A DOUBLE-LOOP LEARNING APPROACH TO CONSTRUCT UNDERSTANDING OF ACCUMULATION PRINCIPLES

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY OF
THE AUSTRALIAN NATIONAL UNIVERSITY

© CHRISTOPHER ANTHONY BROWNE 2015
ALL RIGHTS RESERVED

I declare that the work in this thesis is entirely my own and that to the best of my knowledge it does not contain any materials previously published or written by another person except where otherwise indicated.

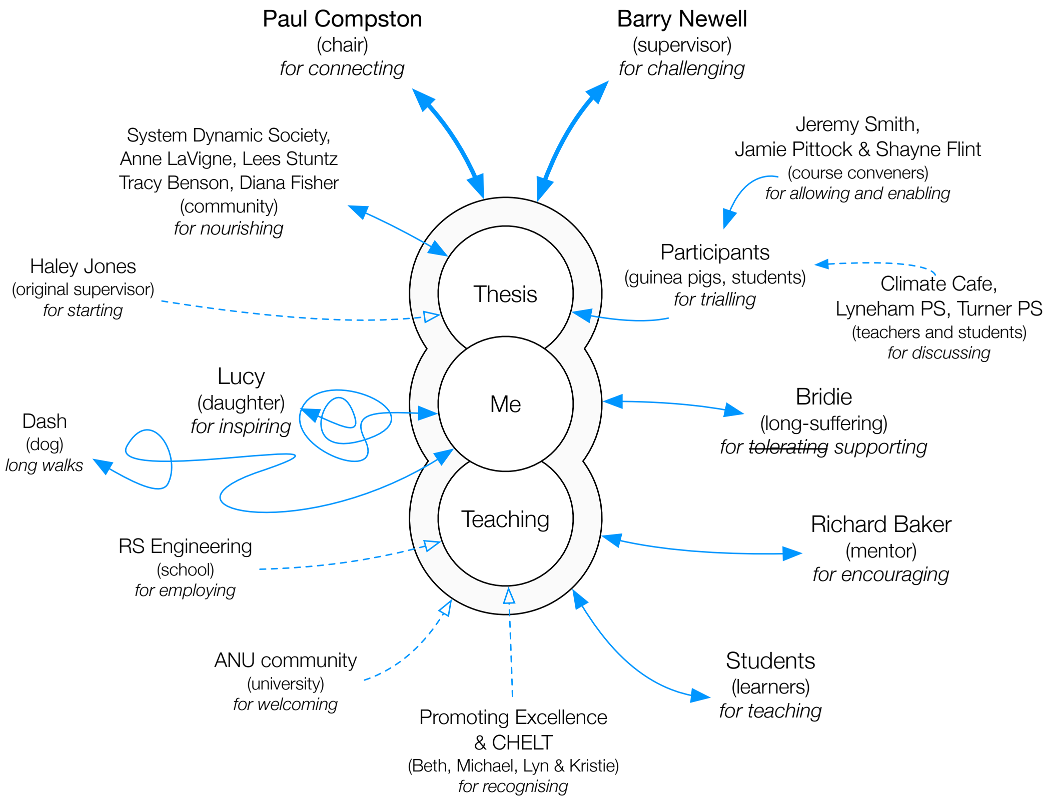
Chris Browne
December 2015

For Lucy,

and the impending generation of Tinker-bells

who will show the Lost Boys how

Acknowledgements



Thank you also to the examiners for their encouraging and thoughtful discussions of both the value and importance of this work.

Abstract

In this thesis, I present a concrete learning activity to assist engineering and STEM students with no formal systems-thinking training develop improved mental models of accumulation principles. This thesis 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.

At the core of this research is a model for double-loop learning through constructionist inquiry. The scenario for the activity is the effect of anthropogenic carbon emissions on the atmospheric carbon concentration. A hands-on activity was developed called Tubs & Pumps (T&P) as a physical analogue of the carbon cycle. However, the activity could be adapted to a range of dynamic problems.

Students manipulate the T&P system guided by a series of prompts, which encourage focused and informed group discussion about the given problem. A range of treatment conditions were used to investigate the effect of prompts and assessment layout in the experiment. The results show that using targeted prompts can drastically improve the likelihood of students demonstrating a sound understanding of accumulation principles.

This finding has implications for how knowledge is constructed in engineering classrooms, and is a valuable approach for educators wanting their students to develop a deep understanding about dynamic systems.

Table of contents

List of figures

List of tables

List of boxes

Glossary

Key terminology

accumulation principles	valid stock-and-flow behaviour
activity	a set of protocols for interaction with the T&P system
carbon cycle	the biogeochemical cycle describing the exchange of carbon between stocks in the Earth system
CS	cognitive science, field of knowledge
double-loop learning	a model for learning based on the continuous refinement of mental models based on observations in the real world
experiment	a set of protocols to describe design of workshops used for data collection
flow	processes that change a stock
L&T	learning & teaching, field of knowledge
mental model	cause-and-effect logic used in one's decision-making
physical simulation	a process of simulation using a physical system; in this instance, the T&P system
response	the answers provided in the assessment sheet
SF failure	stock-flow failure, a violation of accumulation principles
ST	systems thinking, field of knowledge
STEM	science, technology, engineering and mathematics
stock	an entity that accumulates or depletes over time
stock-flow (SF) failure	a perception of stock-and-flow relationships that violate conservation of mass principles
system dynamics	a methodology for describing complex physical and social systems
T&P	tubs and pumps, the activity at the core of this thesis
T&P system	a physical, manipulable model used for simulation
WG confusion	written-graphical confusion, where the graphical and written responses do not match
workshop	a session where students engaged with the activity

Terminology for workshop activities

Note: Additional terminology for the workshop methodology is in Part II.

Experiments

Checklist	the early experiments based on checklist instructions
Simulation	later experiment to test different simulation activities
Dialogue	later experiment to test focussed dialogue prompts

Experimental categories

PC	physical simulation activity with checklist instructions
PW	physical simulation activity with workbook instructions
MW	mental simulation activity with workbook instructions

Assessment sheets

ER	Emissions & Removals
AE	Anthropogenic Emissions
AA	Anthropogenic Additions

Treatment categories

Treatment categories are referred to using the acronym 'experimental category–assessment sheet', for example, PW-AA. Additional categories include:

PW-GD	as PW with group diagram treatment condition
PW-SC	as PW with scenario card treatment condition
PW-GDSC	as PW-GD and PW-SC combined

Graphical-written categories

Summary groups that categorise results for both the graphical and written tasks

Sound AP	sound accumulation principles (AP); graphs that show a decrease, with a matching written description
WG confusion	written-graphical (WG) confusion; where the graph does not match the written description, such as a graph that increases with a description that calls for a decrease
SF failure	stock-flow (SF) failure; where the graph has a matching description, but violates accumulation principles, such as a belief that increasing the flow rate into a stock will result in a decrease of that stock

1 Introduction

Generation by generation universities serve to make students think: [to] learn progressively to identify problems for themselves and to resolve them by rational argument supported by evidence; [to] learn not to be dismayed by complexity but to be capable and daring in unravelling it.

- Boulton and Lucas (2008, p.9)

1.1 Context of the study

Engineers are first and foremost problem solvers. Each engineering specialisation has within it an abstract body of knowledge that makes sense of real-world observations. For example, civil engineers use force equations to optimise the design of infrastructure; electronics engineers simulate circuitry to debug, prototype and create the next generation of devices. There will always remain the need for these skilled problem solvers; however, society will need future engineers to work in truly interdisciplinary teams to solve problems that are increasingly complex, dynamical in nature, embedded in systems of systems, and unrecognised by today's engineering educators.

The future engineer needs to be proficient with a broad range of problem-solving skills in order to approach these complex problems, in addition to the specialist knowledge currently taught in undergraduate engineering programs (King, 2008). The student engineer needs more than technical knowledge; they need room to develop into creative, capable and convincing problem solvers. Students must be able to construct their own knowledge and develop their intuition about problems based on rich experiences, beyond textbook-based instruction. To achieve this, engineering educators need to take up Boulton & Lucas' (2008) challenge to create educational environments where students can be bold in unravelling complex problems. We need to create learning environments that help students understand how to think, so that they can be active learners in their future problem-solving activities.

Perhaps the most important perspective a problem solver can have is the understanding that their mental models—their intuitive understanding of the real world—will always be incomplete. Double-loop learning (Argyris, 1976) is required to continually refine mental models based on feedback from the real world. Second-generation cognitive scientists have shown that human understanding of the world is based on experience, and is largely metaphorical. Modern metaphor theory (Lakoff & Johnson, 2003) can provide strong guidance to those attempting to develop effective teams of problem solvers by building on learners' experiences with the real world.

Newell (2012) argues that the careful design of simple, manipulable activities is a basis for building powerful dynamical metaphors for the development of shared conceptual repertoires. For example, using Cuisenaire rods to understand mathematical operations, or interacting with low-order models to understand dynamic behaviour. These activities allow participants to communicate more effectively and, hence, more likely to create informed and shared understanding on a topic. These hands-on physical experiences allow learners to effectively think through a problem, visualise past and future behaviour, learn from the experience, and in turn develop their own mental models.

Creating a learning environment for both active learning and higher-order thinking is not straightforward, nor commonplace, especially in the context of the changing landscape of higher education in Australia. Technology is becoming more prevalent as a method of delivering content on an increasing scale, shown in the rapid expansion of Massive Open Online Courses (MOOCs) and other online teaching models. In the face of these trends, opportunities for effective concrete, hands-on learning are increasingly important and increasingly rare.

I have developed a hands-on learning activity, called Tubs & Pumps (T&P). The context for my development of the T&P activity has been described as the greatest moral challenge of our time: *climate change*. Sterman & Booth Sweeney (2002; 2007) discovered that educated adults' understanding of the dynamics of climate change is poor, and that cause-and-effect intuition of the carbon cycle often violates conservation of mass principles. This is known as stock-flow (SF) failure.

> People of good faith can debate the costs and benefits of policies to mitigate climate change, but policy should not be based on mental models that violate the most fundamental physical principles. The results suggest the scientific community should devote greater resources to developing public understanding of these principles to provide a sound basis for assessment of climate policy proposals.
>
> \- Sterman & Booth Sweeney (2007 p.236)

Further, education in science, engineering or mathematics does not appear to improve these test results (Sterman 2008). There is a need to re-think the prevailing passive and disconnected approach to STEM (science, technology, engineering and maths) education—from kindergarten through to graduate education—in the face of increasingly complex problems, such as climate change.

This thesis is a record of an attempt to improve undergraduate STEM students' understanding of stocks and flows in the carbon cycle, although the principles and approach has application at all levels of education. Intuition is measured through the participant's causal insight into the future behaviour of the carbon system. I do this through adopting a double-loop learning approach, where students can think, visualise, and improve their understanding through a physical simulation experience from which the learner can draw his or her own conclusions.

1.2 Problem statement

As the Earth's climate changes, today's student engineers will be working in an increasingly complex and uncertain world. Sterman & Booth Sweeney (2002; 2007) argue that the public's poor understanding of climate change mechanisms is in part due to a poor understanding of stock-and-flow dynamics. This assertion holds even for participants who have STEM backgrounds. Sterman (2010) calls for research into more effective methods to teach dynamics to improve people's intuitive systems-thinking abilities. This thesis addresses this problem directly.

1.3 Aim and scope of the thesis

The aim of this thesis is to improve how engineering educators understand of the use of concrete experiences in the learning process. This work builds on the established insights of Argyris's single- and double-loop learning, where effective learning is demonstrated through the development of mental models. In Figure 1.1a, a single-loop learning model is shown, whereby decisions are made based on feedback from the real world. However, it is only in Figure 1.1b that decision-making is improved through the development of mental models. In this thesis, the use of hands-on activities and prompts are investigated as a fundamental part of double-loop learning.

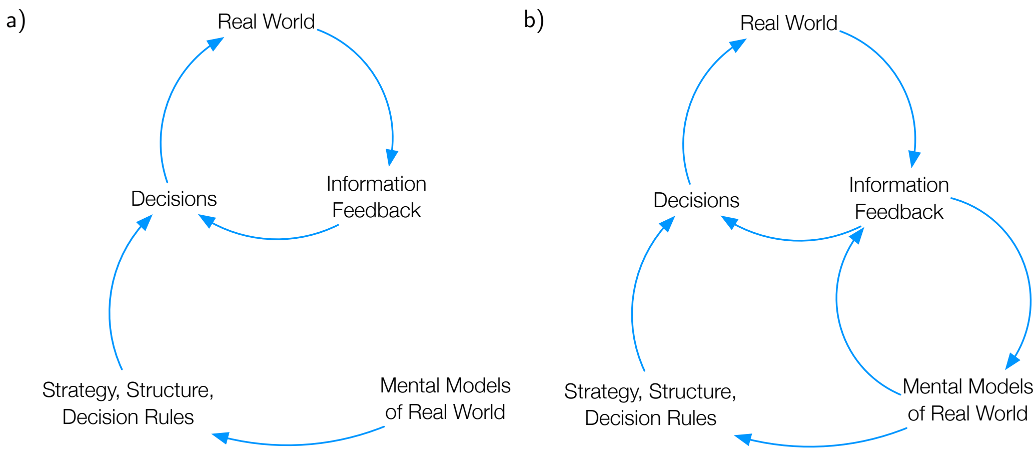


Figure 1.1: Learning loops describing how decisions are made based on observations of the real world from Sterman (2000). a) single-loop learning, where decisions are made without the development of mental models. b) double-loop learning, which requires the evolution of mental models based on experience with the real world. Variables at the head of an arrow are affected by the variables at the tail of the arrows.

The challenge of stabilising atmospheric carbon levels is the specific problem context of learning for this study. The research is focused on the design and evaluation of an effective learning experience, through both a physical simulation activity, and the design of the workshop it is used in. The research is limited to participants at the undergraduate level, and largely those in STEM programs, particularly engineering. Undergraduate STEM students were chosen as the focus group, as they should have the technical ability to understand the mathematical foundation of the assessment task. However, the activity is designed to be accessible to the broader public, especially in an educational context.

1.4 Research questions

The model in Figure 1.1 is developed into a double-loop learning model for physical simulation in Figure 3.12. The loops in Figure 1.1 are named, with the top loop describing *physical simulation*, the bottom right loop describing *scaffolded refinement*, and the bottom left loop describing *focussed dialogue.* From this, three research questions arise, leading to three partially overlapping experiments:

> Experiment 1: Scaffolded refinement
> *Does display of instructions and information have an effect on decision-making?*
>
> Experiment 2: Physical simulation
> *Does representational form of simulation have an effect on decision-making?*
>
> Experiment 3: Focussed dialogue
> *Do opportunities for focussed dialogue have an effect on decision-making?*

1.5 Significance and contribution

The research generates two outcomes of relevance in important areas of learning and teaching in engineering.

The first outcome is the T&P system itself. It is a tool for hands-on learning that systems educators can adopt and use to explore fundamental dynamical concepts. Although the computer has enabled students to simulate dynamical systems, basic concepts such as the ephemeral nature of flows can be easily misunderstood by the learner. This leads to conceptual modelling issues, such as SF failure through confusion between stocks and flows. The T&P activity is a hands-on simulation environment that can reduce this confusion, which makes it a worthy tool in the suite of learning activities for systems educators.

The second outcome is a model for double-loop learning for physical simulation. This model demonstrates the importance of careful design and effective prompts in concrete learning through the T&P activity. Over the series of experiments, the T&P system itself remained relatively unchanged. However, the language used to describe the T&P activity and the prompts that participants used were changed. These subtle changes made a significant difference to the way that students completed the exercise, with correct responses ranging from 11% to 76% depending on the prompts used. With the use of effective instructions, I have been able to complement the T&P activity with a pedagogical approach that uses prompts and social learning to improve the effectiveness of the activity.

The approach for scaffolding of new concepts for the learner through a concrete activity is relevant to engineering educators as they continue to challenge students to become problem-solvers today and into the future.

1.6 Structure of the thesis

This thesis consists of seven chapters, organised into three parts.

- Part I. Background
- Part II. Research Approach
- Part III. Synthesis

Part I. Background (Chapters 2 and 3). I introduce the study, and situate it within the relevant literature. In Chapter 2, I outline a stock-and-flow representation of the carbon cycle. Then I discuss the confusion observed in stock-and-flow thinking, in particular the SF failure around the anthropogenic disruption to the carbon cycle.

In Chapter 3, I discuss three overlapping perspectives for constructing knowledge: systems thinking, cognitive science and learning & teaching. These theoretical frameworks provide the background for the design of the T&P activity, and emphasise the importance of active learning environments and situating learning in the physical world. These concepts are of critical importance to student engineers, especially as they attempt to navigate complex and open-ended problems.

In ***Part II. Research Approach*** (Chapters 4 and 5), I describe the research methodology. In Chapter 4, I present the T&P system as a simplified model of the carbon cycle. I describe the experimental methodology through the design of the T&P activity, the treatment conditions, assessment tasks, and the coding of responses. In Chapter 5, I present summary results, and detailed results for each experiment and for each task by treatment condition.

In ***Part III. Synthesis*** (Chapters 6 and 7), I take the lessons from the current study and apply them to engineering education. In Chapter 6, the results of the workshops are analysed, discussed and expanded in relation to opportunities for enhancing undergraduate engineering education approaches. In Chapter 7, I discuss the conclusions and areas of further work arising from this study.

PART I BACKGROUND

Preface

Approaches for constructing shared knowledge of dynamical systems is the focus of Part I. The emphasis on shared knowledge is due to the fact that the problems future engineers will confront are likely to be more complex than an individual can solve alone.

