Rethinking intuition of accumulation

Chris Browne[a] and Paul Compston[a]
The Australian National University
Corresponding author: Chris.Browne@anu.edu.au

Abstract

We have created a hands-on, concrete learning activity to help people with no systems thinking training develop good mental models around the concept of accumulation. This paper takes up Sterman's 2008 challenge to create new methods to develop intuitive systems thinking capabilities so that "people can discover, for themselves, the dynamics of accumulation and impact of policies." Our case study is the Earth's carbon cycle. Participants in our activity work as a group manipulate a concrete system through a series of prompts, which encourage focussed discussion and development of mental models. One experimental treatment had 75.9% of participants demonstrate a correct response to the problem and sound understanding of accumulation principles. This treatment also had 93.1% of participants drawing a trajectory of future anthropogenic carbon emissions that would see a reversal of its current growth trajectory and greatly exceed the current targets described in international agreements. We propose that this approach is useful for constructing intuition of accumulation principles, and has implications for educators wanting to improve their students' thinking about fundamental systems principles.

keywords: systems thinking, education

A brief and recent history of assessing accumulation

It has been over six years since John Sterman authored a policy forum article in *Science* (Sterman 2008) calling for methods of improving the public's intuition around the accumulation of carbon in the Earth's atmosphere. These observations came from research conducted with Linda Booth Sweeney (Sterman & Booth Sweeney 2002; 2007), which concluded that even highly educated adults had poor understanding of the relationship between carbon accumulation and flows. Since then, nations have continued to negotiate and implement strategies around pricing carbon, scientists have continued to observe a changing climate, and research continues on various methods mitigate the effects of anthropogenic carbon emissions. During this time, however, the atmospheric carbon concentration has continued to increase (Stocker *et al* 2013). It is clear that life on this shrinking planet will be drastically different as this behaviour continues into the future, yet there is little evidence to suggest that the public understanding of accumulation of carbon in the atmosphere has improved.

Distinguishing the behaviour of stocks (an entity that accumulates or depletes over time) and flows (the rate at which that entity changes) has been the focus of many studies in the system dynamics community. These studies reveal some alarming observations. Notable conclusions are that highly educated adults have a poor understanding of the most basic stock-and-flow concepts (Booth Sweeney & Sterman 2000; Cronin & Gonzalez 2007) and that highly educated adults' understanding of stocks and flows violate the law of conservation of mass (Sterman & Booth Sweeney 2002; 2007). Efforts have been made to improve results, such as exploring presentation formats of information (Cronin *et al* 2009; Brockhaus 2013; Sedlmeier *et al* 2014). However, many conclude that these efforts have little effect on results, and further examination is required. Sterman (2010) concludes that formal system dynamics instruction improves understanding of stock-and-flow systems, but that a minority of students still are confused by stocks and flows.

Reframing the 'accumulation problem'

These recent results, though alarming for the system dynamics community, should hardly surprise us as systems thinkers. The system lens that Dana Meadows eloquently describes provides an obvious way to reframe of this problem. For example:

> The flu virus does not attack you; you set up the conditions for it to flourish within you. (Meadows 2008)

By applying the system lens to the accumulation problem, we can effectively reframe it. Our hypothesis is based around the observation that:

> It is not that our students' intuition of accumulation is poor; instead that we set up conditions where students are prone to fail.

We found that we had significantly improved results when we set up an engaging activity that can encouraged clear thinking, the exchange of ideas and the opportunity for participants to challenge their mental models. The goal of the activity was not to assess the participants' intuition of accumulation, but rather to test whether a concrete activity to improve their mental models; we wanted students to not just reach the correct answer without direct instruction, but also improve their understanding of accumulation so that they could transfer the concept to other real-world contexts.

To explain the rationale for our activity, we need to be clear about the relevance of our approach to improving mental models. There are three attributes common to definitions of mental models in Meadows (2008, pp. 86-87), Sterman (2000, pp. 15-29), Richmond (2010), Senge (1997) and Maani & Cavana (2007). Mental models are:

- a subset of the real world *(bounded rationality, simplified complexity)*
- built from experience with the real world *(an experiential abstraction)*
- incomplete and therefore difficult to simulate correctly *(dynamically deficient)*

By explaining our approach to developing these attributes in the following section, we will provide the pedagogical rationale for our activity.

On improving mental models of accumulation

Exploring these three attributes of mental models allows us to connect ideas from within the systems thinking community to other areas, such as cognitive science and learning theory, in the context of improving understanding about accumulation of carbon in the atmosphere.

Simplified complexity and mental models

The flow of carbon through the carbon cycle is itself complicated, and involves many feedback structures. The accumulation of carbon in these stocks causes non-linear behaviour in other systems, such as global temperature change, sea level change and to levels of ocean acidity. The problem of climate change is not limited to the physical science. Add to that the social, political and economic drivers and responses to climate change, and the system becomes too large to intuit through a collection of mental models.

Drawing an appropriate boundary around the system of interest is useful for interrogating the problem at hand. In our activity, we wanted participants to be able to clearly articulate the effect of the anthropogenic disruption to the carbon cycle on the accumulation of carbon in the atmosphere. Sterman's (see Kunzig 2009) carbon bathtub provides a useful simplification for this activity, where the accumulation of carbon in the atmosphere is determined by the difference between the rates of additions to and removals from the atmosphere. The variables considered in our activity are shown in Table 1 in the form of a model boundary chart, where the endogenous and exogenous variables were included in the model, and the excluded variables were considered outside the scope of the problem.

Table 1: Model-boundary chart for design of the activity. The endogenous and exogenous variables are included in the model, and the excluded variables are intentionally not considered further in the model.

Endogenous	Exogenous	Excluded
atmospheric carbon stock terrestrial, geological and oceanic carbon stock rate of natural additions rate of natural removals rate of anthropogenic additions	desired atmospheric carbon level	rate of anthropogenic removals carbon flux between terrestrial, geological and oceanic stocks heat storage in stocks temperature difference between stocks insolation and radiative forcing climate feedback processes social, political, economic intervention

A conceptual diagram of this model, and an equivalent stock-and-flow model is shown in Figure 1. The flows in the natural carbon cycle are approximately equal (Barker *et al* 2007), with natural additions and natural removals effectively cancelling each other out. The anthropogenic additions due to the burning of fossil fuels, although much smaller in magnitude than the natural cycle, mean that carbon is moved from the terrestrial stock to the atmospheric stock resulting in an increased accumulation of carbon in the atmosphere.

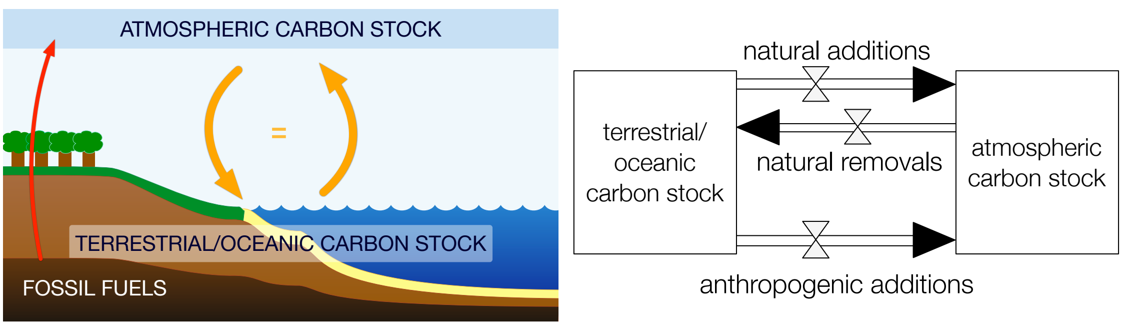


Figure 1: Representation of the simplified carbon cycle model. Left: a conceptual diagram of the key stocks and flows. Yellow arrows show the natural cycle, and the red arrow shows the anthropogenic disruption due to burning fossil fuels. Right: a stock-and-flow diagram of the same system. Carbon moves from the terrestrial/oceanic carbon stock to the atmospheric carbon stock through the natural and anthropogenic additions flows, and is returned through the natural removals flow.

