35$, $p=.003$ (sig. difference, $z = 2.39$, $p=.017$). Very much in line with these results, variances of prediction errors (MAE) of higher and lower WMC participants were identical for both accelerating process, both $p > .1$, but variances were bigger for higher compared to lower WMC participants for both the asymptotic increasing ($SD=22$ vs. $SD=17$), $F(1, 98) =9.4$, $p =.003$, and decreasing process ($SD=33$ vs. $SD=24$), $F(1, 100) = 2.3$, $p =.04$.

WMC and information-use. Concerning how much participants differentiated in their predictions between process twins, there were no differences between higher compared to lower WMC participants in the asymptotic processes, both increasing ($M=2$, $SD=43$ vs. $M=-5$, $SD=51$), $t(97) =.788$, $p=.43$, and decreasing ($M=38$, $SD=110$ vs. $M=43$, $SD=152$), $t(77.7) =.16$, $p=.87$, suggesting that both groups used more or less identical predictions in these cases. Higher compared to lower WMC participants' differentiations were significantly different, however, in the accelerating processes, both increasing, $t(118) =-1.7$, $p=.054$, and—more strongly so—decreasing, $t(76.8) = 3.2$, $p=.001$. The descriptive differentiations (calculated exponential function value – quadratic function value) also show that only for high, but not for lower WMC participants, differentiations were in line with the correct trajectories of exponential versus quadratic processes, namely positive for the accelerating increasing ($M=197$, $SD=1089$ vs. $M=-156$, $SD=1309$), and negative for the accelerating decreasing process ($M=-43$, $SD=149$ vs. $M=48$, $SD=145$).

In sum, higher WMC calibrated their prediction strategies more to training data and used their calibrated strategies consistently for prediction in all process types. Lower WMC participants, in contrast, used (albeit less-)calibrated strategies when predicting accelerating processes, but did not use calibrated strategies at all when predicting asymptotic processes.

WMC and prediction accuracy. To investigate a potentially differential influence of WMC and process type on prediction accuracy in an ANOVA, we z-standardized errors

(MAE) to calculate interaction effects. In line with our expectation, we found a significant interaction between process type (accelerating vs. asymptoting) and WMC (higher vs. lower), $F(1, 222) = 5$, $p = .027$, but no significant interaction between WMC and direction of the process (increasing vs. decreasing), $F(1, 219) = 0.3$, $p = .60$, suggesting that the differential influence of WMC on prediction accuracy holds for both increasing and decreasing processes. Specifically, in the accelerating processes, higher compared to lower WMC participants made better predictions (lower MAEs) for both increasing (M=2109, SD=906 vs. M=2419, SD=1137), $t(147) = 1.8$, p=.03, and decreasing processes (M=211, SD=106 vs. M=270, SD=113), $t(143) = 3.3$, p=.001. Higher compared to lower WMC participants predicted higher values for the accelerating increasing (M=8197, SD=1041 vs. M=7918, SD=1337), $t(147) = 1.4$, p=.007, and lower values for the accelerating decreasing process (M=880, SD=145 vs. M=933, SD=126), $t(143) = 2.4$, p=.001, showing that higher WMC participants' mean predictions followed more the correct trajectories of the accelerating processes.

As implied by the significant interaction between WMC and process type, this pattern of results did not hold for the asymptotic processes. Despite differing prediction strategies, higher and lower WMC participants' predictions were virtually identical for both the asymptotic increasing (M=1506, SD=22 vs. M=1500, SD=17), $t(102) = -1.4$, p=.16, and decreasing process (M=2007, SD=50 vs. M=2011, SD=31) $t(99) = .46$, p=.64. Furthermore, higher WMC participants' predictions were not more accurate compared to lower WMC participants in both the asymptotic increasing (M=20, SD=21 vs. M=17, SD=14), $t(95.3)$ = -1,1, p=.29, and decreasing process (M=33, SD=43 vs. M=24, SD=27), $t(72.4) = -1.4$, p=.16, but were, as descriptive values show, even less accurate by tendency.

Discussion

This study investigated how working memory capacity limits are associated to prediction strategy-use and prediction accuracy. To measure strategy-use, we employed three operationalizations. We found that, first, higher WMC participants' prediction strategies were better calibrated to the training data for every single process, suggesting that higher WMC participants were better able to align their prediction strategies to the structure of the training data. Second, higher WMC participants' calibration and prediction accuracy were correlated in every single process, whereas for lower WMC participants, calibration and accuracy were completely uncorrelated in the asymptotic processes, suggesting that only higher WMC always used calibrated strategies for predictions. And third, only higher, but not lower, WMC participants differentiated between accelerating function twins in the direction of the correct process trajectory: They predicted higher absolute values for the accelerating increasing process compared to its quadratic twin, and lower absolute values for the accelerating decreasing process compared to its twin, suggesting that their prediction strategies differentiated between exponential and quadratic process twins in line with the correct process trajectories in these cases. In sum, both training and prediction results support the central hypotheses that higher WMC facilitates the use of better-calibrated rule-based prediction strategies.