Simulation models are a useful for engineers to expose their thinking about a problem before making changes to complex real-world systems. In this thesis, I investigate the use of a physical simulation activity to improve students' thinking about the real-world carbon cycle, an important component of the dynamic Earth system.

Human activity is changing the Earth system to the extent that we have entered a new geological epoch: the Anthropocene (Crutzen, 2002). It is becoming widely recognised, and acknowledges a new geological period where human activities have a significant global impact on the Earth's ecosystems:

> Human influence on the climate system is clear, and recent anthropogenic emissions of greenhouse gases are the highest in history. Recent climate changes have had widespread impacts on human and natural systems.
>
> - Stocker et al (2013, p. 15)

The scientific evidence on the anthropogenic disruption to the carbon cycle is clear. However, research shows that the dynamics of the carbon cycle are widely misunderstood, particularly the interaction between the rate of anthropogenic carbon emissions and the resultant stock of carbon in the atmosphere. Even well-educated adults often confuse stocks and flows (Sterman & Booth Sweeney, 2002; 2007).

In this thesis, I investigate whether the use of a physical simulation activity (the T&P activity) can provide a basis for adults to better understand the cause-and-effect logic in the carbon cycle. The goal of the activity is to allow participants to correctly infer the required future trajectory of anthropogenic carbon emissions in order to stabilise the atmospheric carbon stock.

Part I is organised across two main themes:

Chapter 2: A discussion of the carbon cycle: the real-world context of the T&P activity. This includes a stock-and-flow representation of the carbon cycle (§2.1) and the previous work that recognised stock-flow (SF) failure (§2.2), specifically SF failure around the carbon cycle.

Chapter 3: The rationale behind the T&P activity: the physical simulation activity used to help participants improve their understanding of the carbon cycle. This includes the background for constructing shared knowledge from three perspectives: systems thinking (§3.1), cognitive science (§3.2), and learning & teaching (§3.3).

2 Stock-and-flow thinking and the carbon cycle

To suggest that we're tackling climate change by building new mines and new power stations is simply absurd. I mean, it's like someone who's drunk at the end of the night at a party saying, "Look, I've switched to low-alcohol beer now. It's OK".

- Richard Denniss (ABC RN Breakfast, 2015)

A fundamental problem for system dynamics educators is helping students to distinguish between stocks and flows. Recently, many studies have highlighted stock-flow (SF) failure in students' mental models of accumulation, in both linear and non-linear systems (Booth Sweeney & Sterman, 2000; Sterman & Booth Sweeney 2002; 2007; Cronin & Gonzalez, 2007; Cronin et al, 2009; Brockhaus et al, 2013; Sedlmeier et al, 2014; Kapmeier, 2004; Kapmeier et al, 2015). This confusion occurs when there is an inconsistency between the perceived behaviour of stocks (an entity that accumulates or depletes over time) when the flows (the rate at which that entity changes) are given in a described system.

Of particular concern is SF failure in relation to the anthropogenic additions to the atmospheric carbon concentration. At the 2007 National Climate Change Summit, then Opposition leader Kevin Rudd proclaimed "Climate change is the great moral challenge of our generation" (Kelly, 2007). However, studies have shown that public understanding of the dynamics of climate change is poor amongst even well-educated adults (Sterman & Booth Sweeney 2002; 2007).

In this chapter, I describe the dynamics of the carbon cycle in stocks and flows. I then describe the current arguments in the area of SF failure and testing of accumulation principles. Then I introduce the previous work investigating specifically SF failure in relation to the carbon cycle. This leads to Chapter 3, where I describe the background theory for T&P activity.

2.1 Dynamics of the carbon cycle

The carbon cycle is one part of the complex and changing Earth system. Rockström et al (2009) argue that the Earth system has been pushed into a higher-risk area with respect to a number of planetary boundaries, including levels of biodiversity loss, the nitrogen cycle and climate change. The 2015 update of their work (Steffen et al, 2015) add the phosphorous cycle and changing land-use as higher-risk systems. These systems at higher-risk increase the uncertainty of the functioning and resilience of the Earth system.

The effects of this change on the Earth system are many. In the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), the changes that are expected to occur towards the late 21st century are described as:

> ***Virtually certain***
> - warmer and/or fewer cold days and nights over most land areas
> - warmer and/or more frequent hot days and nights over most land areas
>
> ***Very likely***
> - warm spells/heat waves. Frequency/duration increases over most land areas
> - heavy precipitation events. Increase in the frequency/intensity/amount
> - increased incidence and/or magnitude of extreme high sea level
>
> ***Likely***
> - increases in intensity and/or duration of drought
>
> \- Stocker et al (2013)

There are a number of indicators that are used to measure the state of the climate system at various locations around the globe. These include land and ocean temperatures, precipitation rates, land-mass snow cover, extent of Arctic ice, ocean heat content, sea level change, the operation of the carbon cycle and other biogeochemical cycles. In this thesis, the carbon cycle is the only system of interest, specifically the anthropogenic disruption to the carbon cycle due to burning fossil fuels.

2.1.1 The carbon cycle in stocks and flows

In system dynamics terminology, a 'stock' is a state variable, such as the amount of money in a bank account, the number of people in a department store, the number of trees in a forest, the value of an inventory, the amount of energy produced by a solar panel, or the reputation of a politician. The level of the stock is controlled by the 'flows' which add to or remove from it. For example, the amount of money in a bank account is controlled by the flow of deposits and withdrawals; the number of people in a department store at any one time is determined by the number of people entering and leaving the store.

The major stocks and flows, as identified and quantified in the IPCC's Fourth Assessment Report, are shown in Figure 2.1. The diagram shows the natural carbon stocks and flows and the anthropogenic perturbation of these stocks and flows. The quantities shown are as estimated in the mid-1990s.

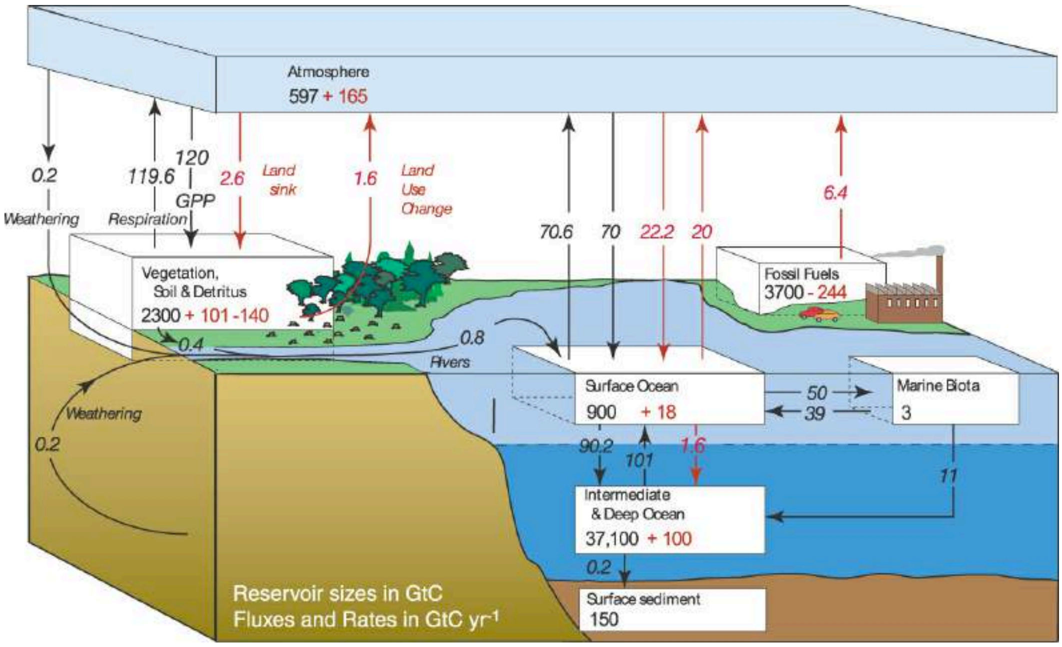


Figure 2.1: Box diagram of Earth's carbon cycle (Barker et al, 2007 Figure 7.3). Stocks (reservoirs) are shown as rectangular blocks; flows (fluxes) are indicated by arrows. Black numbers and arrows represent the natural cycle prior to anthropogenic influence. Red arrows and numbers indicate the anthropogenic perturbation.

Figure 2.1, although a simplification, shows the major flows of carbon around the system. The unit used to describe the stocks is gigatons of carbon (GtC) and the unit for flows is gigatons of carbon per year (GtC/year). The atmospheric CO_2 concentration is often reported using the unit parts per million by volume (ppmv). Ppmv of CO_2 can be converted to GtC using the ratio 1ppm (atmospheric CO_2) : 2.13 (GtC).

The diagram shows a system that is largely in balance, with additions to and removals from each stock largely cancelling out. There are, however, small imbalances. Consider the flows directly connected to the atmospheric stock as seen in Figure 2.1. The natural cycle is in balance, with a circulation of approximately 190.2 GtC/year. The anthropogenic disturbance upsets this balance, with a net amount of 3.2 GtC/year added to the atmosphere, mainly due to the 6.4 GtC/year added through the burning of fossil fuels.

Further, the rate of flows is changing, with the increased rate of burning fossil fuels. Feedback behaviour will also change the relationships in the carbon cycle; for example, as the global surface temperature increases, the temperature of the ocean increases, and its capacity to absorb carbon reduces.

In stock-and-flow notation, stocks are represented by rectangular boxes. An inflow is represented by an arrow pointing into a stock, and an outflow is represented by an arrow pointing out of a stock. The flows are represented by valves (taps), and clouds represent sources or sinks that are outside the system of interest. Variables are connected via influence links (arrows), which determine the nature of the relationships between the variables. Mathematical equations sit behind this graphical representation, and determine the relationships between the variables in a stock-and-flow diagram, allowing the model to be simulated over time.

Sterman (2000, p. 246 after Fiddaman 1997; Sterman & Booth Sweeney 2002, p. 213) presents a stock-and-flow model of the carbon cycle and its influence on the global heat balance. This model is shown in Figure 2.2, depicting the major stocks of carbon, the flows between these stocks and their relationship to atmospheric heat balance.

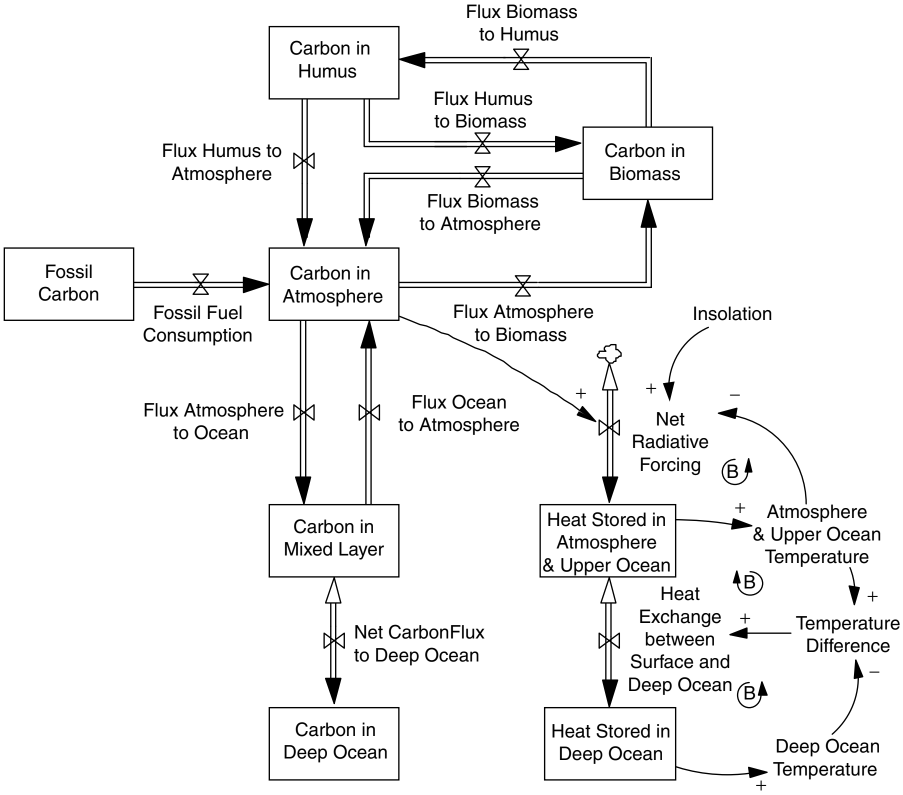


Figure 2.2: Stock and flow diagram of the global carbon cycle and heat balance (Sterman, 2000 after Fiddaman, 1997). The stock of carbon in the atmosphere influences the heat stored in the atmosphere and upper ocean.

This stock-and-flow representation is a valuable tool for identifying how carbon moves through the carbon cycle. It clearly shows that the flows are two way, except the flow from the fossil carbon stock[1]. Even with only six stocks and nine flows, it is difficult to mentally simulate the behaviour of this model and determine what the likely future behaviour will be.

Despite the complex dynamics of the carbon cycle, the issue at the core is the net difference between the carbon additions to and removals from the atmosphere. Sterman offers the bathtub metaphor (Kuznig, 2009) as a simpler model for thinking specifically about this relationship in the carbon cycle, shown in Figure 2.3. This 'carbon bathtub' model highlights the causal logic of the situation: if the additions exceed the removals, then the level of carbon in the atmosphere

1. Although, it is noted that the fossil fuel stock may be replenished on geological time scales

increases; if the removals exceed the additions, then the level of carbon in the atmosphere decreases, and if the additions equal the removals, then the level of carbon in the atmosphere will remain steady in dynamic equilibrium.

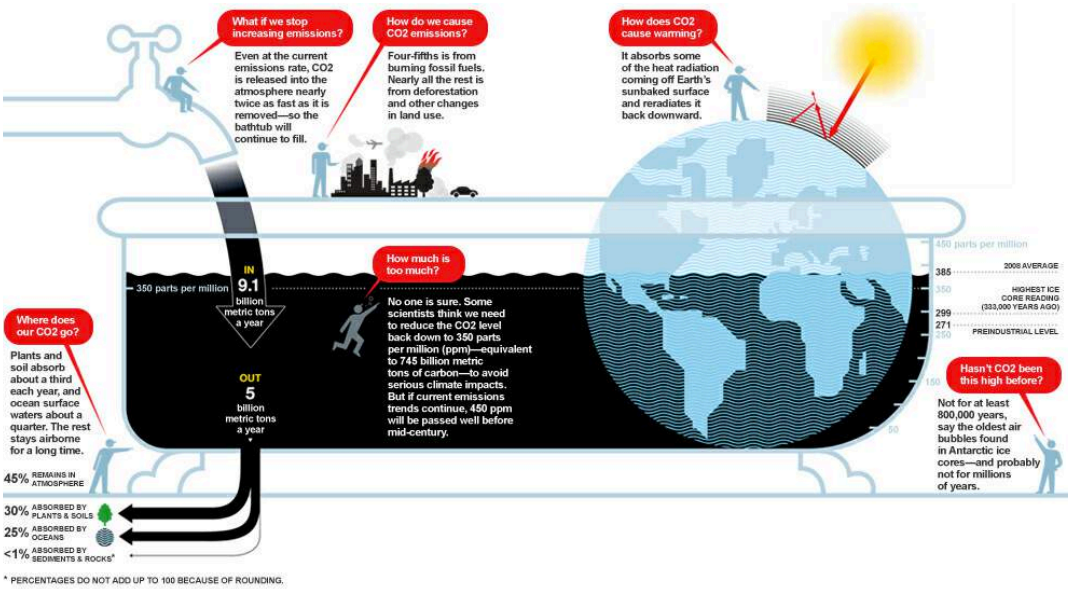


Figure 2.3: Infographic describing the carbon bathtub (Holmes in Kuznig, 2009). The carbon bathtub highlights the important relationship in the carbon cycle: the net flow, which when positive will increase the level of the bathtub, when zero will stabilise the level, and when negative will allow the bathtub to drain.

If we take a bathtub approach to the model shown in Figure 2.1 and consider only those aspects of the carbon cycle that human activities can change directly, we are left with land-use change and the burning of fossil fuels. Rapid afforestation could hypothetically—although not plausibly—balance out the burning of fossil fuels, and would require a significant reversal of current trends to replace the biomass removed through deforestation and other land-use change. Further, an increase in the vegetation carbon stock acts only as a buffer; living biomass that acts as a carbon sink also adds to the carbon in the atmosphere through respiration, and eventually decay, transferring carbon to both the soil and the atmosphere.

One flow that is not considered in any of these models is the anthropogenic removal of carbon from the atmosphere, effectively adding an extra drain to the carbon bathtub. Carbon capture and storage (CCS), where carbon is removed from the atmosphere and injected into underground reservoirs, is being considered by some as a high-tech solution to restoring the carbon balance. The IPCC (2005) discuss the plausibility of geological and oceanic storage methods. Conceivably, if carbon could be captured at the point of emission, or even sucked out of the atmosphere, then balance could be restored. However:

> All models indicate that CCS systems are unlikely to be deployed on a large scale in the absence of an explicit policy that substantially limits greenhouse gas emissions to the atmosphere.
>
> - IPCC (2005 p.43)

Further, there are many gaps in knowledge about how CCS will affect the Earth system, especially in the example of deep ocean carbon storage. With such uncertainty, it would be dangerous to consider this as a solution to reducing the stock of carbon in the atmosphere. Further, doing so could lead to reduced attention to strategies for mitigation.