This, of course, is a simplification of the complex carbon cycle. It is useful, however, for the purposes of improving intuition of accumulation. We propose that this allows the participants in our activity to clearly see the dynamics within our system of interest.

Experiential abstraction and mental models

There is a strong connection between the improvement of mental models and the learning process. This is shown in Argyris's (1976) description of single- and double-loop learning, where double-loop learning involves the active development of mental models in response to feedback and information from the real world.

Perspectives from cognitive science and learning theory complement the systems view in reframing the accumulation problem as a learning problem, specifically in relation to how knowledge is constructed through experience with the real world. Take, for example, these perspectives from the different fields:

> ***Systems thinking observations***
> Any child who can fill a water glass or take toys from a playmate knows what accumulation means. (Forrester 2009)
>
> If you have had much experience with a bathtub, you understand the dynamics of stocks and flows. (Meadows 2008)
>
> ***Cognitive science observations***
> Our experiences with physical objects (especially our own bodies) provide the basis for an extraordinarily wide variety of ontological metaphors, that is, ways of viewing events, activities, emotions, ideas, etc., as entities and substances. (Lakoff & Johnson 1980)
>
> ***Learning theory observations***
> The learner constructs knowledge inside their head based on experience. Knowledge does not result from receipt of information transmitted by someone else without the learner undergoing an internal process of sense making. (Martinez & Stager 2013 describing Piaget's constructivism)
>
> Children learn to speak, learn the intuitive geometry needed to get around in space, and learn enough of logic and rhetorics to get around parents—all this without being "taught." (Papert 1980)

Our understanding of the world is shaped by the conceptual metaphors hidden in the language that we use (see Lakoff & Johnson 1980). One reason that accumulation can be so easily understood and widely applied is because its behaviour is consistent with the *container metaphor*, where objects (stocks, people, ideas, et cetera) are placed into and taken out of a container.

This conceptual metaphor allows imaginative thought processes, where we understand one thing in terms of another. Phrases such as *I am in trouble* or *I got out of trouble* make sense because of the container schema logic, based on our physical experience with containers. Objects can be situated in relation to the container, such as in, out, on, under, in front, behind or to the side. *The trouble is behind me* or *I'm on top of the problem.* Containers have a geometry that allows them to fill and drain. *I am in a lot of trouble* or *The trouble is disappearing slowly.* The container metaphor is universal to human cognition because of our physical experience with, and our shared *a priori* understanding of, containers.

The bathtub metaphor is a useful instance of the container metaphor. It references a concrete, physical experience that allows us to use our understanding of the bathtub system to intuit the behaviour of other systems with similar structures. Newell (2012) demonstrates that this is because of the conceptual mapping between the bathtub (the conceptual source domain) and the system component (the conceptual target domain). Similarly, the carbon bathtub is useful because of the mapping between a bathtub system and the carbon cycle; we can interpret the dynamics of the carbon cycle because of our concrete experiences with a bathtub.

We have built on the bathtub metaphor to create a physically manipulable representation of the carbon cycle. We call this system 'Tubs & Pumps' (hereinafter T&P), where the stocks are represented by tubs of water and pumps allow participants to control the flows between the stocks. A photo of the T&P system in use is shown in Figure 2.

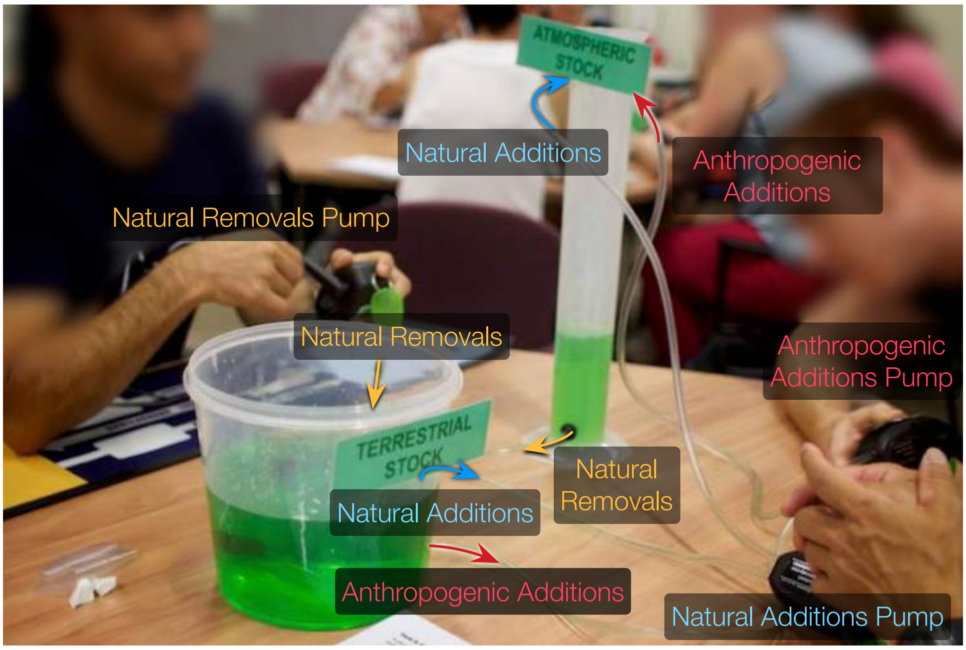


Figure 2: Annotated photo of the Tubs and Pumps (T&P) system. Participants manipulate pumps which change the state of the stocks. Coordination is required by team members to control the system.

Although participants are involved in the activity, the mapping of components between the T&P system and the carbon cycle is done implicitly through labelling the components. This also helps participants to understand who is controlling the flow of each pump. Building on Newell's (see 2012 p779) conceptual mapping of the bathtub metaphor, we have mapped the T&P system components to the carbon cycle in Table 2.

Table 2: Mapping of the Tubs & Pumps Analogue. The components in the conceptual source domain are mapped to the corresponding component in the conceptual target domain, allowing participants to construct an understanding of one system (the carbon cycle) in terms of another (tubs and pumps). Note: arrows " ⇒" represent the expressions "corresponds to" or "maps to".

Conceptual source domain: tubs & pumps system		**Conceptual target domain: the carbon cycle**
The water in each tub	⇒	The accumulation of carbon at time t
The amount of water in the atmospheric tub	⇒	The amount of carbon accumulated at time t in the atmosphere; the atmospheric carbon concentration
The amount of water in the terrestrial tub	⇒	The amount of carbon accumulated at time t in terrestrial stocks, including oceanic and geological stocks.
Water entering the tub through the 'natural additions' pump	⇒	The natural processes that increase the amount of carbon in the atmosphere, such as decomposition of organic matter
Water entering the tub through the 'anthropogenic additions' pump	⇒	The processes that increase the amount of carbon in the atmosphere due to human activity, such as burning fossil fuels
Water leaving the tub through the 'natural removals' pump	⇒	The natural processes that decrease the amount of carbon in the atmosphere, such as afforestation

Participants are able to understand the behaviour in carbon cycle through logic that is created through experience with the T&P system. For example, the speed of the anthropogenic additions pump relates to the rate of anthropogenic additions into the atmosphere. Participants can visualise scenarios in the carbon cycle by interacting with the T&P system; for example, an increase of burning fossil fuels would be represented in the T&P system by pumping the anthropogenic additions pump faster.