It may seem intuitive that the ability to calibrate prediction strategies better to match the training information should always be beneficial. However, we found an interaction between WMC and the type of the to-be predicted process. Higher WMC participants were more accurate predicting the accelerating processes, but not more accurate predicting the asymptotic processes. Instead, higher and lower WMC participants predicted virtually identically values in both the asymptotic increasing and decreasing processes—despite higher

WMC participants' clearly better calibrated strategies. Higher WMC hence was more beneficial for making accurate predictions only in processes that benefit from information-intensive well-calibrated rules, but was not beneficial for making predictions of processes that can be predicted using information-frugal, exemplar-based strategies. However, higher WMC participants generally used well-calibrated rule-based strategies more often, both when predicting accelerating processes (where this is the optimal strategy) and when predicting asymptotic processes (where a simpler and equally accurate strategy is available).

One dissociation between higher and lower WMC participants seems particularly noteworthy. For higher WMC participants, training errors were always predictive of prediction errors, suggesting that the degree to which their prediction algorithms were calibrated to the training data determined their prediction accuracy: Those inducing well-calibrated rules tended to make better predictions than those inducing poorly calibrated rules. For lower WMC participants, these results were mirrored in case of the accelerating processes, only on a lower level: lower compared to higher WMC participants' prediction strategies were less calibrated to the training data, and consequently produced less accurate predictions. In case of the asymptotic processes, however, the difference between higher and lower WMC seemed to be qualitative in nature, since lower WMC participants' training errors were completely unpredictive of their prediction errors. This lacking correlation between training and prediction errors suggests that lower WMC participants relied on simple exemplar-based prediction strategies in these cases. Exemplar-based strategies make only minimal use of the training data, do not imply calibration to the training data at all, and hence are virtually unreflective of training error.

Supporting this interpretation, we found that in case of the asymptotic processes, higher WMC participants’ prediction errors were more variable than lower WMC participants’ errors. The different variances reflect that while higher WMC participants made use of a whole range of comparatively poorly up to comparatively well-calibrated strategies, the exemplar-based strategies used by lower WMC lead to rather consistent results as they rely on only the two last training exemplars, and therefore hardly allow for inter-individual differences. Higher WMC participants' strategies were in need of good calibration, with comparatively badly calibrated rules leading to comparatively bad predictions. Lower WMC participants' simple strategies, in contrast, were robust and unsusceptible towards how much (or how little) one has learned during training. Using robust strategies that do not reflect to what extent one was successful to abstract the systematic structure of the process, therefore proved a particularly adaptive strategy for lower WMC participants. In fact, lower WMC participants' strategies proved robust not only in the sense of being unreflective of learning—that is, when compared to others also applying this strategy—but even in the sense of being equally successful as those applying other, more demanding rule-based strategies.

We found that higher WMC does not always lead to more accurate predictions, but does so only if the structure of the to-be predicted process requires one to induce rules. This is an important qualification of the previous finding on how individual WMC affects the accuracy of predictions showing that higher WMC leads to increased use of rules, and to better predictions (McDaniel et al., 2014). However, McDaniel et al. operationalized rule-induction via, and only via, more accurate predictions. In the present study, a more direct (independent of prediction accuracy) and encompassing (employing three operationalisations) strategy-assessment was used. Results showed that lower compared to higher WMC participants were in fact no less accurate predicting processes where simple information-frugal strategies perform optimal.

The present results are comparable to the phenomena of over- and underfitting in machine-learning (Hawkins, 2004; Todd & Gigerenzer, 2000). Per definition, underfitting occurs when rendering an algorithm more complex and better-calibrated to fit the structure of the training data would still pay off in terms of better predictions. Overfitting in turn occurs when higher complexity and better calibration does no longer pay off, that is, when a complex compared to a simpler prediction algorithm does not produce more accurate predictions. In the present study, both phenomena occurred: In the accelerating processes, lower WMC participants' prediction strategies were calibrated less, and this resulted in worse predictions. In the asymptotic processes, however, higher WMC participants' better-calibrated prediction strategies did no longer produce better predictions compared to lower WMC participants' simpler strategies. Thus, lower WMC participants' prediction strategies underfitted the accelerating processes, and higher WMC participants' prediction strategies overfitted the asymptotic processes.