Carbon-bathtub thinking then leaves only one viable option to stop the atmospheric carbon concentration from increasing: reduce the burning of fossil fuels immediately to zero. A key insight from this observation is that this action will only stop the stock of atmospheric carbon from increasing, and will not reduce it on the required time scale.

This is the educational challenge: participants in the T&P activity should be able to use the system as a physical analogue to think through this problem by experiencing the interaction over time of stocks and flows. This should help them to realise that the atmospheric carbon stock will continue to rise unless the anthropogenic carbon emissions decrease to zero.

2.1.2 Anthropogenic carbon emissions over time

In the simplified carbon bathtub, the natural cycle is largely in dynamic equilibrium. The anthropogenic disruption to the carbon cycle through burning fossil fuels is the primary difference. However, the rate of annual anthropogenic carbon emissions has been growing since the industrial revolution and rapidly since the end of the Second World War.

The trajectory of the carbon emissions from the burning of fossil fuels is shown in Figure 2.4, along with the resultant atmospheric carbon levels. The carbon emissions graph (Figure 2.4a) is a spline curve fitted to the atmospheric CO_2 record. This record is based on ice core data before 1958. From 1958, it is based on the yearly averages of direct observations from the Mauna Loa and South Pole observatories. The atmospheric CO_2 concentration graph (Figure 2.4b) is an estimate based on a compendium of energy use, with a consistent record from the United Nations since 1950.

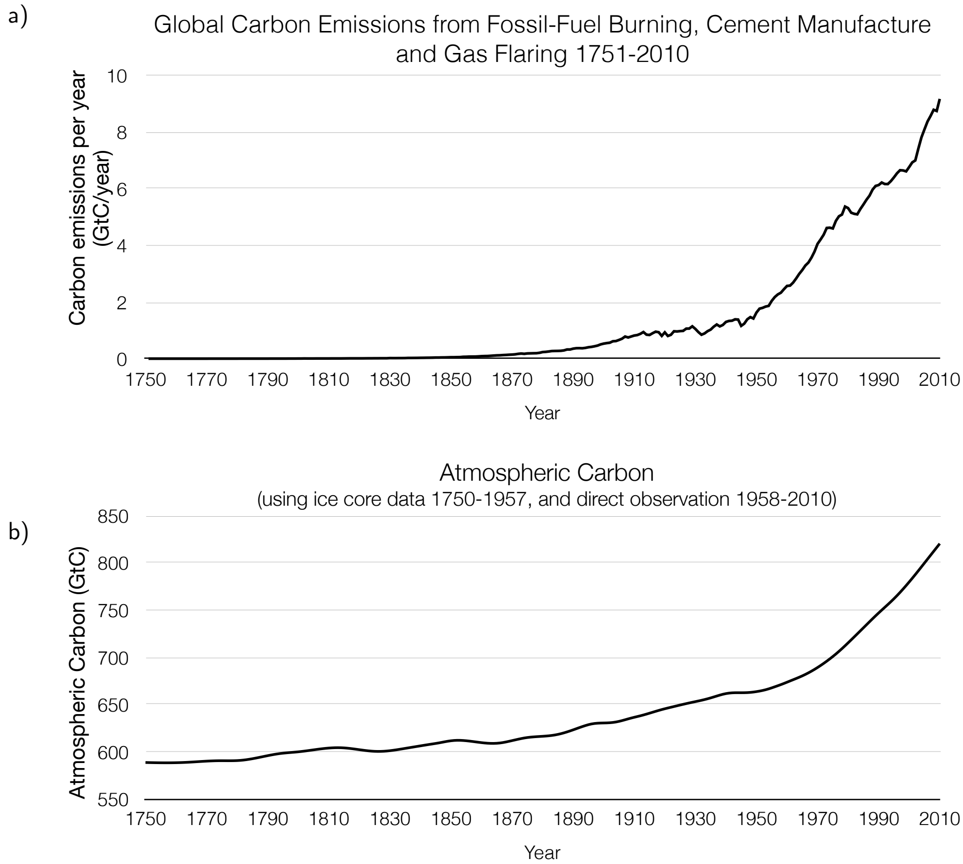


Figure 2.4: Anthropogenic carbon emissions and atmospheric carbon levels (c1750-2010) a) global carbon emissions from fossil-fuel burning, cement manufacture and gas flaring (1751-2010), b) atmospheric carbon levels (1750-2010). Source: global carbon emissions (Keeling et al, 2005; Scripps, 2015); atmospheric CO2 concentration (Boden et al, 2013).

The accumulation of carbon in the atmosphere has risen significantly over recent decades. Over the same period of time, the annual greenhouse gas emissions due to human activity has also risen, and the capacity of the planet to cope with this increase has been reduced through land-use change.

What will happen to the future atmospheric carbon concentration is of major concern to scientists and policy-makers alike. The argument about planetary boundaries put forward in Rockström et al (2009) is primarily that the goal should be to keep all Earth's systems in balance. When any of the boundaries are exceeded, it puts at risk the other components of the system. Given the current circumstances, a 2°C temperature rise is considered to be both the upper limit of this safe operating space, but also a best-case scenario. In 2009, the Copenhagen Accord recognised the extent of the problem[2]:

> We underline that climate change is one of the greatest challenges of our time. We emphasise our strong political will to urgently combat climate change in accordance with the principle of common but differentiated responsibilities and respective capabilities. To achieve the ultimate objective of the Convention to stabilize greenhouse gas concentration in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system, we shall, recognizing the scientific view that the increase in global temperature should be below 2 degrees Celsius, on the basis of equity and in the context of sustainable development, enhance our long-term cooperative action to combat climate change. We recognize the critical impacts of climate change and the potential impacts of response measures on countries particularly vulnerable to its adverse effects and stress the need to establish a comprehensive adaptation programme including international support.
>
> - UNFCCC (2009, p.1)

Despite this apparent political consensus, action has been lacklustre. Political arguments aside, a best-case benchmark for all negotiations is bringing anthropogenic emissions back to 1990 levels by (or after) 2020. In 1990, approximately 6.1 gigatons of carbon were released into the atmosphere (Keeling et al, 2005), which is still a significant annual addition to the atmosphere.

2. This sentiment is extended in the outcomes of CoP21 in Paris, in which climate change is recognised as an "urgent and potentially irreversible threat to human societies and the planet", that urgent "deep reductions in global emissions will be required", that the social and political circumstances should consider obligations to human rights and equity across a range of issues, and that there is an "urgent need to address the significant gap between the aggregate effect of Parties' mitigation pledges in terms of global annual emissions of greenhouse gases by 2020 and aggregate emission pathways". (UNFCC, 2015 p.1-2)

> Cumulative emissions of CO_2 largely determine global mean surface warming by the late 21st century and beyond. Most aspects of climate change will persist for many centuries even if emissions of CO_2 are stopped. This represents a substantial multi-century climate change commitment created by past, present and future emissions of CO_2.
>
> - Stocker et al (2013 p.27)

The problem is, as described, the greatest moral challenge of our generation.

2.2 Stock-flow (SF) failure

In this section, I outline the recent work in the area of SF failure, before describing SF failure in relation to the carbon cycle.

Previous work around SF failure focusses on testing understanding of systems through either identifying features of the system, such as the point in time that a stock is at its minimum or maximum level, or by visually integrating the flows to show the behaviour over time of the stock. Studies examining SF failure to date have largely shown that participants have a poor intuitive understanding of the dynamics in stock-and-flow systems. Notable findings are that:

- highly educated adults have a poor understanding of simple stock-and-flow problems, and results often violate the law of conservation of mass (Booth Sweeney & Sterman, 2000; Sterman & Booth Sweeney, 2002; 2007; Cronin & Gonzalez, 2007), such as the belief that an increase in the rate of flows would lead to a stabilisation of the accumulation (see the correlation heuristic in Figure 2.5).

- varying testing formats shows little demonstrated improvement of understanding; for example, using graphs, multiple choice or written descriptions (Sterman & Booth Sweeney, 2002; 2007; Cronin & Gonzalez, 2007; Cronin et al, 2009) or describing flows as succession of discrete stocks (Brockhaus et al, 2013; Sedlmeier et al, 2014)

- varying cover stories of similar dynamic systems shows little improvement of understanding; for example, using everyday examples such as water in a bathtub, money in a bank account, or people in department store (Kapmeier, 2004, Cronin & Gonzalez, 2007; Cronin et al, 2009; Kapmeier et al, 2015)

In the previous work listed above, a small amount of priming is typically given, and participants use their experience of the real world to inform a decision during an assessment exercise. Typically, the exercise itself does not prompt the participant to question, orientate or improve their mental model of the problem prior to assessment.

There are, however, some promising observations in testing SF failure. Cronin et al (2009) note that outcome feedback helps students to achieve the correct answer, but that improvement is gradual. Formal system dynamics instruction improves understanding of stock-and-flow systems; however, a minority of students after graduate-level training in system dynamics still appear to be confused by stocks and flows (Sterman 2010).

A brief summary of this previous work is shown in Table 2.1. Here, test type describes the context of the study for reference. A common approach in these studies is to provide a range of assessment tasks which involve "calculus without mathematics" through graphical integration and differentiation (as described in Sterman, 2000 §7.1). Success rates, typically defined as the proportion of students providing a correct response, show the range (low to high) of correct results for treatment conditions in the given task.

Table 2.1 should not be seen as a leaderboard, as there is no common or standard approach between the studies, even between studies with similar authorship. This is, perhaps, because the goals and approach of each activity are different, and the threshold for the reporting of correct responses varies. Further, the assessment tasks vary in difficulty, and were targeted at different groups of students. Table 2.1 does, however, provide an indication of the previous work and challenges in this area.

Table 2.1: Overview of key previous SF failure studies, including test type, assessment task, treatment conditions and range of indicative correct response rates

Test type	Assessment task and treatment conditions	correct reported % low	correct reported % high
Booth Sweeney & Sterman (2000)			
bathtub test	graphical integration	46%	83%
cashflow test	graphical integration	51%	77%
manufacturing case	graphical differentiation	32%	50%
Sterman & Booth Sweeney (2002)			
zero-emissions task	graphical integration and written (CO_2)	22%	36%
	graphical integration and written (temperature)	20%	46%
stable concentration task	multiple choice and written	18%	69%
Sterman & Booth Sweeney (2007)			
340ppm task	multiple choice, graphs: emissions with/without removal	'falling': 66%-89%	
400ppm task	multiple choice, graphs: emissions with/without removal	'falling': 30%-49%	
Cronin & Gonzalez (2007)			
department store task	identification; store and bank context (accumulation)	29%	43%
distractor point	identification; line and triangle (accumulation)	23%	42%
Cronin et al (*2009*)			
department store task	identification; multiple representation (accumulation)	31%	69%
	identification; multiple contexts (accumulation)	17%	38%
	identification; feedback and no feedback	12%	21%
	identification; priming with system dynamics	53%	68%
Sterman (2010)			
department store task	before and after system dynamics training improved:	24.7% to 46.1%	
Brockhaus et al (*2013*)			
square wave	succession of stocks, tubs and bath context	44%	67%
triangle	succession of stocks, tubs and bath context	41%	61%
discontinuous task	succession of stocks, tubs and bath context	22%	44%
Sedlmeier et al (*2014*)			
square wave	succession of stocks, animations	~32%	60%

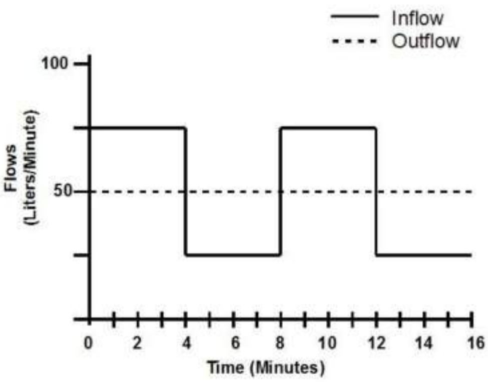
Behaviour over time graphs are common to these tests. Either, a graphical representation of the behaviour over time of a stock, or a graph of the flow rate as a function of time is given, and the participant asked to graph or judge the related behaviour. A written description is typically provided giving further details of the graph. Approaches differ in the representation used and preparatory information provided. Sedlmeier et al (2014) shows a typical example of the bathtub assessment task, which tests participant's bathtub-thinking skills, summarised in Box 2.1.

Box 2.1: Components of a typical bathtub test (summarised and reproduced from Sedlmeier et al, 2014). The written description supplements the graphical representation of the problem, providing numerical information to help students compute the level of the stock over time.

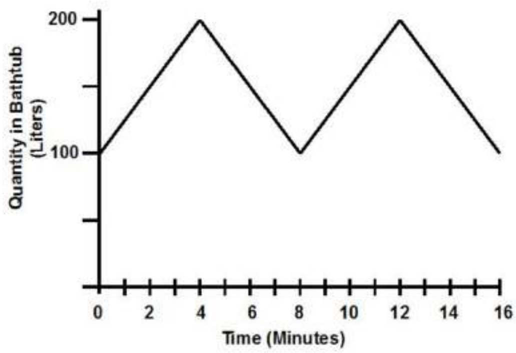
Written description

The flow diagram indicates a constant outflow of 50 litres per minute but a variable inflow of 75 litres per minute for the first four minutes that diminishes to 25 litres per minute for the next four minutes and repeats this pattern once. The bathtub has 100 litres of water at time t=0.

Graphical description



The inflow and outflow rates



A 'correct' solution
showing the corresponding stock level

Although the task shown in Box 2.1 is a relatively simple visual integration task, cohorts of students have failed to perform well. A number of strategies for reducing this type of SF failure have been proposed, such as providing adequate preparation, or instruction designed specifically to reduce the confusion between stocks and flows. However, at this stage no intervention has been able to provide a clear direction for improving participant's understanding of accumulation, including formal education.

> Given these results, the even shorter exposure to stock-and-flow concepts provided in short academic and commercial training workshops is highly unlikely to be effective in overcoming the correlation heuristic and helping people learn the principles of accumulation. Those teaching system dynamics in other formats, and with other groups, should carry out evaluative research to assess the impact of their curriculum and pedagogy on student learning. A second issue relates to the unusual characteristics of the subject population in this study. Graduate students at MIT are highly selected for top academic performance and capability;

> they have far more training in STEM and other quantitative disciplines (economics, business) than the average person. Still their understanding of stocks and flows prior to exposure to the course is extremely poor. Further research into the failure of the educational system to provide such basic concepts, and effective methods to teach these concepts in the K-12 grades, is sorely needed.
>
> - Sterman (2010, p. 331)

There is a passionate movement for teaching systems thinking from an early age in the American education system (K-12). Forrester (2009) tells us that “any child who can fill a water glass or take toys from a playmate knows” about accumulation. Draper’s experience of teaching high-school physics, including topics such as thermodynamics, using stocks leads him to think:

> Systems thinking, like any thinking paradigm, should be invisible—a natural way that people think about the world. In Barry Richmond’s words, it should be “the water we swim in.” Just as most of us don’t really deliberately choose to use inductive reasoning for a specific problem and deductive reasoning for a different problem (we just figure stuff out), people do not have to consciously know that they are using systems thinking for any particular problem. Instead, we should always be thinking in terms of feedback and circular causality. This principle also has to be applied to the teaching and to the learning of systems thinking and system dynamics. People do not have to know they are learning about the science of systems thinking or system dynamics. They just need to be taught, from the beginning, that the world around is in made of dynamic, interconnected systems and that there are tools we can use to understand these dynamic relationships.
>
> - Draper (2010, p. 53)

Some innovative teachers have described their efforts to teach stocks and flows in early primary school. For example, Benson (2010) describes primary school students drawing stock-and-flow models using stocks such as amount of soup in a pot, the number of passengers on a trolley car, the courage a protagonist develops in a story, and the number of people visiting a zoo. LaVigne et al (2015) describe the construction of shared metaphors using behaviour over time graphs amongst preschool children. For example, a storyline that oscillates between good and bad is a ‘crown’ story. Students transfer this metaphor to their everyday experience: “I’m having a ‘crown’ day.”

Previous work from the system dynamics community has shown that the carbon cycle is a topic where innovative teaching methods are needed to address SF failure.

2.2.1 SF failure and the carbon cycle

Developing an intuitive understanding about the dynamics of accumulation of carbon in the atmosphere is the specific problem that the T&P activity examines. Sterman & Booth Sweeney’s work (2002; 2007) shows that there is confusion between the rate of flows of carbon into the atmosphere and the corresponding changes in the level of the stock of carbon in the atmosphere. This confusion can contribute to poor policy decisions, such as the belief that a reduction of the rate of carbon emissions will also lead to a correlated reduction of the atmospheric carbon concentration.

Sterman & Booth Sweeney’s work considered many scenarios around the dynamics of the climate system, including the accumulation of CO_2 in the atmosphere and the global temperature changes resulting from increased radiative forcing. The CO_2 tasks involved questioning participants about the required action to stabilise the atmospheric carbon concentration at different levels with and without consideration of removal of carbon from the atmosphere.

Sterman & Booth Sweeney (2007) used a stable concentration task at 340ppm and 400ppm with three conditions: multiple choice (MC), an emissions graph (EG), and an emissions and removal graph (ER). The graph shown in Figure 2.5 is modelled on their 400ppm EG task. The task asks the participant to sketch the trajectory of future anthropogenic carbon emissions (the bottom graph) given a projected atmospheric carbon concentration (the top graph).

The correct answer to achieve the stabilisation of atmospheric CO_2 concentration is to reduce the anthropogenic emissions to zero, as shown in the bottom graph in Figure 2.5. The correlation heuristic shown would result in a rapid growth of the atmospheric CO_2 concentration. In the Sterman & Booth Sweeney study, 48% of participants in this task drew a graphical response of future emissions falling from current rates.