The learning in this activity is not only concrete, but also social. The groups are required to work together to achieve a common goal - to ensure that the level of water in the atmospheric carbon tub does not exceed the capacity of the tub. The only way to achieve this, with the natural cycle in dynamic equilibrium, is to stop the anthropogenic additions pump. And quickly!

Dynamic deficiency and mental models

A final argument for reframing the accumulation problem is the observation that mental models are dynamically deficient. The obvious solution for the systems thinker is to get the learner to construct the complete, feedback-rich model to fill the gaps present in the mental model. On the one hand, the dynamic deficiency can be seen as a problem to be solved, but on the other it can be seen as an opportunity to exploit.

Maxwell *et al* (1994) suggest that decision-makers were not equipped with mental models that displayed closed-loop cause-and-effect structures, but rather became better decision-makers when they were given dynamical insights described as ‘causal chunks’. These are inferential models that form a basis of decision making. In relation to the activity, the model ultimately needs to be useful for making causal inferences consistent with the principles of accumulation, rather than necessarily being able to mentally simulate the dynamics of the complex carbon cycle. Previous work within the systems thinking community show a number of strategies to improve decision making. Namely, that group model building is a useful way of sharing mental models, and that small models, such as those the size of the system archetypes, are easily transferred between contexts.

Group model building is a way for explicitly integrating individual’s mental models; as the parable shows us, a way of seeing the ‘whole’ elephant. In group model building, participants and relevant stakeholders play an active part in the modelling process

(Anderson *et al* 2007), rather than being a process only undertaken by modellers. Involving stakeholders in the modelling process ultimately improves the quality of the model, and the insights and recommendations from the model are more likely to be implemented (Vennix 1996). Hovmand (2014) furthers this argument, suggesting that models built with high participation are likely to have more public acceptance as a basis for community-based system dynamics.

Andersen and Richardson (1997) outline a series of scripts for running group model building activities. These are not formal scripts, but rather routine prompts that have been built up through experience. These scripts cover strategies for defining the problem, conceptualising model structure, eliciting feedback structure, equation writing and parameterisation, and policy development. Newell & Proust (2012) describe a collaborative conceptual modelling (CCM) process that includes a series of prompts: What is the challenge? What is the story? Can I see how you think? What drives system behaviour? What are the leverage points? Can we have new eyes? These prompts are designed to extend the group's thinking, and look at the problem from alternative perspectives. "Good prompts do not burden a learner, but set them free." (Martinez & Stager 2013) This is a valuable insight for the development of a learning activity.

Simple, generic models that can be transferred between contexts is another useful tool for addressing dynamic deficiency. Of most relevance here are the models commonly known as the system archetypes. The system archetypes are generic structures that demonstrate dynamically complex, yet recognisable behaviour.

Take the treatment of the limits to growth archetype in Senge (1997). Limits to growth describes a situation where rapid growth is followed by stagnation. The reinforcing behaviour that promoted the growth eventually slows down due to a limiting factor. Senge offers a number of examples, such as: affirmative employment strategies; learning a new skill such as tennis; start-up businesses or social movements that grow too fast and lose direction; a city that grows rapidly and in doing so increases housing prices, and; an animal population that grows too fast, which leads to overshoot and collapse. Meadows (2008) provides more examples: a new product will eventually saturate a market; a chain reaction in a nuclear power plant or bomb will run out of fuel; a virus will run out of susceptible people to infect; the economy may be constrained by physical capital or monetary capital or labor or markets or management or resources or pollution. Rogers (2003) describes many further examples, such as the diffusion of innovations (ideas or products), and the growth of a sunflower.

Recognising the archetype in a real-world situation allows us to predict future behaviour and recognise opportunities to alter that behaviour; it equips the decision-maker with a proven approach for effective intervention. Experience with the limits to growth archetype makes the decision-maker aware that the rate of growth they are experiencing today may eventually slow down and stop in the future. In a situation where growth is desired, Senge recommends focussing on removing or weakening the limitation. In the tennis example, this might include more coaching to lift the limitation level; in the product example, this could include opening the product to more customers through a price drop, or developing another product to ensure the company’s growth does not also plateau.

The system archetypes are useful because they provide causal insight into real-world situations. They are transferable, easily recognisable and small. We wanted to create our activity to draw on the simplicity of the small models used in the archetypes. The model used in the T&P system is small, with three flows and two stocks. When participants interact with the T&P system, they are generating the conditions that allow their understanding to became dynamically richer; to address the dynamic deficiency. The rules that control the flows are straightforward: the natural cycle is in dynamic equilibrium, and the anthropogenic additions get faster to simulate the reference mode of burning fossil fuels over the last hundred years. The activity is designed to set up conditions where the decision-making is obvious because of their experience with the physical activity; to allow the participants to construct their own knowledge and reach their own conclusions.

Experimental methodology

We conducted a workshop with 96 participants from a first-year environmental science course at a leading Australian university in 2014. This workshop was one of seven workshops undertaken to investigate intuition of accumulation principles during 2013 and 2014 involving over 1,000 participants. Data were collected using a paper-based profile questionnaire and test. The approach for collection of data had human ethics approval, and no personally identifiable information was collected. Participants were not paid, graded or otherwise incentivised, and participation was encouraged but not required for their coursework. Workshops ran as one part of the first tutorial in a course, and lasted approximately 30 minutes including a debriefing discussion. The generic structure of the workshop is shown in Table 3, which also provides an indication of the running time. These times were not adhered to strictly to allow exploration as required, but no workshop went for more than 30 minutes before the collection of results.

Table 3: Structure of the workshops. Times were not adhered to strictly to allow for groups to explore the activity as required.

Running Time	Activity
0-3 mins	Introduction, human ethics explanation and opportunity for questions
3-8 mins	Video introduction and group formation
8-18 mins	Groups undertake activity, guided by workbooks, with minimal intervention
18-25 mins	Upon completion of the group activity, groups given assessment to complete
Remaining time	Present possible and correct solutions and guide class discussion

To control for consistency across the sessions, a video introduction was given to provide a small amount of background for the participants. The video was captioned to aid comprehension, especially for students who were not native English speakers. The video ran for 1min 20sec, and explained the activity, the simplified carbon cycle and the challenge for the activity. The text used in the video mirrored the text used on the first two pages of the activity workbook to ensure that all participants had been given this information before beginning the activity.

The groups completed the activity guided by a workbook (see the supporting materials). Typically, three group members controlled a pump each, and one group member read the workbook instructions. Once the groups had completed the workshop, they reset the equipment and completed the assessment. The assessment included a profile questionnaire which asked for the participants age category, sex, field of study, degree progress and language most comfortable communicating in. The assessment also included three main tasks: a graphical task, a written task, and a causal-logic matching task. These tasks are shown in Box 1. The graphical task is derived from Sterman & Booth Sweeney (2002; 2007), but has had the y-axis values removed to more closely connect with the experience of the T&P activity.