Perhaps counterintuitively, higher WMC participants used suboptimal strategies particularly when performing the comparatively simple task of predicting asymptotic processes. What explains this finding, however, is that predicting asymptotic processes is not only comparatively simple, but actually *surprisingly* simple. The steep training region suggests structural complexity (and hence the use of more complex rules), but their extrapolation region runs relatively flat (and hence is best predicted with simple linear strategies). Wagenaar and Sagaria (1975) described this phenomenon as the cognitively "easy part" and "hard part" of exponentiality. Our results hence suggest that in cases where "easy parts" directly follow "hard parts", higher WMC may not be a helpful resource. Quite the contrary: The higher cognitive control performance may effectively impede one from using more elegant approaches. This finding is analogous to the concept of "mental set" in problem solving, where simpler solutions for a problem are overlooked if a more complex solution routine has been built up first. Analogously to how the study by Beilock and DeCaro (2007) has shown that the tendency to inflexibly maintain mental set is especially pronounced for higher WMC participants, we found that higher WMC participants were prone to overlook simple—and accurate—prediction strategies.

This effect may also explain the contradictory results on why higher cognitive capacity sometimes leads to more adaptive strategy-use (Bröder, 2003; Lewandowsky et al., 2012), and sometimes to the use of overly complex, non-adaptive strategies (Beilock & DeCaro, 2007). In the experiments by Bröder (2003) and Lewandowsky et al. (2012), there was one strategy that performed optimal *throughout* each task, whereas in Beilock and DeCaro (2007) and in the asymptotic processes in the present study study, complex strategies were successful during the first trials, but a simpler solution became available later on. It was then that higher WMC participants persisted in using their more complex strategy. Higher WMC participants seem to be able to construct a more initially suitable, complex strategy for a given task, but may have difficulty giving up this strategy when a simpler strategy becomes available.

Cowan (2010) suspected that the positions of working memory capacity limits as a weakness versus as a strength may not be incompatible, but that each one may have its merits. With respect to predicting non-linear dynamic processes, we find that whether higher capacity is beneficial, or not, depends on the type of process to be predicted. Higher working memory capacity limits are generally associated with better-calibrated strategies that are more aligned with the structure of the processes. This leads to better predictions of difficult, non-linear processes, but can result in overfitting and applying overly complex strategies when predicting simple processes. In a word: Finding structure is good, but finding more structure is not necessarily better.

References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133–137.

Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983-998.

Bott, L., & Heit, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 38-50.

Bröder, A. (2003). Decision making with the“ adaptive toolbox”: Influence of environmental structure, intelligence, and working memory load. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 611-625.

Bröder, A., Newell, B. R., & Platzer, C. (2010). Cue integration vs. exemplar-based reasoning in multi-attribute decisions from memory: A matter of cue representation. *Judgment and Decision Making*, *5*(5), 326–338.

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K.Lamerts & D.Shanks (Eds.) *Concepts and Categories* (pp. 405–437).

Busemeyer, J. R., & Myung, I. J. (1992). An adaptive approach to human decision making: Learning theory, decision theory, and human performance. *Journal of Experimental*

Psychology: General, *121*(2), 177-192.

Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57.

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology*, *65*(3), 439–464.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968-986.

DeStefano, D., & LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, *16*(3), 353–386.

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*(1), 1–12.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2013). Deliberation's Blindsight: How Cognitive Load Can Improve Judgments. *Psychological Science*, *24*(6), 869–879.

Hoffmann, J. A., von Helversen, B., & Rieskamp, J. (2014). Pillars of judgment: How memory abilities affect performance in rule-based and exemplar-based judgments. *Journal of Experimental Psychology: General*, 2242–2261.

Janacsek, K., & Nemeth, D. (2013). Implicit sequence learning and working memory: correlated or complicated? *Cortex*, *49*(8), 2001–2006.

Kalish, M. L. (2013). Learning and extrapolating a periodic function. *Memory & Cognition*, *41*(6), 886–896.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: knowledge partitioning and function learning. *Psychological Review*, *111*(4), 1072-

1099.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71.

Kang, S. H., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review*, *18*(5), 998–1005.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*(3), 278.

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute d ecision making: Contingent, not automatic, strategy shifts. *Judgment and Decision Making*, *3*(3), 244–260.

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 811-836.

Kwantes, P. J., Neal, A., & Kalish, M. (2012). Item order matters in a function learning task. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 90–97.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83-97.

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 720-738.

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 881-904.

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in psychology*, *5*.

MacKinnon, A. J., & Wearing, A. J. (1991). Feedback and the forecasting of exponential change. *Acta Psychologica*, *76*(2), 177–191.

Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive Processing*, *11*(2), 103–121.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: an ecological model of strategy selection. *Psychological Review*, *118*(3), 393-437.

Mcdaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule-based and associative-based models. *Psychonomic Bulletin & Review*, *12*(1), 24–42.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*, *143*(2), 668-693.

McDaniel, M. A., Dimperio, E., Griego, J. A., & Busemeyer, J. R. (2009). Predicting transfer performance: a comparison of competing function learning models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 173-195.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700-708.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534-552.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences*, *23*(05), 727–741.

Wagenaar, W. A., & Sagaria, S. D. (1975). Misperception of exponential growth. *Perception & Psychophysics*, *18*(6), 416–422.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 256-263.

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, *21*(4), 258–262.