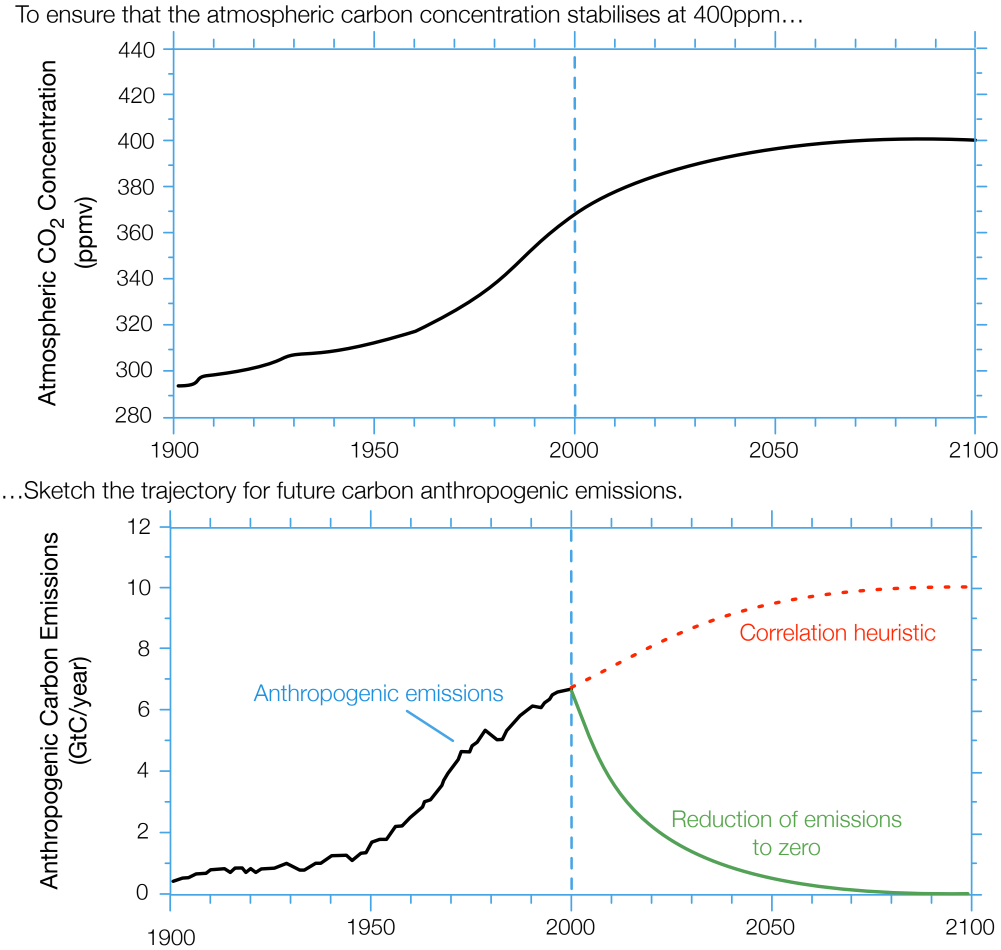


Figure 2.5: Typical stock-and-flow assessment task, adapted from the 400ppm task (EG condition) in Sterman & Booth Sweeney (2007).

Further, Sterman & Booth Sweeney observed the correlation heuristic in the 340ppm task, where the atmospheric CO_2 concentration falls rapidly from 2000 levels to 340ppm. 79% of participants drew a trajectory that had carbon emissions falling in a similar shape. However, to achieve a reduction in the atmospheric CO_2 concentration, the carbon emissions would have to fall below zero. This example demonstrates the prevalence of SF failure in the area. Results for their ER condition, where a removals trajectory was also considered, showed further confusion, leading to their concern about SF failure in relation to the carbon cycle.

Rather than varying the scenarios as in the previous work, the T&P activity will focus on the single 400ppm EG task shown in Figure 2.5, and instead I will vary the educational approaches to investigate whether participants can reach the correct solution through double-loop learning.

3 Constructing shared knowledge

The perceived threat from exposing our thinking starts early in life and, for most of us, is steadily reinforced in school—remember the trauma of being called on and not having the "right answer"—and later in work.

- Peter Senge (2006)

As Sterman & Booth-Sweeney (2002; 2007) note, adopting a wait-and-see approach about climate change is likely when the dynamics of the carbon cycle are not well understood. In this chapter, I provide the background rationale for using the T&P activity to address Sterman's (2010) call for effective methods to teach stock-and-flow concepts.

There are three overlapping perspectives through which the background is presented. The first is the *systems thinking* (ST) perspective, specifically with improving the ST capabilities of the educated public. Second, approaches in *cognitive science* (CS) explain how our understanding of the world is influenced by conceptual metaphors and how knowledge is framed through our embodied mind. And third, approaches in *learning and teaching* (L&T) provide inspiration for a new way of framing ST learning. ST is, at its simplest, an approach for interpreting the world, just as the theoretical foundations of the design sciences, physical sciences and the social sciences are.

Although theoretical underpinnings of the three differ significantly, there is an intersection of ideas when it comes to understanding how knowledge is constructed through experience with the real world.

Take, for example, these observations 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, 2003)
>
> ***Learning & Teaching 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)
>
> 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)

The intersection of these ideas is the focus for my review into background theory for the T&P activity.

Mental models

The starting point for this investigation is how the systems thinking community approaches thinking about thinking, particularly through the ST community's description of mental models. There are three attributes common to definitions of mental models (Meadows, 2008, pp. 86-87; Sterman, 2000, pp. 15-29; Richmond, 2010; Senge, 2006; Maani & Cavana 2007). These attributes are that mental models:

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

Through the lenses of the three perspectives, I consider how each approaches the development of mental models, described through the double-loop learning model. I use the first attribute—bounded rationality—to describe the type of activity in the single-loop learning, shown in Figure 3.1. In this feedback loop, decisions are made reactively to the information feedback observed in the real world. In this model, the learner doesn't change their thinking about the problem, and continues to make the same mistakes without learning.

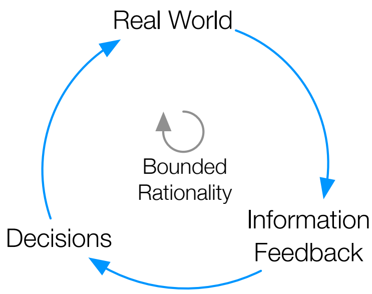


Figure 3.1: Influence diagram of bounded rationality as the first attribute of mental models, where information is derived from the real world, and decisions are made based on information from the real world observation.

I use the second attribute—experiential abstraction—to describe the refinement of mental models based on experience with the real world. In this feedback loop, information from real-world decisions are used to improve mental models. With the inclusion of the third attribute—dynamical deficiency—comes the capacity to change strategy and decision rules. In a learning context, this can be described as process learning, rather than learning as the skill of reciting content and facts. The student learns how to learn, and is readily able to adjust their decision-making process based on changes to their mental models. The addition of these two loops are shown in Figure 3.2.

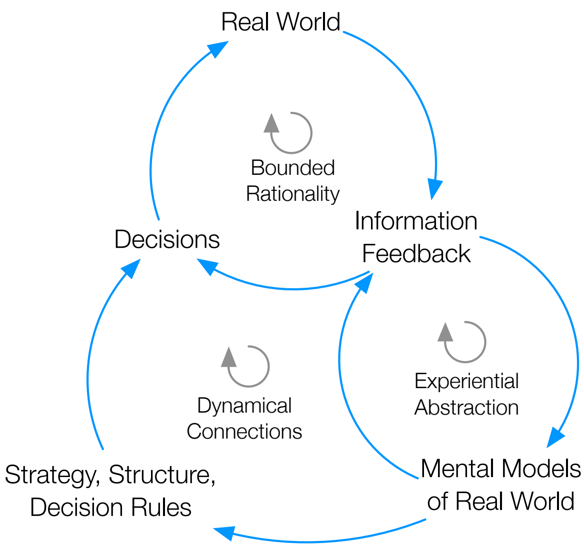


Figure 3.2: Influence diagram of the double-loop learning model. Experiential abstraction as the second attribute of mental models, where information from the real world changes the mental models used for decision-making in the real world, and Dynamical connections, where strategies are continuously refined and improved.

In this chapter, I examine the positive approaches that each perspective brings to facilitating development of these attributes of mental models, and in turn double-loop learning. The core ideas presented in this chapter are shown in Figure 3.3, and will be explored in the sequence: systems thinking, cognitive science, then learning & teaching.

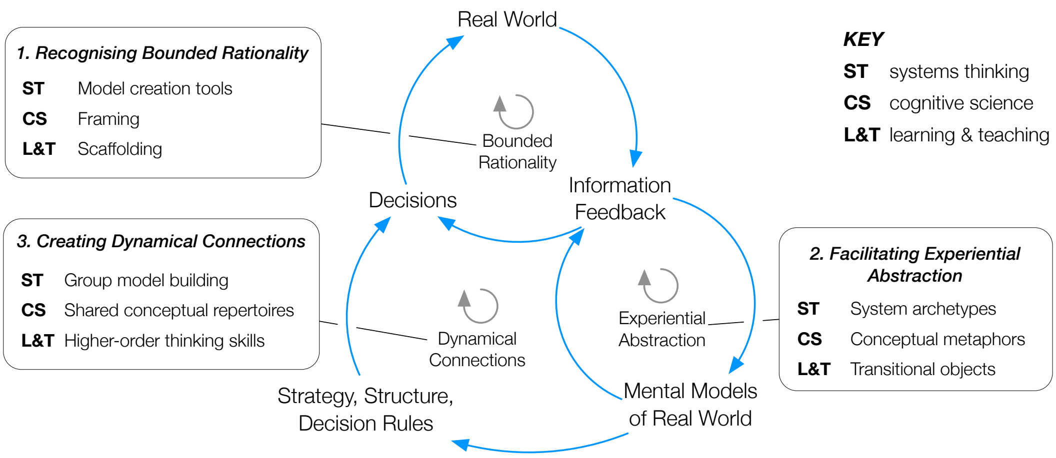


Figure 3.3: Annotated influence diagram of background theory for aspects of the double-loop learning model, as seen from the ST, CS and L&T perspectives. This diagram represents the structure of this chapter.

Throughout this chapter, I refer to the carbon cycle to demonstrate concepts. Understanding the anthropogenic impact on the carbon cycle requires a clear, shared understanding of the problem in order to make informed decisions about possible pathways for the future of our planet. I also refer to the development of the T&P activity, described in Chapter 4, the goal of which is for participants to construct their own intuitive understanding (reliable mental models) of the relationship between anthropogenic carbon emissions and atmospheric carbon levels.

3.1 A Systems Thinking perspective on constructing knowledge

The ST community sees the construction of knowledge as the development of mental models. Mental models are used by individuals to interpret the world around them. Richmond (2010, pp. 4-5) describes thinking as a two-step process:

constructing mental models, then simulating the models to draw conclusions. Mental models provide a framework for thinking, understanding and learning, and hence their study is central to building effective learning environments.

ST texts commonly introduce mental models as a background for describing how cause-and-effect logic is built. Mental models are commonly used in this context to explain why systems behave unexpectedly. When a formal model that has been carefully created behaves in a way that is not expected during simulation, it is likely due to the modeller's incomplete mental model that guided the construction of the formal model. When our mental models are challenged, there is an opportunity for learning by understanding how they could be improved.

An example is a student seeking to perform well in an exam. His or her previous experience with exams may lead to forming a mental model that cramming before an exam is an efficient way to perform well. This strategy may work for some time and on some topics, but when exams become more difficult, the strategy will no longer work. The reflective student will revisit their mental model about exams in order to improve their performance by, for example, adopting better study habits.

I will discuss how the systems thinking community develops the three attributes of mental models. First, I will discuss tools that the ST practitioners use to make bounded rationality explicit, such as model boundary charts, dynamic hypotheses and behaviour over time graphs. Then, I will discuss group model building as a tool for creating dynamical connections. To complete the ST perspective, I will discuss system archetypes and other small-order feedback structures as a tool for facilitating experiential abstraction.

3.1.1 Systems Thinking methods to recognise bounded rationality

All models, including mental models, are necessarily a simplification of the real-world, used to improve understanding of a situation. The well-known George Box quotation comes to mind: "Essentially, all models are wrong, but some are useful" (Box, 1987). All models are wrong because, by definition, a model represents a subset of the real world. This is true for mental models too. However, mental models are further flawed because they exist in our minds. They are perpetually in draft form, as the causal chains are challenged and revised continuously to match the world we experience.

Mental models exist in one's mind, and are not usually formally described. Although they may have been built from vast experience of the system of interest, it is difficult for a single person to understand the world from multiple perspectives. Because of this, an individual's mental models are often an incomplete representation of a given problem – they exhibit bounded rationality. The ancient Sufi story of the blind men and the elephant is commonly used to explain this (Meadows, 2008). In this story, many men each hold a different part of the elephant and each have a completely different understanding of both what it is and what its purpose might be. This story provides an example of why sharing, and then challenging, mental models of a problem can result in a deeper understanding of the system of interest.

The bathtub metaphor is a fundamental framework for systems thinking and the T&P activity. A visualisation of the bathtub metaphor is shown in Figure 3.4, showing the main components and relationships. This visualisation shows only some of the factors that could be included in a bathtub system; it is a simplification that represents a subset of the real world.

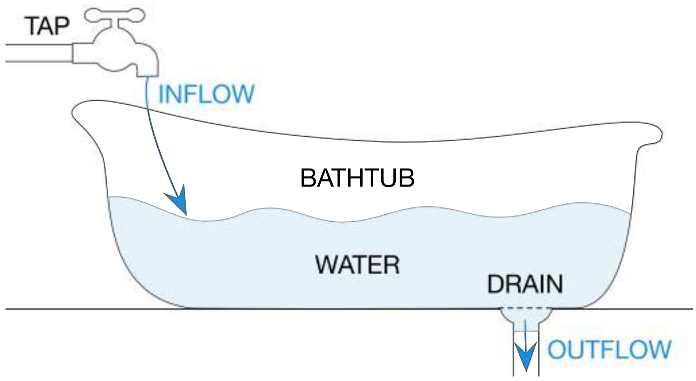


Figure 3.4: A generic bathtub visualising the components of the bathtub metaphor. Water enters the bathtub through a tap (the inflow), and leaves the bathtub through the drain (the outflow). The level of water in the bath (the current state of the stock) is limited by the size of the bath, and controlled by the net flow.

The bathtub metaphor will be used to examine how the ST community use a range of tools and methodologies to make models explicit. These tools show how ST modellers simplify complex problems so that they can draw useful outcomes from their models. By making a model explicit, the modeller's bounded rationality also becomes explicit.

Model boundary charts

A model boundary chart explicitly describes the scope of the system of interest. The modeller sorts relevant factors and variables into three categories:

- *endogenous*; inside the system, factors that are controlled by the relationships within the system
- *exogenous*; outside the system, external inputs to or outputs of the system
- *excluded*; not considered as part of the model

This process establishes the scope of the model, but also shows the underlying assumptions of the modeller (Sterman, 2000, p. 98). The model boundary chart helps to keep the system of interest useful. A model boundary chart for the bathtub metaphor is shown in Table 3.1.

Table 3.1: Model-boundary chart of the bathtub metaphor. Endogenous and exogenous variables will be included in the model, and the list of excluded variables describe the system components that are not going to be considered further, or are outside the scope of the model.

Endogenous	Exogenous	Excluded
water	desired water level	price or availability of water
volume of bath		temperature of water
rates of flow through taps		people/toys/bubbles
rate of flow through drain		gravitational constant
		alternative means of water addition
		alternative means of water removal
		evaporation rate
		bathtub material and surface finish

Of course, there is no single, 'correct' model boundary chart for a given system, as the modeller may choose to add or remove variables, or place them in different categories. He or she might choose to describe a different instance of the system, built on different assumptions. For example, the *desired water level* has been placed in the exogenous category in Table 3.1 because it is considered to be an external value or goal placed on the system by the person filling the bath. The *desired water level* could well be an endogenous variable in, for example, an automatic bathtub-filling system.

The variables in the excluded category in Table 3.1 may well have an effect on the system in certain situations; for example, the availability of water during a drought may become a strong influence on the desired water level. However, by placing them in the excluded category the modeller's underlying assumptions and intentions are made explicit, and their influence is not considered in further modelling.

By using a model boundary chart, the systems thinking modeller simplifies complexity by concentrating only on factors of interest. Once a problem has been scoped in a clear way, models can be constructed to represent the behaviour of the system.

Dynamic hypotheses

In order to construct and simulate useful mental models of a problem, the learner needs to be aware of the deficiencies of his or her models and employ strategies to continuously improve and refine them. Examining how mental models can be formalised provides direction on the design of the T&P activity.

The process of developing mental models into simulation models eventually requires the explicit description of components and relationships within the model. That is, moving the model from something that exists only in one's mind into something that can be shared, simulated, explained and improved. Sterman (2000) simplifies the modelling process into five key steps.

1. Problem articulation
2. Formulation of dynamic hypothesis
3. Formulation of simulation model
4. Testing
5. Policy design and evaluation

The formation of the dynamic hypothesis (described in Richardson & Pugh, 1981) is an important step in making a mental model explicit. The use of tools such as causal-loop diagrams and stock-and-flow maps helps to formulate such hypotheses (Sterman, 2000). Ford (1999) points out that being specific about the problem at this stage is critical to the success of the modelling process.