Individual Questionnaire

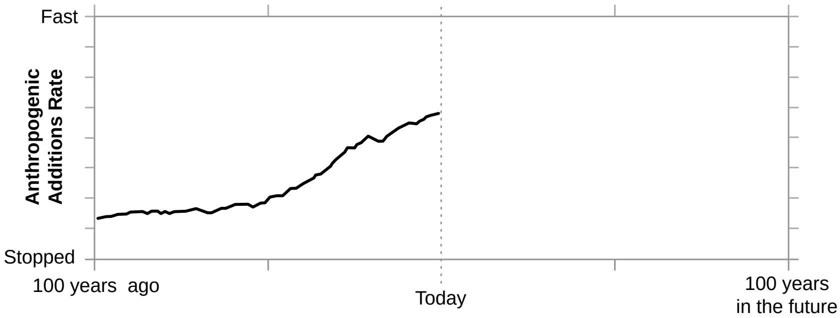
Q1. Do you agree with your group's answers? (please circle)

Not at all ——— Somewhat ——— Half ——— Mostly ——— Completely

Q2. Describe in your own words what the Anthropogenic Additions pump should do to ensure that the goal level is not exceeded:

Q3. Sketch what you think the future Anthropogenic Additions pump rate should be to maintain the goal level.

Future Anthropogenic Additions Rate



Sketch the Anthropogenic Additions rate to reach the atmospheric carbon level goal

Q4. Consider the activity with the tubs and pumps. Circle the box you think is most correct in the table below to complete each sentence:

If the Anthropogenic Additions pump…	…the water level will…		
continues to get faster	go down	stay level	go up
slows to a stop	go down	stay level	go up
remains steady	go down	stay level	go up
slows slightly	go down	stay level	go up
gets faster, then remains steady	go down	stay level	go up

Box 1: Assessment tasks given after the activity. The goal level indicated in Q2 refers to a stabilisation of atmospheric carbon, discussed as part of the last task in the workbook.

Colour coded cards were used to create groups within the activity, handed out around the room in sequence. Up to four groups were created, based on the colour of their card. Upon collection of the assessment sheets, we checked only that this identifier was completed, and asked participants to record this identifier if it was missing. No other checks for completeness or observations concerning the results were made during the workshop.

Once all assessment sheets were collected, an open-ended discussion of the activity was facilitated. The correct graphical response was discussed. Common items of discussion involved what factors may inhibit the ideal solution, whether action is best placed at personal, community, national or global scales and what that action looks like at each level. A typical discussion ended on what take-home messages did students get out of the activity, some of which were shared back with the class.

Treatment groups

Previous work from earlier workshops had shown that a significant proportion of incorrect graphical responses demonstrated a ‘mismatch’ or confusion between the graphical and written responses (see Browne *et al* 2013). For example, a graphical response could show the rate of anthropogenic additions stabilising above today’s levels, but be accompanied with a written description that described the anthropogenic additions slowing down to a stop (see Box 4 for examples). A key motivation for testing treatment groups in this workshop was to investigate whether prompts or activities could help reduce the frequency of this confusion and in turn demonstrate consistent understanding.

Each workshop session was separated into four treatment groups – two treatments with two levels. The first treatment was the use of *Scenario Cards* (SC), which were used to show a range of possible futures, shown in Box 2. The second treatment was the use of a *Group Diagram* (GD), which asked that the group draw a diagram as a group before they complete the workshop. The group sizes are shown in Table 4, with the designations and number of participants. Two sessions did not have enough participants to have all four groups, accounting for the variation of number of tutorials the treatment conditions ran in.

Table 4: Summary of group sizes for sessions used in the results

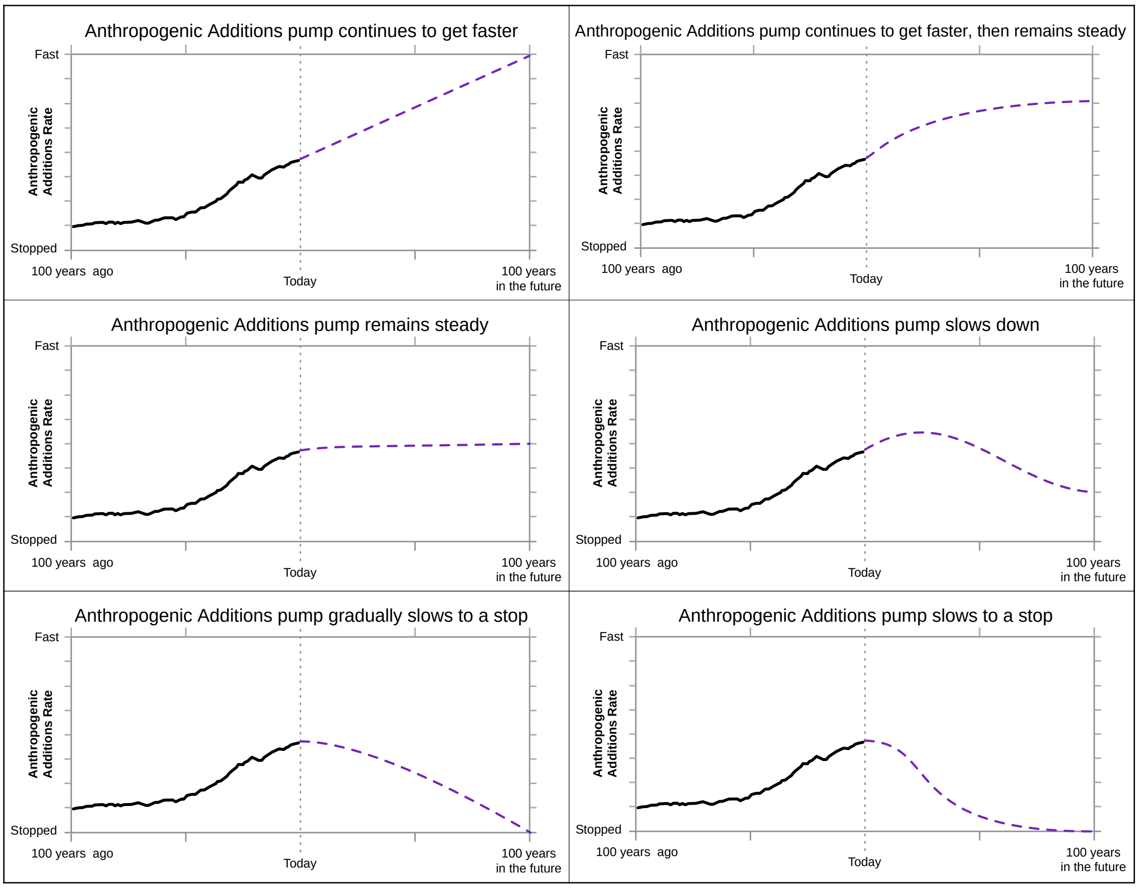
Treatment	Designation	# of Groups	# of Participants
Group Diagram and Scenario Cards	GDSC	8	29
Group Diagram	GD	7	24
Scenario Cards	SC	8	25
Neither GD or SC; the base activity	Base	6	18
		Σ=29	Σ=96

To accommodate the different treatments, four slight variations to the workbook tasks were made. Groups were not made aware that they were undertaking slightly different activities until after the assessment sheets had been collected. The sequence of the workbook activities are shown in Table 5. Groups were randomly allocated to the treatments based on the colour of their allocated cards.