Causal-loop diagrams describe the nature of the relationship between the variables that the modeller selects. This allows the modeller to explore the logic of the feedback structures. In a causal-loop diagram, a series of variables are stated with series of causal links. These describe the relationship between the variables, shown using arrows. The variable at the tail of the arrow affects the variable at the head of the arrow.

The relationship between the affecting variable and the affected variable can be described as either positive or negative link polarity, shown with either a plus or minus symbol at the head of the arrow. A positive link polarity indicates that the causal effect pushes the variable in the same direction: an increasing variable at the tail of the link will increase the variable at the head of the link, and a decreasing variable at the tail of the link will decrease the variable at the head of the link. A negative link polarity indicates that the causal effect pushes the variable in the opposite direction: an increasing variable at the tail of the link will decrease the trajectory of the variable at the head of the link, and a decreasing variable at the tail of the link will increase the trajectory of the variable at the head of the link.

The polarities in a causal-loop diagram assist in describing the structure of the system. The polarities also determine whether the loop has a reinforcing or balancing behaviour. The logic here is similar to multiplication of positive and negative numbers. If all the link polarities are positive, or if there are even numbers of negative links, the behaviour of the closed loop will be reinforcing. If there are odd numbers of negative links, the behaviour of the closed loop will be balancing.

As an example, the endogenous and exogenous variables used in Table 3.1 are shown in the causal-loop diagram presented in Figure 3.5. The observed water level is controlled by the net inflow rate, which is an adjustment in or out of water according to the gap between the actual and desired water level. The behaviour of the system is balancing, driving the system towards the point where the gap between the actual and desired water level is zero because of the single negative link.

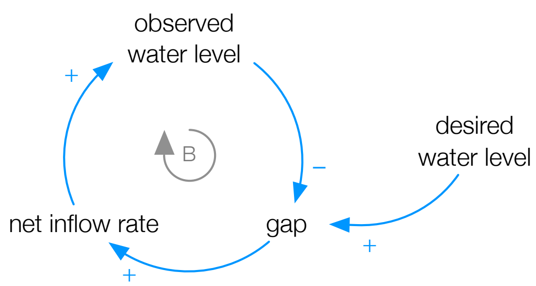


Figure 3.5: Causal-loop diagram of the bathtub metaphor dynamics. The system has a balancing feedback behaviour, as the net inflow rate is adjusted to meet the desired water level. The desired water level is assumed to be exogenous to the internal dynamics of the system.

Although causal-loop diagrams allow for the logic of a system to be explored, they are not able to be simulated until values and relationships are explicitly defined using mathematical equations. Stock-and-flow diagrams allow the modeller to define these relationships through structures that have embedded mathematical equations, enabling the simulation over time.

A simplified stock-and-flow representation of the dynamics of filling a bathtub adapted from the causal-loop diagram in Figure 3.5 is shown in Figure 3.6a. Here, the flow of water entering the tub is controlled by the gap between the observed water level, and making adjustments until it—approximately—reaches the desired water level. The resulting behaviour over time of this balancing-loop system is shown in Figure 3.6b.

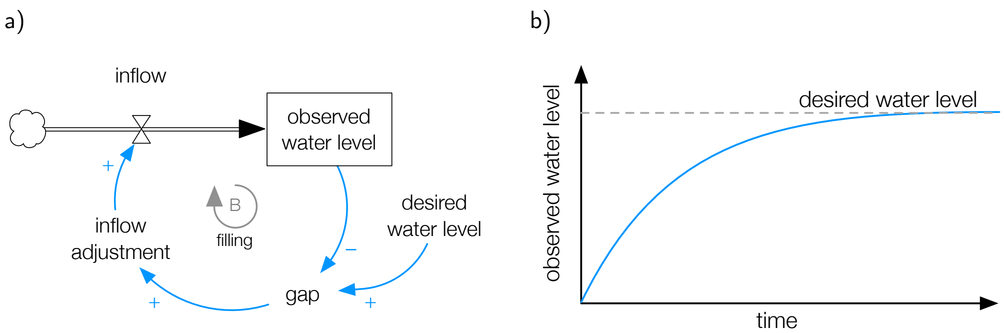


Figure 3.6: Stock-and-flow representation of filling a bathtub. a) Note that this instance of the system does not describe control of flow through removal of water from the bath. The boxed variable 'observed water level' represents a stock. The double arrow with valve represents a flow into the stock. The cloud is outside the boundary of the system of interest. Influence links are shown with single arrows. The system has a balancing behaviour. b) The behaviour over time of the balancing system described in (a).

After a system has been described through stock-and-flow notation, an ST practitioner will describe the system in equations that govern its behaviour. With these equations, the model is made formal and ready for simulation. This is how an ST practitioner makes the bounded rationality of his or her mental models explicit.

3.1.2 Systems thinking methods to facilitate experiential learning

It follows that if mental models are built from experience, then mental models can be changed through experiences, and that it should be possible to design a learning experience to do so. In this section, I present how systems thinking practitioners use small, useful models and the system archetypes to abstract everyday experiences into generalised ST models.

Small, useful models

This brings us to the second part of the George Box quotation: *some models are useful*. To include all the components and relationships of a situation can detract from understanding the key relationships within a system of interest. When all the factors of a problem are considered, such as those in the Excluded category in Table 3.1, the model becomes too complex to be useful.

An extreme example of this is the Afghanistan Stability / COIN[3] dynamics influence diagram (PA Consulting, 2009). This was largely ridiculed in the press, which reported that General Stanley McChrystal declared, “When we understand that slide, we'll have won the war” (Rogers, 2010). The ‘spaghetti diagram’ model was far too complex for decision-makers to understand, and was no longer useful, save for clearly demonstrating that the problem of counterinsurgency is complex.

Simple models can still be dynamically complex. They can become controversial because they are counterintuitive to fixed, static mental models. Two classic studies demonstrate this. Forrester’s urban dynamics model contained only nine major stocks (see Forrester, 1969) and allowed him to see many insights that later proved to be true. Forrester’s model showed that the introduction of low-cost housing in urban areas meant that large numbers of people were moving into areas where few jobs were available. Rather than alleviating poverty, this

3. COIN is an abbreviation of counterinsurgency

intervention created it and entrenched it in the community. This view was at odds with the otherwise politically popular intervention, but has repeated itself where low-cost housing has been introduced.

The World model of *Limits to Growth* (Meadows et al, 1972) considered stocks such as non-renewable resources; service, industrial and agricultural capital; land use; population, and; pollution. The model suggests that there is a limit and balance to these stocks. It was controversial at the time and is controversial still, in a global economy where growth and abundance are considered normal and worthwhile goals. Even so, its 'normal' run continues to exhibit accuracy to this day (Turner, 2008). The usefulness of models comes down to both the art and science of effective modelling, as well as the ability to communicate the outcomes.

System archetypes

Using smaller, more manageable models can be extremely useful in thinking about systems of interest. The system archetypes are small, generic structures that demonstrate dynamically complex but recognisable behaviour.

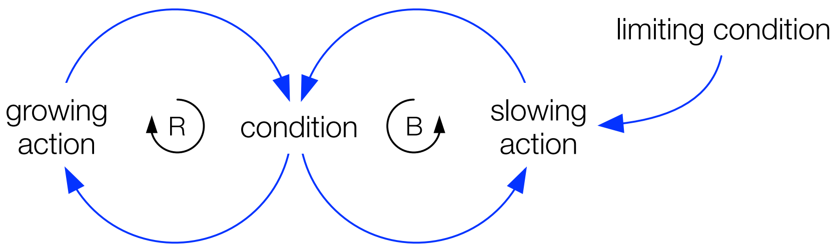
The most commonly used system archetypes (as described in Meadows, 2008; Senge, 2006; Maani & Cavana, 2007; Braun, 2002) are:

- balancing loops with delays
- limits to growth; limits to success
- policy resistance; fixes that fail
- the tragedy of the commons
- eroding goals; drifting goals; drift to low performance
- escalation
- success to the successful; competitive exclusion
- shifting the burden; addiction

These typically involve an interaction of one or two feedback loops, and at most four[4]. Generic system archetypes provide important examples that can improve our mental models of situations. Through understanding the closed-loop dynamics of the archetypes, we can infer causal logic structures that can be applied to other real-world situations.

The dynamics of these small models are driven by their feedback structures. The output at one time step feeds back into the system and becomes the input at the next time step. There are only two types of feedback: balancing and reinforcing. Balancing feedback, usually represented by the letter B, stabilises the level of a stock or exhibits goal-seeking behaviour. Reinforcing feedback, usually represented by the letter R, pushes the stock to grow or decline.

a)



b)

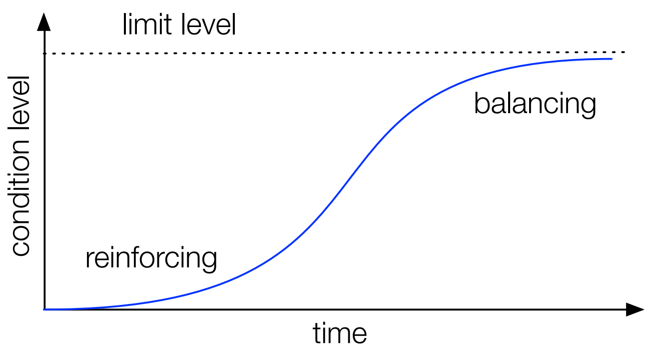


Figure 3.7: Influence diagram and behaviour over time graph of the limits to growth archetype a) an influence diagram showing the structure of the limits to growth archetype. The words represent variables and the arrows indicate influence. The variable at the tail of the arrow influences the variable at the head of the arrow. The R surrounded by an arrow represents a reinforcing behaviour; the B surrounded by an arrow represents a balancing behaviour. b) a typical plot of the behaviour over time in the limits to growth structure, which is limited by the level of the limiting condition. This behaviour is commonly called S-shaped growth.

4. the tragedy of the commons archetype has four feedback loops, one set of balancing and reinforcing loops for the individual and one set for community.

These feedback structures can be seen clearly in the system archetypes. Take the treatment of the limits to growth archetype in Senge (2006). 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. An influence diagram and behaviour over time graph of the limits to growth archetype is shown in Figure 3.7.

Senge offers a number of examples: affirmative employment strategies; learning a new skill such as tennis; the growth of start-up businesses or social movements; a city that grows rapidly and in doing so increases housing prices; an animal population that grows too fast, which leads to overgrazing and starvation. 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, monetary capital, labor, markets, management, resources or pollution. Rogers (2003) describes many other examples, such as the diffusion of innovations (ideas or products), and the growth of a sunflower.

Recognising an archetype in a real-world situation allows us to draw connections about future behaviour and recognise opportunities to alter that behaviour. The behaviour in the limits to growth archetype indicates that growth will eventually slow down and stop. In a situation where growth is required, Senge recommends focussing on removing or weakening the limitation. In the tennis example, this might include more coaching to identify and lift the limiting condition; 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.

From the ST perspective, the system archetypes are useful because they are small yet can provide the modeller with causal insights into real-world situations.

3.1.3 Systems thinking methods to create dynamical connections

System dynamics modelling assumes that the behaviour of a system is caused by its internal feedback structure (Meadows & Robinson, 2007; Richardson, 2011); the endogenous behaviour of the feedback model. However, mental models are not formal models and are typically incomplete, which makes them difficult to simulate. Maxwell et al (1994) suggest that decision-makers are often not equipped with mental models that display closed-loop cause-and-effect structures, and become better decision-makers when they are given dynamical insights described as 'causal chunks'. These are inferential models that provide a basis for decision making. That is, when X happens, Y is likely to happen.

This observation is critical in the design of the T&P activity. The goal of the activity is not necessarily to enable participants to be experts in modelling, or even to become aware of their own mental models. Rather, the goal may well be to equip participants with better causal chunks to improve or formalise their intuitive understanding; their ability to use their experience to make better decisions based on available data.

When mental models become explicit and closed-loop, they can be simulated: a process that can create dynamical insights that may later be remembered only as causal chunks. Further, understanding and informing mental models from multiple perspectives around a problem is important to assemble a complete picture of the problem at hand. Insights from group model building in system dynamics can help us to understand the process of building shared models and collective insights.

Group model building

In group model building, participants and relevant stakeholders play an active part in the modelling process (Anderson et al, 2007), rather than that being a process undertaken only 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 (CBSD):

> Models that are developed in [the] same place around public issues often have a common referent of experiences within the same community, so even when participants have not been involved in the modeling process, there can be a connection to a model developed by others from their community. Recognizing and drawing on the potential of communities to learn and build capacity over time through multiple system dynamics and group model building projects is a central idea in CBSD
>
> - Hovmand (2014 p.18)

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. Hovmand (2014 p.27) outlines activities within the *group model building with participants* phase of CBSD: introducing system dynamics, variable elicitation, defining the reference mode, structure elicitation, model formulation and testing, policy analysis, and transfer of ownership. Warren (2014) describes an agile model for group model building: problem definition; identify stock-drivers; capture stock-accumulation, and; identify interdependence. All contribute to working models and actionable findings. 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? Where are the leverage points? Can we have new eyes?

Each of these group model building examples describes a process for taking individuals' mental models and turning them into explicit, shared models to describe a situation of interest. Of relevance to the learning activity here is not necessarily the specific activities described in group modelling practice, but rather the common attribute between them: that they require the use of effective and well-timed prompts to get the most out of the participants. 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.

Good group model building also relies on a repertoire of activities to help the group form a shared understanding. A Think-Pair-Share (see Lyman, 1987) activity gets individuals to think through their ideas and make them explicit, before partnering with someone to share and combine their ideas. In a convergent activity, such as a problem set, the *pair* component is an opportunity to check individual solutions and work through any differences before sharing with the broader group. In a divergent activity, such as building a discussion paper, the *pair* component provides an opportunity to argue, justify and perhaps see new perspectives.

Group collaboration can also be encouraged using the jigsaw approach (first described by Aronson, 1978). In this technique, individuals leave their home group to become an expert on a given topic, by working in a short-lived group with 'experts' from other groups. They return to their home group with the relevant insights for their area of expertise, making individuals more accountable for their learning and at the same time broadening their home group's perspectives on a problem. Another method used to broaden perspectives within a group is six-hat thinking (De Bono, 2000). This gets the group to move through different points of view (wear different hats):

- the white hat to identify information known or needed
- the red hat to explore hunches and intuition
- the black hat to play devil's advocate
- the yellow hat to approach the situation optimistically
- the green hat to explore creativity
- the blue hat to facilitate the thinking process

In a facilitated process, this approach provides reference points for the group to think through problems and provides a different way for individuals to see beyond their own perspectives and opinions.

3.1.4 Application of Systems Thinking perspectives to the T&P activity

In this summary, I conclude the ST perspective by demonstrating how the ST methods described here relate to the T&P activity, which is described in detail in Chapter 4.

Model boundary chart and dynamic hypothesis

A model boundary chart was used to explicitly define the variables included in the simulation model in the T&P activity. The model boundary chart is shown in §4.1, alongside the dynamic hypothesis for the variables, relationships and expected dynamical behaviour.

Small, useful models and system archetypes

The T&P activity is an extension of a small, useful model with behaviour similar to the bathtub metaphor. Efforts were made to ensure that the carbon cycle (explained in Chapter 2) was described through the smallest number of variables and relationships, whilst still making the T&P activity useful and relevant.

Group model building

The components of the T&P activity require a small amount of construction and prompted interaction, which is in line with the observations that groups included in the modelling process take ownership of the model. The idea of targeted prompts that guide the group model building process were the basis for the instructional prompts for the activity, described in §4.2.

3.2 A Cognitive Science perspective on constructing knowledge

Second-generation cognitive science provides additional and complementary perspectives for thinking about the development of the key attributes of mental models. In this section, I will explore four key ideas. First, I discuss the bounded rationality of communication in the exchange of ideas, and how ideas are framed. Second, I relate our physical experience with real-world objects to the use of conceptual metaphors, which allow us to understand one thing in terms of another. Third, I discuss how dynamical connections are created through shared conceptual repertoires. Fourth, I discuss I2S, an interdisciplinary framework that builds on many aspects of a cognitive science approach to constructing knowledge. To conclude the cognitive science perspective, I summarise the approaches from CS in relation to the T&P activity.

3.2.1 Cognitive Science methods to recognise bounded rationality

Communication involves interaction for the exchange of ideas and knowledge. At its simplest, it is the process of sending and receiving signals with a previously agreed set of conventional 'meanings'. When the intended messages are coded using a finite set of discrete signals, such as those of mathematics, communication is unambiguous, providing that both the sender and receiver both are fluent in mathematics. Weaver (Shannon & Weaver, 1949, p. 4) describes problems in communication occurring at three levels:

> LEVEL A. How accurately can the symbols of communication be transmitted? (The technical problem.)
>
> LEVEL B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
>
> LEVEL C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

Reddy (1979) notes that an *a priori* shared context, such as fluency in mathematical concepts, between the transmitter and receiver is required to allow for the coding and decoding of signals. However, communication becomes significantly more difficult in human communication, where both the coding and decoding can be ambiguous. For example, the speaker will transmit a series of signals (words), and the listener will receive these signals and construct a message. This message may or may not match the speaker's intention. An *a priori* shared context in human communication requires iterations around a communication loop to ensure that the conceptual repertoire is shared (Newell, 2012).

Framing

The field of cognitive linguistics offers some insights into how language can be used to frame human interactions in particular ways. With respect to the T&P activity, if the language used to describe the task is framed in a particular way, it may change the behaviour of the participants.

Lakoff (2004) discusses how values can be framed in the language that we use. For example, *tax relief* immediately places a positive frame around a policy of reducing taxes. Although essential government-run services require adequate funding generated through taxes, the word *relief* frames the activity of cutting taxes as a virtuous activity, and places anyone who is against cutting taxes in the wrong.

Lakoff describes how this frame, created by the conservative side of US politics, then pervades the media and public psyche and in turn makes it difficult for the progressive side to attack.

> This is what framing is about. Framing is about getting language that fits your worldview. It is not just language. The ideas are primary—and the language carries those ideas, evokes those ideas.
>
> - Lakoff (2004)

Careful attention to the language used to frame a problem or viewpoint can make it easier, or harder, to gain support for ideas. This is important consideration for the design of the T&P activity, where it is used to help participants understand the dynamics of stocks and flows in the often politicised topic of climate change. For example, participants may come to the T&P activity with a set of values, which frames their thinking about the problem of climate change. The environmentalist and the economist may frame the problem in different ways, using different language, and will have a set of solutions that are consistent with their worldview. Careful consideration is required to frame the activity in a neutral way so that unintended biases do not surface and dominate the responses.

3.2.2 Cognitive Science methods to facilitate experiential learning

Our mental models are built from experience with the real world. In a learning environment, with multiple people with multiple backgrounds, it is important to pay careful attention to the creation of a shared understanding between participants. If this is not done adequately, then participants will unnecessarily have different interpretations and understanding of the purpose of the lesson.

Conceptual metaphors

Lakoff & Johnson (2003) argue that human thought occurs through underlying metaphorical constructs.

> Human conceptual systems are metaphorical in nature and involve an imaginative understanding of one kind of thing in terms of another.
>
> \- Lakoff and Johnson (2003, p. 194)

Here, *metaphor* is taken to mean *conceptual metaphor*, referring to the underlying concepts in language that our inferences about the world are based on. An identical structure underlies the use of metaphor used in poetry and rhetoric; these poetic uses only make sense to us because of our pre-existing repertoire of metaphorical constructs (Lakoff & Turner, 1989).

A relevant example of how communication occurs through underlying metaphorical constructs is the container metaphor. The container metaphor, describes the construct where objects (things, people, ideas, words, et cetera) are placed into and taken out of a container, and is fundamental to human cognition. The container metaphor provides the basis for Reddy's conduit metaphor, which is used to describe the structure of communication:

> The speaker puts ideas (objects) into words (containers) and sends them (along a conduit) to a hearer who takes the idea/objects out of the word/containers.
>
> \- Lakoff & Johnson (2003, p. 10)

Newell (2012) builds on Reddy's ideas in describing an elemental communication loop. Effective communication between two people (a transmitter and receiver) occurs through an iterative loop, where the transmitter and receiver compare the sent and constructed messages. The signal is required to be translated through both the transmission process and the conceptual repertoires of the two people. If the conceptual repertoires of the individuals are not identical, then communication is ineffective.

A conceptual metaphor has an underlying organisation of categories and relationships, called a schema. Consider the container metaphor further. Phrases such as *I am <u>in</u> trouble* or *I <u>got out of</u> trouble* make sense because of the container schema logic, based on our physical experience with containers. Objects can be situated in a spatial relationship to the container, and containers can fill and drain. *The trouble is <u>behind</u> me* or *trouble is <u>building</u>.* The container can be used to represent many abstract concepts. *She helped me <u>get out of</u> trouble, and we fell <u>in</u> love.* The container is universal to human cognition because of our physical experience with, and our shared *a priori* understanding of, containers.

In a conceptual metaphor, concepts from one context (the source domain) can be mapped into another context (the target domain). The mapping between the source and target domains is partial, and the use of vocabularies between the source and target domains can highlight or obscure concepts and in turn change the way that concepts are understood (Lakoff & Johnson, 2003). The mapping between domains can enable people to think through cause-and-effect structures in the source domain without requiring expert knowledge in the target domain.

This, together with Newell's elemental communication loop, reveals an important observation for building a shared understanding in a group learning environment: communication using conceptual metaphors can be effective without the need for prior instruction, providing that the conceptual metaphors are part of a shared conceptual repertoire.

3.2.3 Cognitive Science methods to create dynamical connections

Conceptual metaphors with their underlying logic can be used intuitively. Lakoff & Johnson (1999) explain this as a function of the embodied mind. For example, the container schema logic is transferable between many contexts and our understanding of the schema is embodied and intuitive because of our spatial experiences with containers.

Attempts to improve communication require careful selection of conceptual metaphors. However, problems can occur even when communication involves simple concepts. For example, orientational metaphors have a schema that organise whole systems of understanding, such as *more is up* and *less is down*. This leads to phrases such as *She is at the top of her game* and *He has hit an all-time low*.

The orientational relationship is not universal. *More is up* is often linked with the rationale that *up is good* and *down is bad.* There are situations where the orientational relationship is reversed. For example, in a business or economics domain, growth is accepted and commonplace (good), whereas in an environmental domain, economic growth is viewed with caution (bad). Examples can be drawn from some of the most complex problems facing society today, and can be seen in newspaper headlines daily:

> Obesity soars to 'alarming' levels in developing countries
> - Mark Tran in the Guardian Online, 3 January 2014
>
> Sea level rise quickens more than thought in threat to world's coasts
> - The Sydney Morning Herald, 15 January 2015, via Reuters
>
> Rising income inequality harms growth, OECD says
> - Paul Hannon in the Wall Street Journal, 12 August 2014

These counter-relationships challenge our metaphorical understanding of problems: here up is not good. These headlines require nuanced shared conceptual repertoires to understand. The orientational metaphor—up is good—behind the third example headline frames inequality as a problem for growth.

Shared conceptual repertoires

Modern cognitive science offers insights into ways that shared conceptual repertoires can be created, based on the observation that our understanding of the world is largely metaphorical. The words we use to communicate themselves have no universal meaning (Lakoff & Johnson, 2003; Newell, 2012). Conceptual metaphors have schemata that organise systems of information and understanding, and these schemata allow us to map concepts from one context to another.

The ability to understand a problem and organise thinking, even at a basic level, makes conceptual metaphors a *powerful idea* (Newell, 2012). A relevant example of metaphorical mapping is shown in Table 3.2, demonstrating how the bathtub metaphor maps bathtubs to systems concepts, such as accumulation and state-change processes.

Table 3.2: Conceptual mapping of the bathtub metaphor between the source and target domains (Newell, 2012, p. 781) Note: * the arrows "⇒" represent the expressions "corresponds to" or "maps to".

Conceptual source domain: a bathtub		Conceptual target domain: a system component
The water in the tub	⇒*	An accumulation, a 'stock', a variable that measures the state of the system component at time t
The amount of water in the tub	⇒	The amount accumulated, the 'level' of the stock, the 'magnitude' of the state variable at time t
Water entering the tub from the tap	⇒	An 'inflow', a process that increases the amount accumulated
Water leaving the tub through the drain	⇒	An 'outflow', a process that decreases the amount accumulated
Water leaving the tub through splashing	⇒	Another outflow, another process that decreases the amount accumulated

The bathtub metaphor itself is an instance of the more generic container metaphor. A container can only be filled to its capacity, and the level within the container is controlled by processes that add or subtract from the stock.

System dynamics modelling relies heavily on the container metaphor in its visualisation of stocks and flows. There have been many visual representations of the stock-and-flow notation as technology and software packages used in system dynamics have changed. Although each look different, they carry the same stock and flow conventions. Stocks are represented as boxes (the container) and flows are represented by emphasised arrows (pipes) with a valve or tap. Clouds represent stocks outside the system of interest, and single-lined arrows indicate causal relationships between variables. Examples have been compiled in Figure 3.8 to show different representations of stocks and flows. The universality of the container metaphor helps these different representations to be easily understood.

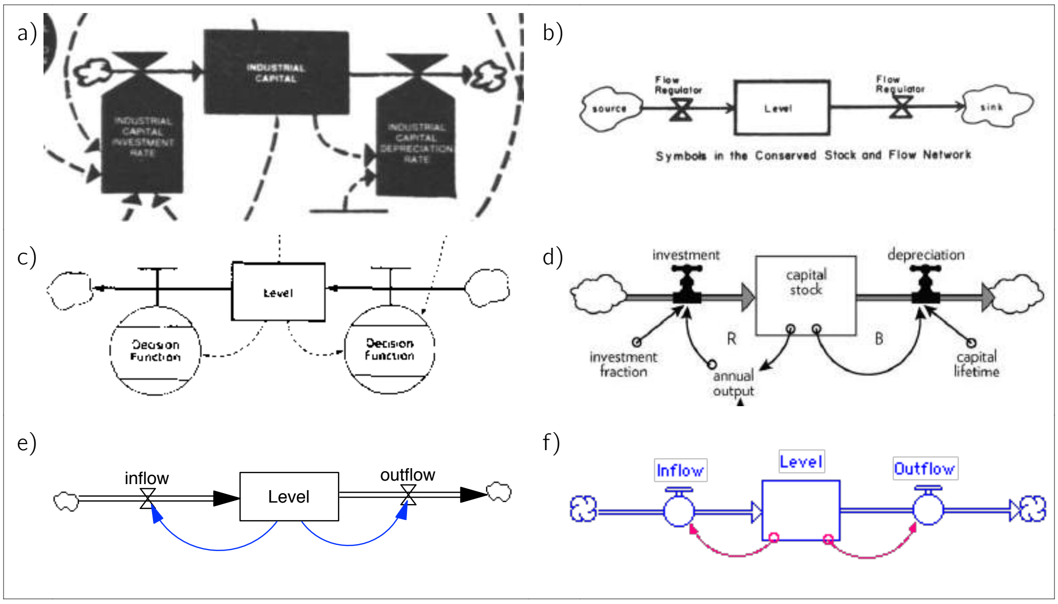


Figure 3.8: Different representations of stock and flow components. Stocks are represented as a box, arrows with a valve or tap indicate flows, clouds represent stocks outside the system of interest, and connection arrows indicate causal relationships between variables. Diagrams are reproduced from a) Meadows et al (1972); b) Morecroft (1982); c) Morecroft (1988); d) Meadows (2008); e) diagram drawn in VensimPLE v6.3; f) diagram drawn in STELLA v10.0.6.

The bathtub metaphor is an instance of the container metaphor, based on the concrete and largely universal experience of using a bath. The physical experience, however, has attributes that are not necessarily useful if mapped within the conceptual metaphor. For example, a bathtub has many attributes that could be measured: temperature of water, turbidity of water, number of people or toys in the bath, or the bubble-to-water ratio.

Considering the level of water as the only stock in Table 3.2 means that the mapping is selective. Just as in the system boundary chart in Table 3.1 for creating a dynamic hypothesis, it is necessary for the conceptual mapping to be partial if it is also to be useful. If, for example, the water temperature was considered in Table 3.2, it would add an extra dimension to the system, and the conceptual metaphor may be less applicable to general systems concepts. It could, however, be more applicable to other concepts, such as an analogue for understanding entropy.

Some systems thinkers, however, would continue to argue that the mapping in Table 3.2 is incomplete and requires improvement. A standard addition to descriptions of the bathtub metaphor (see Sedlmeier et al, 2014) is to replace the tap and drain with pumps to remove the non-linearities involved in the removal of water through the drain as a function of the height of the water. While this revision might provide a clearer mapping with the systems concepts in the target domain, it can detract from the metaphor as it moves the source domain away from the familiar, concrete experience of using a bath.

Others use alternatives to the bathtub to allow for more flexible mapping to specific source domains. Newell (Dyball & Newell, 2014) describes the water tank metaphor, an extension of the bathtub metaphor, to help understand ecological systems. Rusty water tanks are iconic to Australian rural life. The orientational logic of the water tank is different to the bathtub. An overflowing water tank can represent good times—a desirable state—whereas an overflowing bath is a sign of chaos or danger; an undesirable state. The emphasis on the flows are different. In the water tank metaphor, the water in the tank is precious, and the outflows are carefully observed due to the uncertainty of when the next rainfall would be. The differences are subtle, but important in creating consistent conceptual metaphors.

Physical experience can also put artificial bounds on the effectiveness of the conceptual metaphor. In the physical experience of using a bath, it is unusual to be continuously both filling and draining. In a typical use case, the bath would be empty, filled, in use, and then drained. Because of the behavioural activities that go alongside physically using a bath, the bathtub metaphor may obscure certain ways of thinking in the target domain. For example, the water level of the bath usually sits safely within user-defined bounds, and so its value in helping one to think through unusual problems may be limited. Implicit in the use of the bathtub metaphor is also an assumption that there is a universal bathtub experience: many cultures don't use baths, or use baths differently.

The acknowledgement that understanding is constructed through our embodied experience with the physical world has many implications for the way that learning environments should be created to build shared conceptual repertoires for successful educational outcomes. Experiential learning situates knowledge in the embodied mind, and provide frameworks for learners to think about problems.

3.2.4 A Cognitive Science basis for interdisciplinary group work

An example of how the ideas from cognitive science are being used generally to improve a shared understanding between disciplines can be seen with the recent establishment of the integration and implementation sciences (I2S). The goal of this field is to identify how to conduct collaborative research across disciplinary boundaries to solve complex, real-world problems. Although this is not a movement from within the field of cognitive science, it does draw heavily on ideas from both the ST and CS perspectives, and as a result is worth discussing in relation to building and sharing better mental models.

Bammer (2013) describes three domains as the basis for this approach:

1. Synthesising disciplinary and stakeholder knowledge
2. Understanding and managing diverse unknowns
3. Providing integrated research support for policy and practice change

Within these domains, Bammer describes six key, interrelated categories of concepts and methods are explored:

1. Taking a systems view
2. Scoping
3. Boundary setting
4. Framing
5. Dealing with values
6. Harnessing and managing differences

Bammer states the importance of paying careful attention to framing collaborative research, "so that the research team can accurately convey what it is setting out to do." (Bammer, 2013, p. 40). Bammer also outlines the importance of understanding blind spots that exist because of our values, arguing that a disproportionate amount of time can be spent exploring certain perspectives because of our values.

Understanding that framing and values can create blind spots in a group dynamic is important, but Smithson (Bammer & Smithson, 2009, p. 21) argues that blind spots are actually a form of social contract. In fact, blind spots around unknowns within a group are required for learning, creativity, social relations and cultural practices. An example that Smithson provides is *trust*. When two people trust each other, they create a blind spot in their relationship through the quasi-contractual arrangement. This allows the pair to be productive without the need to continuously question the trust relationship. That is, until that trust is breached.

The field of I2S is still in its developing stages, but promises to be a rich area of practical application of the concepts from cognitive science discussed in this section.

3.2.5 Application of Cognitive Science perspectives to the T&P activity

In this summary, I describe how cognitive science perspectives and approaches described here relate to the carbon cycle, described in Chapter 2, and to the T&P activity, described in Chapter 4.

Communication and Framing

As mentioned, careful attention is required around the language that is used to described the T&P activity. The topic of climate change is often politicised, and efforts have been made in the activity to present the carbon cycle in neutral language. In the design of the experiment, two discrete sets of language were used in the descriptions of the components of the T&P activity. In the first set, the terms ‘emissions’ and ‘absorptions’ were used to describe the flows in the carbon cycle. During the workshop facilitations, it became clear that some participants did not have the same *a priori* understanding of these concepts. In the second set of experiments, the terms ‘additions’ and ‘removals’ were interchanged into the T&P activity. This is described in more detail in §4.2.

Conceptual metaphors and Shared conceptual repertoires

The T&P activity represents a physical analogue of the carbon cycle. It is itself a conceptual metaphor; that is, it is a tool for understanding one thing (the carbon cycle) in terms of another (the T&P system). The interaction with the physical analogue as a group establishes a shared conceptual repertoire for discussion and communication of ideas. Participants can use the T&P system as a basis for discussion about the carbon cycle, such as an increase in the speed of the pumps to represent an increase in the rate of emissions. Participants can also discuss the dynamics of the carbon cycle in terms of the T&P activity, such as a reduction of emissions being represented by slowing the speed of a pump. The mapping of the conceptual metaphor is discussed further in §4.3.

3.3 A Learning & Teaching perspective on constructing knowledge

The L&T perspective is the final perspective I explore in relation to double-loop learning. Learning is more than the simple process of knowledge transfer from instructor to student. Effective learning occurs when mental models are developed, adjusted and enhanced by the learner. This process is innately experiential, embodied and continuous.

In this section, I will explore three key ideas. First, I discuss scaffolding and social learning as a way of escaping bounded rationality. Second, I relate experiential abstraction to concrete learning, and the role of transitional objects as a tool for relating the physical world to systems of knowledge. Third, I explore active learning as a tool for developing higher-order thinking skills. To conclude the chapter, I relate the concepts from the L&T perspective to the T&P activity.

3.3.1 Learning & Teaching methods to escape bounded rationality

One way to investigate the nature of learning is to look at the structure of the formal learning system: the education system. The education system is broad and complex, with many different goals, perspectives and approaches. The education system has a reputation for passive approaches, where the process of learning is often a process of storing and retrieving information, rather than construction of knowledge.

> I think it's an exaggeration, but there's a lot of truth in saying that when you go to school, the trauma is that you must stop learning and you must now accept being taught.
>
> - Seymour Papert in discussion with Paulo Freire
>
> (*Afternoon Journal* c1980)

Meadows (2008) notes that the over-arching purpose of a university is to discover and preserve knowledge and to pass it on to future generations, but sub-purposes can pervert that goal. For example, if the goal of a student is to get a good grade, this does not necessarily require her or him to learn, provided they can achieve a good grade by other means. If, instead, the goal of the student was to discover knowledge, she or he could indeed learn quite a lot, but fail the course.