Table 5: Summary of workbook sequence with treatment differences described

Workbook sequence	Description
Overview, Introduction to the Carbon Cycle, Behaviour over Time	*Introduction to and orientation of the activity and the carbon cycle, similar to the video introduction*
Task 1: Assemble Your System	*Participants label the components of the system*
Task 2: Getting Started	*Each pump takes a turn to operate independently, matching a graphical representation*
Task 3: Simulate the Natural Carbon Cycle	*The natural additions and removals pumps operate in dynamic equilibrium*
Task 4: Simulate the burning of Fossil Fuels	*The natural carbon cycle continues in dynamic equilibrium, and the anthropogenic additions pump starts slowly and continues to get faster until the tub is almost full. Discuss strategies on how to stop the water from exceeding the capacity of the tub*
Scenario Card (SC) treatment	*Rather than discussing strategies, participants consult Scenario Cards for ideas on strategies*
Task 5: What now?	*Participants are asked to give a group written description of the successful strategy that the group used, just as in the individual task*
Group Diagram (GD) treatment	*Task 5 included an additional prompt to draw a graphical representation of the strategy as a group, just as in the individual task*

The Group Diagram treatment used the same graph as in the graphical task in Box 1. The Scenario Cards presented six scenarios of future anthropogenic additions, together with the written description of the scenario. They are approximately the size of a postcard. The six scenarios are shown in Box 2, and range from a scenario of increasing rate of anthropogenic additions through to a rate of zero.



Box 2: Information presented on the scenario cards.

Coding of responses

The data from the profile questionnaire and assessment tasks were entered first using a digital form. A blinding process was undertaken through a digital card sort. Discrepancies between the two data entry processes were resolved case-by-case. Here, the questions in the assessment task will be explained.

Graphical task. The graphical task was coded into five categories according to the trajectory of the graphical response: *Correct*, *Decrease*, *Stabilise*, *Increase* and *Other*. Responses where the trajectory transversed multiple categories were coded according to the category the trajectory most agreed with. Where this was extreme or where the response was discontinuous, it was coded as *Other*. Figure 3 shows the rubric used to code the graphical responses.

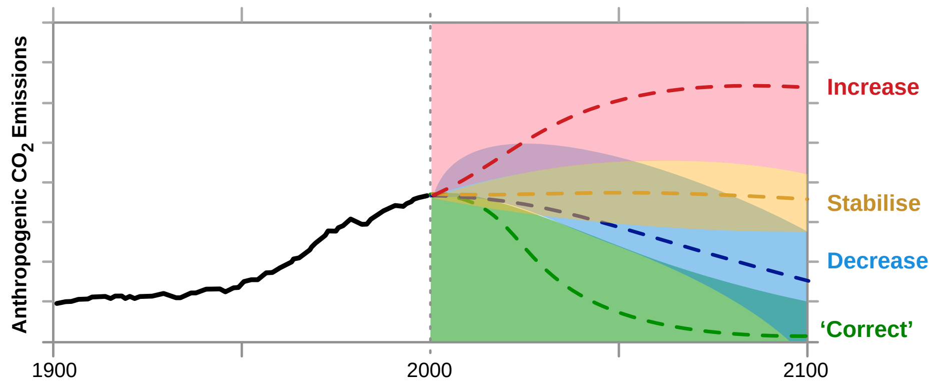
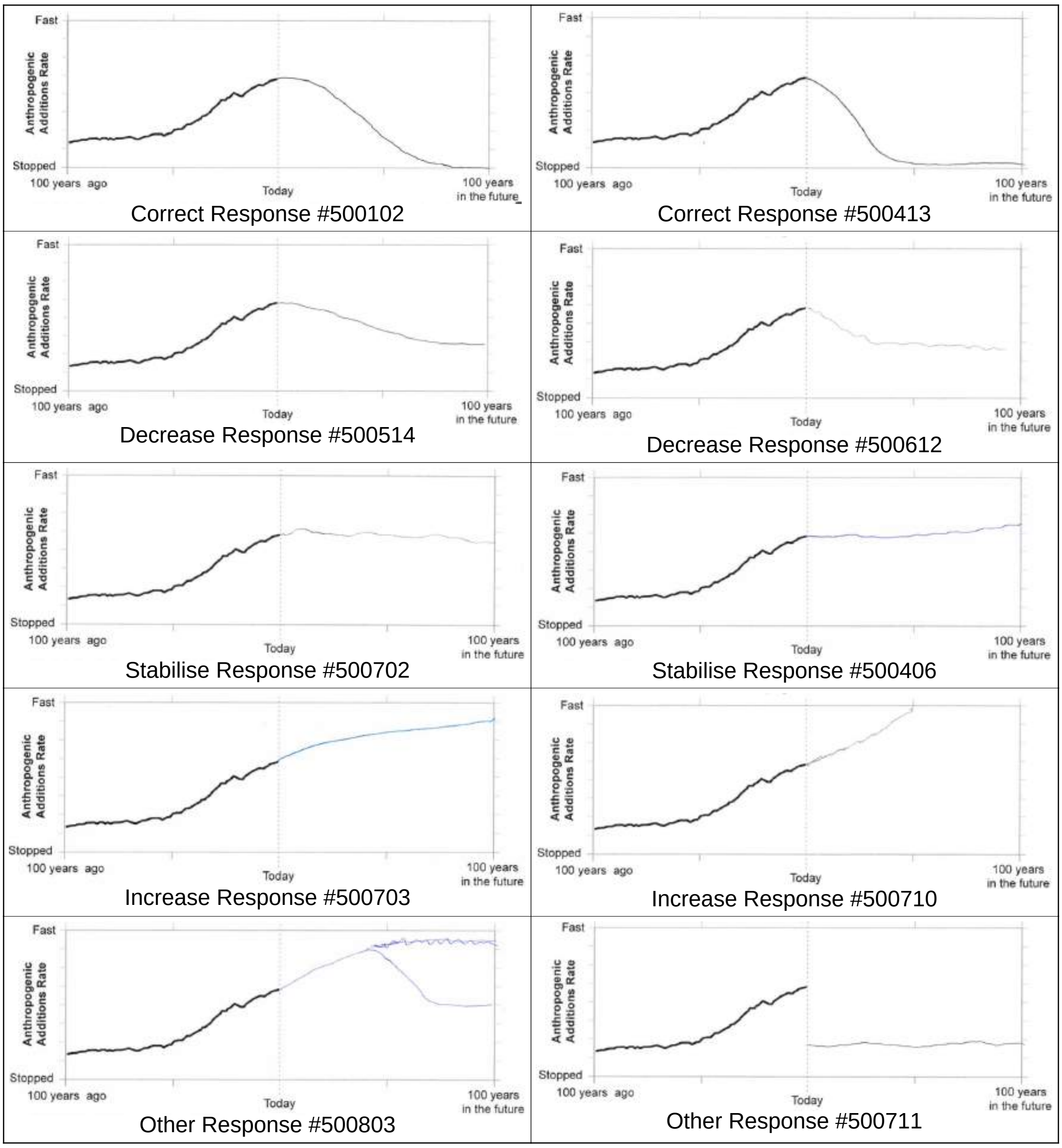


Figure 3: Rubric used to categorise graphical responses, with dotted lines showing typical responses. An 'other' category was used where responses did not fit clearly into the rubric. Y-axis values omitted.

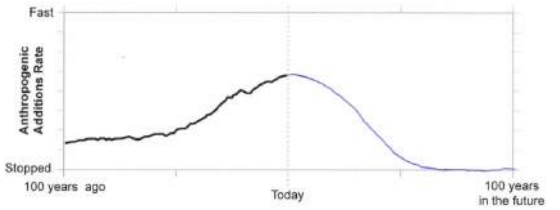
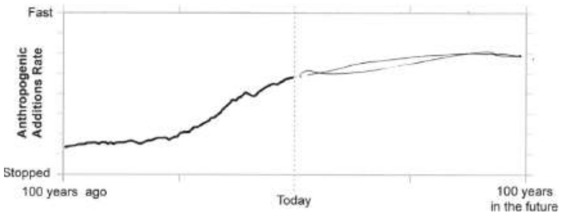
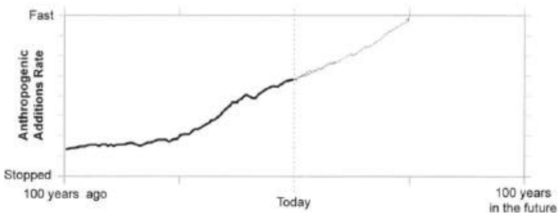
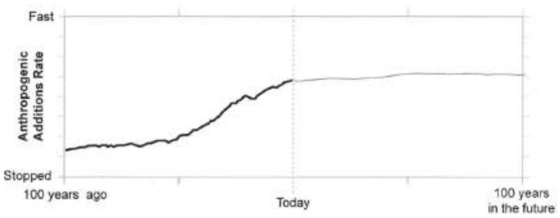
Two example graphical responses from each category are shown in Box 3. For the purposes of the analysis, these five categories were reduced to three categories: *Correct*, *Decrease*, and *Incorrect*, where the incorrect category combined *Stabilise*, *Increase*, and *Other.*

Written task. The written responses were coded according to whether the written explanation agreed with the graphical response. This approach is taken from Browne *et al* (2013) to investigate whether there was a discrepancy between the written and graphical representation.

The written responses were sorted in three categories: *Match-Strong*, *Match-Weak*, and *Mismatch.* Responses in the *Match-Strong* category described the behaviour in the graph correctly. Responses in the *Match-Weak* category described the behaviour in the graph partially, but not incorrectly. A common example would be a description that described anthropogenic emissions stopping completely, but the graph demonstrating a *decrease* trajectory; these descriptions had a graph that had a trajectory which followed the same direction. Responses in the *Mismatch* category, however, had a trajectory which followed the opposite direction than that described in the description; the descriptions did not match the graphical response. Where responses did not fit these criteria, they were categorised as *Other*. Results where there was no written or graphical response were marked as *NA*. The *Match-Strong* and *Match-Weak* responses are combined in the results as *Match*. Examples of these categorisations are show in Box 4.



Box 3: Typical graphical responses for each of the five categories

Example ‘match’ descriptions	
	Response #500106- correct graph To slow down and eventually it needs to stop
	Response #500614- increase graph The speed to pump the Anthropogenic Additions pump should be fast
Example ‘mismatch’ descriptions	
	*Response #500710 - increase graph* Slow down
	Response #500502 - stabilise graph The Anthropogenic Additions pump needs to slow down or really stop because otherwise it will keep going up and eventually exceed the goal level.

Box 4: Coding of graphical and written descriptions. A matching description does not necessarily mean that a response is correct, as shown in Response #500614. The mismatching description show confusion, most likely with the graphical representation as shown in Response #500710.

The combined coding of graphical and written responses means that there are six possible categories of interest in the analysis - *Correct*, *Decrease* and *Incorrect* graphs with either a description that in a *Match* or a *Mismatch*. This categorisation allows analysis of whether the participant was consistent in their description. The *Correct-Match* combination is desired, whereas an *Incorrect-Match* combination would demonstrate that the participant’s mental models didn’t match with the ideas within the activity. A goal of the workshop is to reduce the frequency of *Mismatch* responses, regardless of category, which would demonstrate a reduction of confusion between the graphical and written tasks.

Causal logic matching task. The causal logic task was straightforward to code. The causal logic task was used to further examine the mismatch between written and graphical responses, and triangulate whether the mismatch was due to incorrect causal logic or graphical misunderstanding. The correct responses are shown in Figure 4.

Q4. Consider the activity with the tubs and pumps. Circle the box you think is most correct in the table below to complete each sentence:

If the Anthropogenic Additions pump...	...the atmospheric carbon level will...			
continues to get faster	go down	stay level		go up
slows to a stop	go down	stay level	OR	go up
remains steady	go down	stay level		go up
slows slightly	go down	stay level		go up
gets faster, then remains steady	go down	stay level		go up

Figure 4: Correct responses for the causal logic matching task. *Slows to a stop* could have a correct response of *stay level* or *go up*, depending on the time scale that the pump slows to a stop. It was not uncommon to see a response where both were circled.

Correct responses for this task all involved the atmospheric carbon level increasing. In the *slows to a stop* scenario, responses that indicated the atmospheric carbon level remaining stable were also counted as correct, as it depends on the time scale that the participant is considering. Many responses had both circled, with an indication that the level would go up and then stay level. For the reporting of results, the number of correct responses in the causal logic for each participant was summed out of 5.

Results

The profile data are shown in Table 6. Most participants were under 22, female, identified in a STEM-related field, were early in their degree, and primarily communicated in English.

Table 6: Summary of profile data, with the frequency of correct graphs or matching descriptions. The total participants percentages show the category percentage as a proportion of total participants. The correct graph and matching description percentages are a proportion of participants in that category.

		total participants		**correct graph**		**matching description**		**correct causal logic**	
age	21≥	73	76.0%	34	46.6%	61	83.6%	47	64.4%
	22≤	23	24.0%	10	43.5%	16	69.6%	12	52.2%
	NA	0	-	-	-	-	-	-	-
sex	female	55	57.3%	23	41.8%	45	81.8%	35	63.6%
	male	41	42.7%	21	51.2%	32	78.0%	24	58.5%
	NA	0	-	-	-	-	-	-	-
field	STEM	69	71.9%	35	50.7%	57	82.6%	44	63.8%
	other	26	27.1%	9	34.6%	20	76.9%	14	53.8%
	NA	1	1.0%	0	-	0	-	1	-
degree progress	early	82	85.4%	38	46.3%	66	80.5%	50	61.0%
	later	14	14.6%	6	42.9%	11	78.6%	9	64.3%
	NA	0	-	-	-	-	-	-	-
language	English	80	83.3%	39	48.8%	64	80.0%	**54**	**67.5%**
	other	16	16.7%	5	31.3%	13	81.3%	**5**	**31.3%**
	NA	0	-	-	-	-	-	-	-
	Σ	96	100%	44	-	77	-	59	-

Note: bolded results show profile categories that are statistically significant indicators of results at a 90% confidence interval.

There are a number of noteworthy profile indicators for producing a correct graph. Male participants were 9.4%, students from STEM degrees were 16.1%, and native English speakers were 17.5% more likely to draw a correct graph than their category counterparts. Younger participants were 14% more likely to write a matching description than older participants. Younger participants were 12.2%, students from STEM degrees were 10.0% and native English speakers were 36.2% more likely to get all five causal logic correct. This last result is the only profile indicator that is statistically significant at a 90% confidence interval. The treatment groups were randomly allocated.

The treatment groups are listed in Table 7, along with the corresponding number of correct graphs, matching written responses and number of participants. The GDSC[1] treatment group produced the best results across all three tasks, by 25.9% for the graphical task and 13.9% for the written task. These result are not significant at a 90% confidence interval, but are an indication of improved understanding.

1. See Table 4 for treatment group descriptions

Table 7: Results by treatment group. Percentages show the percentage of responses for each treatment group.

Treatment	SC	GD	correct graph		written match		correct causal logic		n
Base	No	No	9	50.0%	12	66.7%	11	61.1%	18
GD	No	Yes	5	20.8%	19	79.2%	13	54.2%	24
SC	Yes	No	8	32.0%	19	76.0%	16	64.0%	25
GDSC	Yes	Yes	22	75.9%	27	93.1%	19	65.5%	29

The results from Table 7 are also shown in Figure 5, with 90% confidence intervals. The group diagram and scenario cards by themselves did not improve the likelihood of drawing a correct graph; however, when used in conjunction produced the best results. The group diagram and scenario card treatments produced better results in the written description than the results in the base condition. The scenario card treatment groups produced marginally better causal logic matching results than those that did not use scenario cards.