> Education thus becomes an act of depositing, in which the students are the depositories and the teacher is the depositor. Instead of communicating, the teacher issues communiques and makes deposits which the students patiently receive, memorize, and repeat. This is the 'banking' concept of education, in which the scope of action allowed to the students extends only as far as receiving, filing, and storing the deposits. They do, it is true, have the opportunity to become collectors or cataloguers of the things they store. But in the last analysis, it is men themselves who are filed away through the lack of creativity, transformation, and knowledge in this (at best) misguided system. For apart from inquiry, apart from the praxis, men cannot be truly human. Knowledge emerges only through invention and reinvention, through the restless, impatient, continuing, hopeful inquiry men pursue in the world, with the world, and with each other.
>
> – Friere (2005, p. 72)

A learner can escape bounded rationality through creativity. Creativity is central to problem solving, and a necessary skill for problem solvers of the future. Creativity in education is a topic often discussed in various fora, and examples from popular culture are common. Sir Ken Robinson's (2006) TED talk titled 'How schools kill creativity' describes how creativity is as important as literacy, how schools push students out of creative subjects, and how intelligence is and should be recognised as diverse and dynamic. “We have to rethink the fundamental principles on which we are educating our children.” This talk is also the most-watched TED talk, with over 31 million views (May 2013; TED n.d.).

American folk singer Harry Chapin describes the story of a young boy entering the American education system, who progressively has his creativity squashed in the teacher’s pursuit of uniformity and standardised teaching practices.

> The little boy went first day of school
> He got some crayons and started to draw
> He put colors all over the paper
> For colors was what he saw
>
> And the teacher said, “What you doin' young man?”
> “I'm paintin' flowers,” he said
> She said, “It's not the time for art, young man
> “And, anyway, flowers are green and red
>
> “There's a time for everything, young man
> “And a way it should be done
> “You've got to show concern for everyone else
> “For you're not the only one.”
>
> And she said, “Flowers are red, young man”
> “Green leaves are green
> “There's no need to see flowers any other way
> “Than the way they always have been seen.”
>
> But the little boy said, “There are so many colors in the rainbow,
> “So many colors in the morning sun
> “So many colors in the flower
> “And I see every one.”
>
> - ‘Flowers are Red’, Harry Chapin (1978)

This theme is common in the educational literature too. Despite China’s extraordinary and consistent results in standardised testing, especially the computational abilities of students from its education system, the nation is facing a dire creativity shortage. Zhao (2014) argues that a focus on creativity is essential to China’s future economic growth. Martinez & Stager (2013) are direct in describing what educators need to do to promote creativity in the classroom: "students learn creativity by being creative." By being creative, they are active participants in the knowledge-construction process.

Piaget proposes that an education should involve the full development of an active and independent mind:

> The traditional school offers the student a considerable body of knowledge and gives him the opportunity to apply it to various problems or exercises. It "furnishes" the mind and subjects it to various "intellectual gymnastics" which are entrusted with strengthening and developing it. In cases where this type of learning is forgotten ... there remains at least some satisfaction in having exercised the intelligence ... having once known [it] becomes the important objective. A student who achieves a certain knowledge through free investigation and spontaneous effort will later be able to retain it; he will have acquired a methodology that can serve him for the rest of his life, which will stimulate his curiosity without the risk of exhausting it.
>
> – Piaget (1973, pp. 92-93)

Piaget's active learning approach emphasises the construction of knowledge in the learner's mind, rather than the learning and application of the content itself. Hattie (2013) rates Piagetian programs in early education far more effective relating to achievement than traditional methods employed in engineering, such as direct instruction, and even more progressive methods, such as problem-based learning and mastery learning.

Scaffolding

Mental models represent a subset of the real world. In order to move beyond the bounded rationality of mental models, active learning approaches are required to foster deep learning and understanding about the topic at hand. Learning is a social, collaborative process between peers and with the facilitator.

> The traditional school recognizes only the social exchange that is connected with a teacher, who is a kind of absolute ruler in control of moral and intellectual truth over each individual student. Collaboration among the students and even direct communication among them are thus excluded from classwork and homework (because of the examination atmosphere and 'grades' to be met). The active school, on the other hand, presupposes working in common, alternating between individual work and work in groups, since collective living has been shown to be essential to the full development of the personality in all its facets - even the more intellectual.
>
> – Piaget (1973, p. 108)

Vygotsky (1978; 1986) recognised learning and reasoning as a product of both a practical and social activity. The relationship between the learner's context and the learner's interaction within that context is fundamental to Vygotsky's sociocultural philosophy. An important part of this is the interaction between learners at different stages of development. More advanced learners can assist less advanced learners through situating learning just outside of their immediate capability, a location he described as the Zone of Proximal Development. The more advanced learner provides a scaffold and environment for effectively assisting a less-advanced learner.

This, of course, can occur on an individual or group level. Effective scaffolding in a group environment does not necessarily result in all members thinking the same thing: indeed, this would be disastrous for creativity. Rather, the sharing of scaffolded mental models can be used as an opportunity for individuals to expand their own understanding and assimilate relevant information into new mental models.

In a university setting, reframing the student body as a community of learners, and lecturers as facilitators, can have an enormous effect on how knowledge is constructed and mental models are shared. Here, I speak briefly from my short experience as a university lecturer, where I have done exactly this. My approach has been recognised with an Office of Learning & Teaching Australian Award for University Teaching in 2014.

In my systems engineering classroom, learning is distributed. In any given week, I facilitate a theory topic with a subset of students using a case study. These students collaborate in groups of 4-5 through a concrete learning activity. This activity has been designed by an undergraduate research student, whom I have worked with to design the activity and conduct research on its effectiveness (see Browne & Rajan 2015 for an example). The students become facilitators, and run their version of the lesson with their peers. The tutors in my class act as mentors. There is no exam, removing the poor learning behaviour that that mode of assessment encourages. Instead, students peer review each other's work throughout the semester and work towards a research portfolio on a topic that is relevant to them.

At this time my observations of the effectiveness of this approach are largely anecdotal; however, they are powerful. I often feel like I, as lecturer, could be removed from the whole process. I had one instance where a tutor had forgotten to attend a tutorial. Instead of the class sitting idle or perhaps leaving, the students agreed that they would send me a message, and forge ahead. When I arrived towards the end of class, the students offered me a video recording of the tutorial from a smartphone, and individual feedback sheets so that I could assess the facilitation. The students had become powerful advocates for their own learning. The power of a distributed learning approach turned what could have been an awful mistake into a fantastic learning experience.

This sits in stark contrast to tea room conversations with colleagues. Recently, a senior colleague talked at me about how he could not understand why his lecture attendance was so low, as he hadn't changed his approach to teaching the course in the many years that he had taught it. He also offered up his own critique of my approach: that students don't like it because they have to *think*.

I share this experience because it offers an example of why it is important to see how we create learning environments within our own bounded rationality. My colleague is stuck in the bounded rationality of his own mental models of what learning should look like, and at the same time my students were actively challenging and sharing their mental models by engaging in peer-facilitated concrete activities.

3.3.2 Learning & Teaching methods to facilitate experiential learning

I will use two similar terms in this section, and distinguishing between them is important. First is 'constructivism', a pedagogical approach where learners construct knowledge in their own mind. Second is 'constructionism', which is a pedagogical approach aligned with constructivism, but where physical interaction, such as making and tinkering, is used to facilitate the construction of knowledge. Where constructivism is discussed, both terms could be used interchangeably because of their underlying philosophy, whereas constructionism will be used when specifically discussing the process of constructing knowledge with physical objects.

Core to this section is the idea of *powerful ideas*. Seymour Papert used this phrase to describe ideas created through constructionism. Powerful ideas organise thinking and can be used for problem solving. They can be big ideas, such as the fundamental ideas that ground disciplines, or small ideas that become powerful because they are useful to the problem solver.

Newell (2012) uses Papert's term *powerful ideas* to describe simple, generic models that have an underlying metaphorical mapping. Examples are the bathtub metaphor or low-order stock-and-flow models. He notes that the most powerful dynamical metaphors are based on concrete experience.

Conceptual metaphors are a feature of many engineering disciplines, and provide students with an opportunity to think about otherwise abstract concepts. A common example is used in DC electronics, where the movement of electrons flowing from a positive terminal to a negative terminal is described as similar to the movement of water through a pipe. The experience of playing with the physical T&P system—water and pipes—provides the source domain for thinking about the concepts in the target domain.

In this section, I extend this idea of using conceptual metaphors for problem solving to using physical, transitional objects to think through problems. I use the term physical analogue to describe a transitional object that makes abstract thinking concrete. The interaction between the physical world and abstract concepts is fundamental to constructionism.

Transitional objects

Papert (1980) describes a transitional object as something that connects concrete experiences with formal reasoning, such as the reasoning required in mathematics. Papert introduces the term when describing the 'gears of [his] childhood'. His fascination of the physical gears aroused an interest in mathematics. Mathematical concepts, such as multiplication tables and algebra, were assimilated by thinking back-and-forth between the domain of the gears and the abstract concept. The physical experience with the gears gave him an inferential logic that allowed him to think through mathematical equations.

There are great opportunities to build shared conceptual repertoires using physical analogues. Newell describes the *Cuisenaire Metaphor*, based on the rods devised by Georges Cuisenaire. These rods have physical lengths between 1 and 10 centimetres, and can be used as a physical representation of numbers. Manipulation of the rods can be used to represent arithmetic operations. For example, addition operations can be modelled by placing the rods end-to-end and counting the units, and subtraction operation can be modelled by looking at the difference between two rods. Multiplication operations can be modelled by either placing the rods perpendicularly and summing the area projected by the two rods, or by counting the area covered by the multiples, as shown in Figure 3.9.

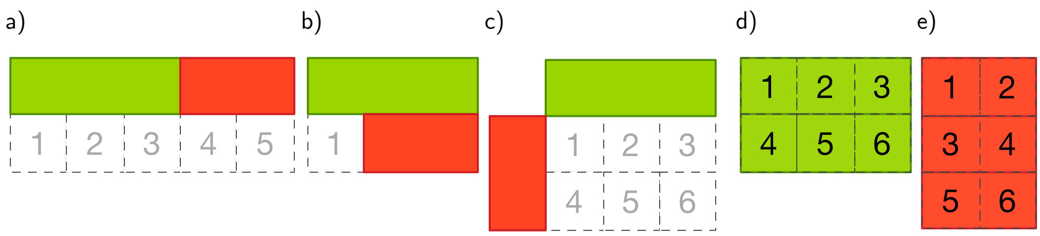


Figure 3.9: Diagram demonstrating how Cuisenaire rods can be used to explore arithmetic operations. The coloured rectangles represent the rods, and the numbered squares represent the count. Multiplication can be achieved by projecting the area (c) or covering the area (d and e). a) addition 2+3=5; b) subtraction 3–2=1; c) multiplication 2x3=6; d) two lots of 3 = 6; e) three lots of 2 = 6

Distinct from the rote learning involved in addition and multiplication tables, the Cuisenaire rods are a tool that can be used to think spatially about numbers. This gives the learner a strategy to think through an arithmetic problem without prior knowledge of the answer. The Cuisenaire rods are a transitional object for concrete learning.

Professor Stephen Cox has developed the use of a physical analogue in geology at the Australian National University to allow students to think through the behaviour of geological structures. The physical analogue uses coloured layers of salt in a large tank, much like a fish tank. A pneumatic rod moves one wall closer to the other, simulating the movement of tectonic plates but over a period of time measured in minutes rather than millennia. Fault lines begin to emerge, and students can predict the outcomes.

The logic of the salt tank becomes a tool for concrete learning when students are out in the field. Cox (2014) describes how students begin to think about geological structures in terms of the layers of salt. Further, graduate students experiment with different materials, such as honey, to make it generalisable to other geological structures. By doing so, they contribute to the development of the physical analogue as a transitional object for thinking through a range of geological deformation problems.

Papert's (1980) turtle geometry is another example of using a physical analogue to build understanding of mathematical concepts. The turtle can move forward, backward, left and right in the same way that that humans move on a plane. In this way, it is body-syntonic: one can use their experience with their own body in space to issue instructions to the turtle. The turtle is a transitional object controlled by programmed code that draws a trailing line as it moves. Programmers can think through the challenge of drawing shapes by thinking about their own physical experience of moving through space.

Consider Papert's Total Turtle Trip Theorem as a powerful idea. To instruct the turtle to draw a complete shape, it must return to its starting state. This means that according to the turtle, all shapes require the algebraic sum of turns to total 360°. With three equal turns of 120°[5], the turtle will draw an equilateral triangle, as shown in Figure 3.10.

In Euclidean geometry, the internal angles of a triangle add to 180°; a quadrilateral, 360°; a pentagon, 540°; a hexagon, 720°, and significant efforts are made to teach students these relationships. These values are soon forgotten by many students, and teaching these relationships is, in Piaget's words, an exercise in

5. Three equal turns of 120° (360° ÷ 3 = 120°)

once-knowing them. For the turtle, the movements are always relative to where it is situated. The Total Turtle Trip Theorem holds for any shape, including more advanced shapes, such as the five-pointed concave decagon, where angles that the turtle turns are both added and subtracted to a total of 360°.

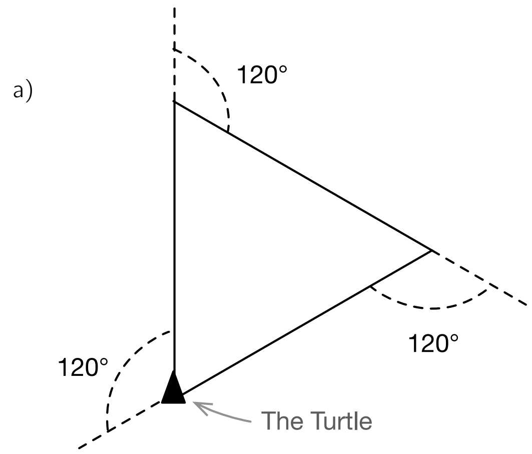


Figure 3.10: Drawing an equilateral triangle using Turtle and Euclidean geometry. a) Turtle geometry, the turtle moves forward 1 unit and turns 120° three times to arrive back where it started, totalling 360°. b) Euclidean geometry, the internal angles of the triangle equal 180°.

Inherent in any programming task is the ability to think in the same way that Richmond describes (§3.1), through simulating mental models. Papert (1980) describes the joy of students and teachers alike using the Logo programming language to move the turtle, and solving 'bugs' in their code when it didn't behave as expected. When the turtle does something unexpected, programmers are encouraged to *think like the turtle* in order to solve the problem.

The turtle is a concrete object for exploring the world in new ways. As a simulation device, the turtle gives instantaneous feedback about mental models expressed as programmed code. As a physical representation, the turtle robot allows the translation of code into physical drawings, and the exploration of robotics and design. As a character, the turtle becomes a collaborator in thinking and dialogue for problem solving (see Newell, 1988).

As an idea, the turtle needs to be cared for. The powerful idea exists only if the instructor allows it to, if the instructor creates an environment for students to think like the turtle. The instructivist teacher might provide the steps necessary to complete the task correctly in a worksheet. This could, perhaps, be a more efficient way of conducting an activity in the classroom, but an opportunity for learning and inquiry has been missed. The constructivist would pose a question that extends the student at each stage; for example, *How could you draw a circle? How could you use your knowledge of drawing a circle to draw a flower? Now, how could you draw a field of flowers?*

Like the other examples in this section, the concrete experience with the turtle gives the learner a framework to think about problems intuitively and creatively, without the need for memorising vast amounts of information. These analogues are tools for the concrete learning that can help learners develop an intuitive understanding of dynamical problems. This intuition is important to develop in our future engineers, who will need strategies to work through undefined problems without clear solutions.

3.3.3 Learning & Teaching methods to create dynamical connections

A constructionist approach places the learner in an active environment where mental models are developed as part of the synthesis of knowledge and interaction with the real world to build understanding. This places the learner actively in a position to exercise higher-order thinking skills.

Higher-order thinking skills

Developing higher-order thinking skills is important in a university setting. Many educational philosophers have described thinking skills along a continuum of lower-order to higher-order. Bloom's taxonomy (Bloom, 1956) is the most frequently used arrangement of thinking skills, with activities such as remembering, imitating and recognising sitting at the lower end and activities such as synthesis, construction and evolution of knowledge sitting at the upper end.

The role of the educational institution and the development of higher-order thinking skills is described differently by other philosophers. Piaget (1973) describes this process as *invention*. Dewey (1938) describes this progressive education as arousal of an active quest for information for the production of new ideas. Biggs (Biggs & Collis, 1982) describes these qualities in the evaluation of understanding as *generalising to a new domain*. Kolb (1984) describes this as *the construction of new knowledge*. Papert (1980) describes this as assimilation and construction of knowledge by the learner. Constructivist and experiential approaches encourage imaginative thought and the transfer knowledge between contexts by building understanding through experience.