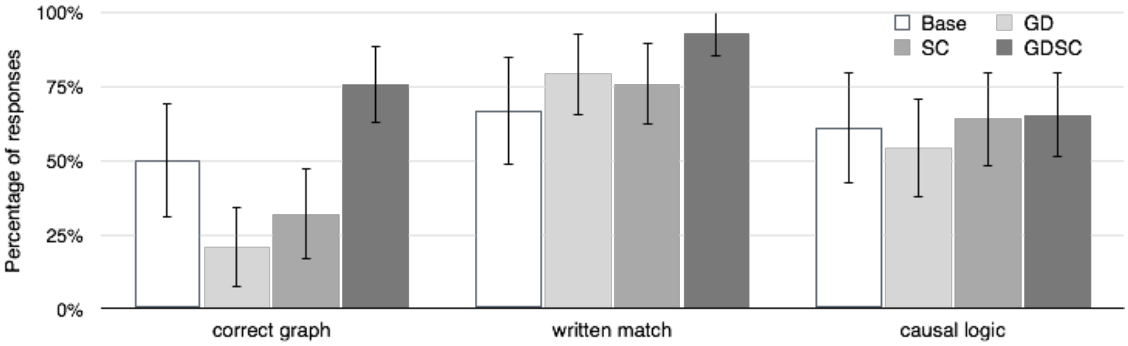


Figure 5: Results graphed by treatment group. Percentages show the percentage of responses for each treatment group.

Some interesting observations can be made when the graphical and written responses are considered in combination. All correct graphs have a matching description; that is, when a participant draws a graph that tends towards zero, then their description matches the graph. However, participants that draw an incorrect graph often have a mismatching description, similar to the mismatching examples in Box 4. The written description typically describes a reduction in anthropogenic additions, but this is not shown in the corresponding graph.

The improvement of mental models should be demonstrated both by the increased frequency of correct graphs, but also a reduction of incorrect graphs with mismatching descriptions. Table 8 shows the frequency of responses for each graphical and written combination. No treatment groups showed any *Mismatch-Incorrect* responses, indicating that all correct graphs were supported with a matching written statement. The

GDSC treatment group did not have any *Match-Incorrect* responses, which would demonstrate a coherent yet incorrect graph with a supporting description. The SC treatment had the only instances of the *Mismatch-Decrease* combination.

Table 8: Results by treatment group for graphical and written response categories. Percentages show the percentage of responses in each treatment group.

Treatment	Written Response	Graphical Response					
		Correct		Decrease		Incorrect	
Base	**Match**	9	50.0%	2	11.1%	1	5.6%
	Mismatch	0	0.0%	0	0.0%	6	33.3%
GD	**Match**	5	20.8%	11	45.8%	3	12.5%
	Mismatch	0	0.0%	0	0.0%	5	20.8%
SC	**Match**	8	32.0%	10	40.0%	1	4.0%
	Mismatch	0	0.0%	3	12.0%	3	12.0%
GDSC	**Match**	22	75.9%	5	17.2%	0	0.0%
	Mismatch	0	0.0%	0	0.0%	2	6.9%

The *Match-Decrease* category in both the GD and SC treatments had the highest proportion of respondents for their respective treatments, rather than the *Match-Correct* category seen in the other treatments. This is not necessarily a demonstration of misunderstanding, and is explored further in the discussion.

The causal logic matching task was used to further investigate the understanding of participants in the activity. Participants were asked to provide the correct response to five causal logic propositions about the system. The average causal logic scores for participants in each category are shown in Table 9. Higher scores tend towards the *Match-Correct* category, except in the Base and SC treatments, which also had higher scores in the *Mismatch-Incorrect* category, supporting an argument that the *Mismatch-Incorrect* category demonstrates graphical confusion rather than intentional responses.

Table 9: Average of causal logic scores for graphical and written response categories. Categories with no responses are shown with a dash. Numbers of each category can be seen in Table 8.

Treatment	Written Response	Graphical Response		
		Correct	Decrease	Incorrect
Base	**Match**	4.0	4.0	3.0
	Mismatch	-	-	4.5
GD	**Match**	4.6	4.2	2.7
	Mismatch	-	-	3.2
SC	**Match**	5.0	3.5	3.0
	Mismatch	-	3.7	5.0
GDSC	**Match**	4.5	3.2	-
	Mismatch	-	-	2.0

Discussion and observations

The approach discussed in this paper are a response to John Sterman's call for new ways to improve people's intuitive systems thinking capabilities.

> We need new methods for people to develop their intuitive systems thinking capabilities. Bathtub analogies and interactive "management flight simulators" through which people can discover, for themselves, the dynamics of accumulation and impact of policies have proven effective in other settings and may help here. (Sterman 2008, p533)

We have created a concrete learning activity to help people with no systems thinking training develop better mental models around the concept of accumulation. We tested four treatment groups based around the same hands-on activity. We found that the GDSC treatment produced the best results, suggesting that scaffolding the hands-on activity with effective prompts that deliberately encourage focused group discussion was a successful strategy for facilitating double-loop learning.

The result shown in Figure 5 for the SC and GD treatment groups is somewhat surprising. Both treatments had fewer correct graphical responses than the Base treatment. However, the *Match-Decrease*[2] category showed the highest frequency of responses, suggesting that participants in this category had a good idea of the required action, but did not draw a graph brought the future trajectory close enough to zero to be categorised as correct. Our observations through the activity suggest that a likely reason for this is that participants bring their assumptions about the real world to the activity; following the recent history of action on climate change makes it difficult to imagine a zero-emissions world. Table 10 shows a summary of the data from Table 8 with the *Match-Correct* and *Match-Decrease* categories combined, alongside the *Mismatch-Incorrect* category.

Table 10: Summary results by treatment group for graphical and written response categories. Percentages show the percentage of responses for each treatment group, and do not necessarily not add up to 100 due to the exclusion of categories not shown.

Treatment	Correct or Decrease Graph and Written Match		Incorrect Graph and Written Mismatch	
Base	11	61.1%	6	33.3%
GD	16	66.6%	5	20.8%
SC	8	72.0%	3	12.0%
GDSC	22	93.1%	2	6.9%

2. Rather than the *Match-Correct category*

The results in Table 10 demonstrate some remarkable outcomes. First is that over 90% of participants in the GDSC treatment group provided a graphical response supported by a written statement that corresponds to a future with reduced anthropogenic additions of carbon to the atmosphere. This is the very future that we need on our shrinking planet.

It could be argued, however, that students in an environmental science course would be more likely to have attitudes towards a zero-carbon world. In our earlier study (see Browne *et al* 2013), the environmental science students demonstrated very poor results performing the worst out of all groups on the graphical and written tasks. Using the same hands-on activity with different instructions, students in the same class in 2013, only 11.3% of responses were *Match-Correct* and 28.6% were *Match-Decrease* whilst 45.1% of responses were *Mismatch-Incorrect*. This, together with the range of responses in the different treatment groups in study, suggests that the scaffolding around an activity plays a significant role in improving participants' understanding.

It could also be argued that GD and SC activities were either spoon-feeding or cheating. This perspective would miss the point. If framed in the prevailing educational paradigm of competitive, exam-based evaluation of learning, then, yes, working together to get to a solution in the group diagram or making the answers visible through the scenario cards could be seen as biasing the results. The rationale for our approach in addressing the accumulation problem from a learning perspective has been articulated in earlier sections. We argue that the learning in this experiment has occurred without direct intervention from the instructor, that mental models have been developed socially, and that the knowledge has been constructed in each individual's head.

Further, we believe that the relatively positive results using the hands-on activity are not the major contribution of this work. If our work has appealed to your creative, playful nature, and you plan to invest in a class set of tubs and handheld pumps, then we urge you to understand one thing. We believe that the major contribution is not the concrete activity itself, but rather that we have been able to complement it with a pedagogical approach that unlocks its effectiveness and tethers it to the real world. We would expect to see similar results if we effectively used tools from the existing systems-thinking toolbox, such as simulation models or flight simulators.

The importance of creating an effective learning environment is shown simply by looking at the range of *Correct* responses, using either the categorisation of results in Table 8 or Table 10. The use of prompts around the activity has a large effect on the results. The use of a hands-on activity itself is not enough for the participant to assimilate knowledge, and that a sequence of small challenges or tasks provide a useful avenue for exploration and improving mental models.

It should be noted that if one were to compare these results to the studies described at the beginning of this paper, she would be comparing two different things. Our objective was not to test understanding, but construct it. It would be a mistake to rush out and assemble T&P systems for every dynamical modelling exercise. We have approached a single problem—to improve participants' intuition around atmospheric accumulation of carbon—and have tried to do this in a creative way that observes the best practices of learning and developing mental models through a shared, concrete experience. We believe that this activity provides a good platform for systems thinkers to then explore more sophisticated models, such as through computer simulation.

Challenges for systems thinking educators

We ran an activity with 96 environmental science students that involved a hands-on dynamical model of a simplified carbon cycle. Students were divided into four treatment groups that used a combination of prompts to navigate the exercise. We found that prompts that encouraged focussed and informed dialogue produced the best individual results. The combination of scenario cards and completing the graphical response as a group gives promising results, with over three-quarters of participants able to draw a correct graph with a matching written description.

We approached the challenge of improving participants intuition of accumulation through designing an interactive, hands-on workshop. We believe the T&P activity is a useful tool for concrete learning about simple dynamic systems. However, the design of the workshop—the learning environment—appears to be a more significant factor than the T&P activity itself, as shown in the range of correct graphical response categories, from approximately 20% correct responses to 75% correct responses all with the same T&P equipment. Our results show that this activity provides a novel method of teaching introductory systems-thinking concepts.

This leaves some key considerations and challenges for further research. We have demonstrated that we can have a positive effect on a relatively small cohort of students. Further work is required to understand whether this activity has any transferability to an individual's decision-making process outside of the activity. The T&P activity has the building blocks to be applied to many dynamical problems, not just the carbon cycle. Further work is required to identify and test its effectiveness to other problems, and with participants from other demographic groups, such as primary school students. We argue that the concrete activity itself is less important than the social and creative learning environment it is embedded in. Further work is required to help educators construct effective learning environments for students to become better systems thinkers, and in turn better decision makers.

Our advice for educators working in this area is that careful attention needs to be paid to building powerful, shared understanding if your goal is to help learners transfer their knowledge about abstract concepts to the real world.

Acknowledgements

I (Chris) would like to acknowledge the contribution of Barry Newell in the conceptualisation of the tubs and pumps system, and the many and ongoing discussions, comments and thoughtful critiques in improving my own understanding of how knowledge is constructed. I would also like to thank Tracy Benson, Diana Fisher, Anne LaVigne, and Lees Stuntz for our engaging discussions about creating better systems thinkers.

Bibliography

Andersen, D. & Richardson, G. P., 1997, Scripts for group model building, *System Dynamics Review*, 13(2), pp. 107-29.

Andersen, D. F., Vennix, J. A. M., Richardson, G. P., & Rouwette, E., 2007, Group model building: problem structuring, policy simulation and decision support. *The Journal of the Operational Research Society*, 58 (5), 691–694

Argyris, C., 1976, Single-loop and double-loop models in research on decision making, *Administrative science quarterly*, pp. 363-75.

Barker, T., Davidson, O., Davidson, W., Huq, S., Karoly, D., Kattsov, V., Liu, J., Lohmann, U., Manning, M. & Matsuno, T., 2007, *Climate change 2007: Synthesis report*, Valencia; IPPC.

Booth Sweeney, L. & Sterman, J.D., 2000, Bathtub dynamics: initial results of a systems thinking inventory, *System Dynamics Review*, 16(4), pp. 249-86.

Brockhaus, F., Arnold, J., Schwarz, M. & Sedlmeier, P., 2013, International System Dynamics Conference, Does the modification of the representation format affect stock-flow-thinking?.

Browne, C., Newell, B. & Compston, P., 2013, Proceedings of the 21st Annual Conference for the Australasian Association for Engineering Education, Developing understanding of the carbon cycle through physical analogues..

Cronin, M., & Gonzalez, C., 2007, Understanding the building blocks of system dynamics. *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), pp. 116-30.

Forrester, J. W. (2009). Some basic concepts in system dynamics. Sloan School of Management. Massachusetts Institute of Technology. Cambridge, MA.

Hovmand, P.S., 2014, *Community Based System Dynamics*, Springer, .

Kunzig, R., The Carbon Bathtub, *National Geographic*, December 2009

Lakoff, G. & Johnson, M., 1980, *Metaphors We Live By*, University of Chicago Press,.

Maani, K.E. & Cavana, R.Y., 2007, *Systems thinking, system dynamics: managing change and complexity*, Prentice Hall, Auckland, N.Z..

Martinez, S.L. & Stager, G., 2013, *Invent to learn: Making, tinkering, and engineering in the classroom*, Constructing Modern Knowledge Press, Torrance, CA.

Maxwell, T., D. F. Andersen, G. P. Richardson and T. R. Stewart., 1994. Mental Models and Dynamic Decision Making in a Simulation of Welfare Reform. *Proceedings of the 1994 International System Dynamics Conference, July 1994*, Stirling, Scotland, 11–28. Vol: Social and Public Policy, Albany, NY: System Dynamics Society.

Meadows, D., 2008, *Thinking in systems*, Wright, D. ed. Chelsea Green Publishing,.

Newell, B., 2012, Simple models, powerful ideas: Towards effective integrative practice, *Global Environmental Change*, 22(3), pp. 776-83.

Newell, B., and Proust, K., 2012, Introduction to Collaborative Conceptual Modelling, Working Paper, ANU Open Access Research. https://digitalcollections.anu.edu.au/handle/1885/9386

Papert, S., 1980, *Mindstorms: Children, computers, and powerful ideas*, Basic Books, Inc.

Richmond, B., 2010, *Tracing Connections*, Richmond, J., Stuntz, L., Richmond, K. & Egner, J. eds. isee systems, incorporated,.

Rogers, E., 2003: *The Diffusion of Innovations*. Fifth Edition. The Free Press, New York

Sedlmeier, P., Brockhaus, F. & Schwarz, M., 2014, *Using Tools for Learning Mathematics and Statistics*, Springer,.

Senge, P.M., 1997, *The Fifth Discipline Fieldbook*, Random House Digital, Inc.

Sterman, J.D. & Booth Sweeney, L., 2002, Cloudy skies: assessing public understanding of global warming, *System Dynamics Review*, 18(2), pp. 207-40.

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), pp. 213-38.

Sterman, J.D., 2000, *Business dynamics*, Irwin-McGraw-Hill.

Sterman, J.D., 2008, Risk communication on climate: mental models and mass balance, *Science*, 322(5901), pp. 532-3.

Sterman, J.D., 2010, Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review*, 26(4), pp. 316-34.

Stocker, T.F., Dahe, Q. & Plattner, G.-K., 2013, Climate Change 2013: The Physical Science Basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013).

Vennix, J., 1996, *Group model building*. New York: Wiley.