With these concepts in mind, the dominant activities in university education stand out as ineffective in addressing challenges around effective communication, as discussed in §3.2. Take the university lecture hall, for example. Given Newell's iterative loop required for effective communication, it would be unrealistic to imagine that students left a lecture having completely constructed adequate understanding of the lecturer's intended meaning. Over four decades ago[6], Donald Bligh questioned the effectiveness of lectures:

> 1. The lecture is as effective as any other method for transmitting information, but not more effective.
> 2. Most lectures are not as effective as discussion methods to promote thought.
> 3. Changing student attitudes should not normally be the major objective of a lecture.
> 4. Lectures are ineffective to teach behavioural skills.
>
> - Blight (2000, p. 4)

Efforts have been made to address the technical problems of communication, such as by making lecture notes and recordings available. This may be reinforced through other class activities, such as tutorials or the summative comments given on marked assignments or tests, but the delays in this back-and-forth process and the inherent power relationship between lecturer and student make lectures a suboptimal activity for communication and learning.

Technology may provide solutions to these basic communication problems. There is recent evidence that the use of 'clickers', a tool that allows students to provide real-time responses to multiple-choice questions, in large lectures can promote active learning and help provide some feedback to both students and the instructor (MacArthur & Jones, 2008; Lantz, 2010; Buskes et al, 2010); however, there are also acknowledgments that lectures are "a relic of the passive teaching paradigm of the past" (Buskes et al, 2010). Others have created much more student-centred environments by providing opportunities to leverage learning as a

6. in his 1971 book

social process. Movements, such as problem-based learning and the flipped classroom, put students in charge of their learning by making them active in the learning process, through active problem solving and student-centred activities. Peer-to-peer teaching further encourages students to engage in the learning process, such as the student-planned and delivered tutorials (Baker 1996; Smith & Browne, 2013). These learning environments allow students to become active stakeholders in the quality of their learning.

Active learning loops

There is a strong link between the way that ST practitioners view developing mental models as a continuous process based on experience, the way that many L&T practitioners describe learning, and the continuous improvement approaches seen in systems and process engineering. Kolb (1984) describes the process of experiential learning as a cycle of continuous activity of processing and perception. The learner moves through four stages in an iterative loop:

- concrete experience: physical experience of something in the real world
- reflective observation: creating meaning out of the concrete experience
- abstract conceptualisation: distilling out the key, transferable ideas
- active experimentation: modification of decisions based on what has been learnt

Kolb's model is common to design and engineering practice, which can provide a useful parallel. Deming (2000) proposed similar cycles for process and quality improvement. His approach is often called plan-do-study-act[7]:

- plan: a process is selected
- do: a process is implemented
- study: the outcomes are observed and behaviour is understood
- act: modification of decisions based on what has been learnt

7. Plan-do-study-act. The original sequence was plan-do-check-act, but this was later changed by Deming to better describe his original intent: the analytical behaviour of studying rather than the process behaviour of checking

Using the Deming cycle as a model for process improvement was a major contributor to the boom in Japanese post-World War II manufacturing. Today, many similar frameworks exist, such as DMAIC[8] used by the Six Sigma movement, and TMI[9] used by Martinez & Stager (2013). Simplified models of Kolb's experiential learning cycle and the Deming cycle are shown in Figure 3.11.

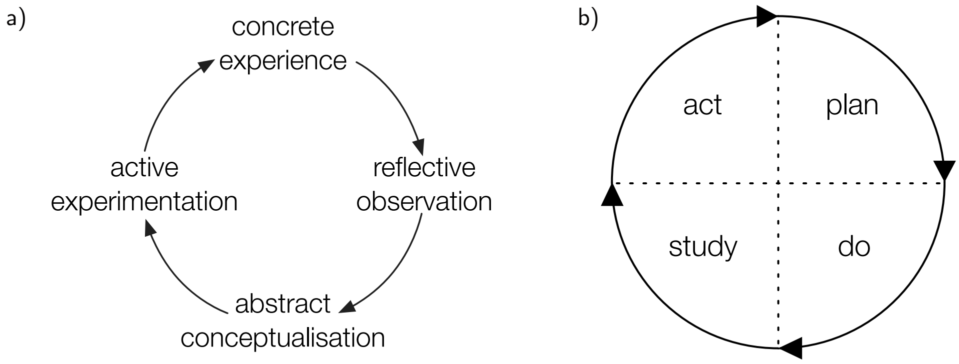


Figure 3.11: Continuous improvement loops. a) Kolb's experiential learning cycle shows processes in constructing knowledge b) the Deming cycle moves through four stages: plan-do-study-act

These cycles describe how learning can be improved and refined over time: the ideas can be applied at both an individual or group level. The experience of going through these cycles iteratively results in new levels of understanding. This cyclic experience provides the basis for improvement.

However, in these models, decisions about a particular system or problem are made based on observations and information from the real world. The decisions are influenced by strategies that are a result of the learner's mental models of the real world. If the learner is not receptive to changing their mental models, based on real-world observations and information, their learning is limited as it is not readily challenged. This limits the effectiveness of their decisions (Argyris & Schön, 1974).

An extension of these single-loop models is the double-loop model fundamental to the design of the T&P activity, described in Figure 1.1 and Figure 3.2. In double-loop learning, the learner embraces information feedback and modifies his or her mental models of the real world. This interaction closes a second feedback loop,

8. Define, Measure, Analyze, Improve and Control

9. Think, Make, Improve

and makes the learner more effective. Argyris (1976) identifies two factors that inhibit learning: the ability to measure the effectiveness of decisions (the first loop), and; the receptivity to corrective feedback (the second loop).

This observation places learning and the accumulation of knowledge in a dynamical context, where decisions are made, action is taken, outcomes are observed, mental models are revised and new decisions made based on the revised mental models. The key difference between single- and double-loop learning is that mental models are explored, challenged and improved continuously in the latter. This offers insights into the design of the T&P activity. First, that the activity should provide an experience that promotes an opportunity to engage in iterations of learning cycles, and, second, that there is an opportunity to challenge the mental models that participants bring to the activity.

3.3.4 Application of Learning & Teaching perspectives to the T&P activity

In this section, I summarise the L&T perspective by providing a pre-emptive outline of how the L&T perspectives and approaches described here relate to the T&P activity, which is described in detail in Chapter 4.

Creativity and Scaffolding

In the T&P activity, the instructions have been carefully constructed to gradually increase in difficulty. Early tasks orientate participants to the physical functioning of the system, and then tasks build in cognitive difficulty. The tasks in the T&P activity are described in §4.2. Answers to prompts are open-ended, and playing with the physical system is encouraged, allowing for scaffolded, creative thinking.

Transitional objects and concrete learning

As discussed in §2.3.5, the T&P system represents a physical analogue of the carbon cycle. The pumping of water between tubs using hand-held pumps is a physical experience of stocks and flows. Similar to Papert's Turtle, participants can use the T&P system to think through the problems they are given through the activity, and use the T&P system to physically simulate alternative strategies around the problems.

Higher-order thinking skills and active learning loops

The goal of the T&P activity is for the participant to be able to transfer their knowledge from the activity to the real-world, thus engaging in higher-order thinking. In the design of the T&P activity, careful selection of prompts were used to encourage multiple cycles around learning loops, and indeed double-loop learning. The prompts are described in §4.2. Treatment conditions were designed around the use of different prompts, including the use of scenario cards, group diagrams and pictorial representations.

3.3.5 Constructing knowledge in engineering education

Here I briefly summarise the L&T perspectives in relation to university engineering education. Creating environments where student engineers can develop their mental models in meaningful ways requires understanding what we consider to be the goal of a university engineering education.

The technology that engineers engage with and use to solve problems changes over time. Consider the speed at which telecommunications have developed over recent decades. New generations of communication infrastructure are continually being researched, developed, provisioned, and replaced. However, little has changed in many aspects of engineering pedagogy over recent history, despite the massive leaps forward in technology.

The process of learning is more important than the content we teach student engineers. Trevelyan (2010) notes that major parts of engineering practice are the rules-of-thumb that engineers use as part of their vernacular. This draws parallels with the earlier observation from Maxwell et al (1994) that decision-makers carry ‘causal chunks’ around in their minds. These rules-of-thumb come from a range of experiences. Trevelyan (2014) further argues that important skills required in professional practice, such as life-long learning, are not learnt at university. This is a key challenge for engineering educators.

In reality, engineering professional practice is team-based. Senge (2006) describes three core learning capabilities for a team: fostering aspiration, developing reflective conversation, and understanding complexity. Fostering aspiration can be built through personal mastery (technical skills development) and having a shared group vision. Reflective conversation calls for the sharing of mental models and dialogue. Systems thinking is a useful tool for understanding complexity.

Engineering faculties have an important role to play in preparing our students for the transition from young adults to graduate engineers. A university education should equip students with the working knowledge and technical vocabulary to enter the workforce, but more importantly equip students with the disciplinary tools they can use to grapple with complexities regardless of their final vocation. However, it is still the case that:

> Too much pedagogy is concerned solely with the transfer of information. Even an education directed towards immediate vocational ends is less than it could be, and graduates are left with less potential than they might have, if it fails to engage the student in grappling with uncertainty, with deep underlying issues and with context.
>
> - Boulton & Lucas (2008)

A university engineering department should be a place where a student's mental models are continuously challenged, and where double-loop learning is integrated and continuous in order to prepare graduates for the challenges of addressing complex real-world problems.

3.4 Towards a double-loop learning activity

The purpose of this chapter was to establish considerations for the pedagogical approach behind the T&P activity. To summarise this chapter, the model in Figure 3.2 is revisited and instantiated into a learning environment for the T&P activity in Figure 3.12. The three loops: *physical simulation, scaffolded refinement* and *focussed dialogue* represent a synthesis of the concepts and ideas described in this chapter.

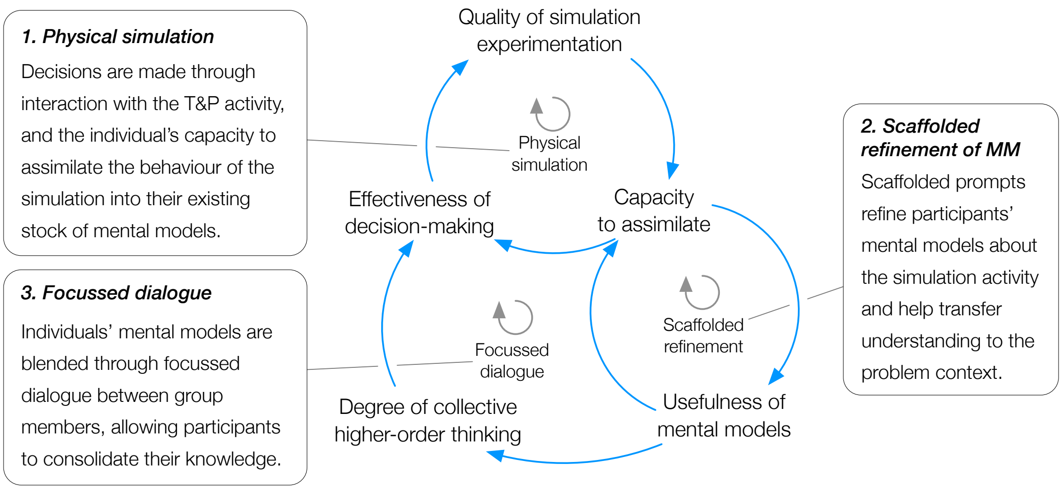


Figure 3.12: Annotated influence diagram of a double-loop learning model for the T&P activity.

At this stage, I will discuss the diagram in terms of a general philosophy for the T&P activity. The learning activity should aim to achieve the following:

- **Physical simulation:** the activity should be the source domain for understanding the carbon cycle. Interactions with the simulation environment should be analogous to the behaviour of the real world, and easily transferred to the target domain
- **Scaffolded refinement:** consideration of students' existing knowledge is important so that students can readily assimilate the new knowledge with their existing mental models. The act of physical simulation should help improve mental models of the simulation environment, and be used as a platform for testing new mental models.
- **Focussed dialogue:** the activity should prompt students to critically evaluate or reflect on the assumptions and relationships in the activity, and help them to collectively draw meaningful connections between the activity and the real world.

The double-loop learning model for the T&P activity shown in Figure 3.12 will be explored further in Chapter 6.

PART II RESEARCH APPROACH

Preface

In Part II, the workshop methodology and results are presented, representing the majority of the original research in this thesis.

Before I discuss the methodology, I will briefly outline the specific terminology I use in this chapter to describe the research activities. An overview of these activities are shown in Table 4.1.

- **Experiments** refer to discrete sets of similar workshops, targeted at answering a research question. Three experiments were conducted, referred to as the Checklist, Simulation and Dialogue experiments. The Simulation and Dialogue experiments used a Workbook as instructions, and are collectively referred to as the Workbook experiments.

- **Treatment conditions** refer to deliberate differences within experiments. The Checklist experiment had treatment conditions with respect to the assessment task; the Simulation experiment had treatment conditions with respect to physical and mental simulation, and; the Dialogue experiment had treatment conditions with respect to different instructions that prompted dialogue within groups.

- **Assessment** refers to the tasks in assessment sheets used to examine decision-making in the activity. There are three distinct assessment sheets that were developed: Emissions & Removals (ER), Anthropogenic Emissions (AE), and; Anthropogenic Additions (AA).

- **Instructions** refer to the instructions used by the participants to guide the T&P activity. There are two distinct instruction types: checklist and workbook. These terms are also used to describe the two categories of experiments referred to in the experiments section. The term 'workshop instructions' is used to describe the additional instructions in a workshop, such as the introductory video.

- **Tasks** refer to individual instructions in the workshop; for example, "Ask the Natural Emissions to pump to a level of half the capacity of the atmospheric stock, then stop".

- **Groups** refer to a grouping of 3-6 students that manipulated one T&P system as a team. There were typically 3-5 groups in one workshop session.

- **Workshops** refer to the experience that a participant had completing the task. In all instances, a workshop could be described as a hands-on tutorial.

In Chapter 4, I present the workshop methodology, including a description of the T&P system, workshop instructions, treatment conditions and assessment tasks. This leads to Chapter 5, where I present the results from the assessment tasks by treatment condition, course cohorts, assessment sheets and profile data.

4 Workshop methodology

Recent trends in system dynamics aim at changing those mental models that people use to represent the real world. To do this a person must become sufficiently involved in the modeling process to internalize lessons about dynamic feedback behavior. Such exposure to dynamic thinking should start at an early age before contrary patterns of thought have become inflexibly established.

- Jay W. Forrester (1989, p.13)

The workshop methodology is described through key aspects of the double-loop learning model proposed in Figure 3.12. The T&P system is central to the *physical simulation* loop, and is presented in §4.1. The development of workshop instructions are central to the *scaffolded refinement* loop, and are outlined in §4.2. The *effectiveness of decision-making* variable is the point where the assessment tasks fit, and are explained in §4.3. Finally, in §4.4 I discuss the differences in treatment conditions through the *focussed dialogue* loop.

The three research questions introduced in §1.4 are applied specifically to the T&P activity to address SF failure below:

Experiment 1: Scaffolded refinement
Does display of instructions and information in the T&P activity reduce SF failure?

Experiment 2: Physical simulation
Does representational form of the T&P activity reduce SF failure?

Experiment 3: Focussed dialogue
Do opportunities for focussed dialogue during the T&P activity reduce SF failure?

These research questions led to three distinct experiments, referred to as the Checklist, Simulation and Dialogue experiments. An overview of the experiments is shown in Table 4.1.

Table 4.1: Overview of experiments with respect to aspects of the double-loop learning model

Experiment	Simulation	Instructions	Assessment	Dialogue	Treatment
Checklist (Scaffold)	Physical	Checklist	ER	-	PC-ER
	Physical	Checklist	AE	-	PC-AE
Simulation	Physical	Workbook	AA	-	PW-AA
	Mental	Workbook	AA		MW-AA
Dialogue	Physical	Workbook	AA	Group Diagram	PW-GD
	Physical	Workbook	AA	Scenario Cards	PW-SC

Note: the Dialogue experiment was a 2x2 factorial design. Hence, there was a PW-AA treatment (without either GD or SC) and a PW-GDSC treatment (with both GD and SC) in addition to the single-factor treatments listed in the table.

Workshop Terminology

After the Checklist experiment, it was clear from observations and results that there was confusion about the activity because of the terminology used in the instructions. Particularly, there was confusion around the terms 'emissions' and 'absorptions'. In the later experiments, this terminology was changed to 'additions' and 'removals' respectively. These changes in terminology also affected the T&P system, workshop instructions and assessment tasks, and will be described in the discussion of variations in each relevant section.

4.1 The T&P system

The T&P system was developed from a simple activity that Barry Newell and Katrina Proust developed to help explain stocks and flows to community leaders on the south coast of New South Wales, Australia, involving buckets of water and pumps.

In the T&P system, participants physically manipulate pumps to change the state of the system, and observe the physical changes to the stocks over time. A photo of participants interacting with the system is shown in Figure 4.1. The analogue provides a concrete source domain for the bathtub metaphor. For the participants, the learning is tacit. The physicality of the system encourages a sensorimotor experience of stocks and flows, allowing participants to develop a shared conceptual repertoire about the cause-and-effect logic of the system in the process of playing with it.

Figure 4.1: Photo of participants manipulating the T&P system. Water is pumped between two tubs in a system analogous to the flow of carbon in the carbon cycle.

I will discuss three key aspects of the T&P system as the physical simulation activity in this section: the variations between experiments, the dynamic hypothesis, and a description of the metaphorical mapping between the T&P system and the carbon cycle.

4.1.1 Variations in the T&P system between experiments

The description of the T&P system that follows in this chapter generally refers to the system as described in the PW-AA treatment group[10]. The T&P system itself remained unchanged through the three experiments; however, here I note the key differences to the way that participants interacted with the T&P system between the Checklist and Workbook experiments.

10. See Table 4.1 for an overview of treatment conditions

Assembly

In the checklist experiment, the T&P system was pre-assembled, filled with water, and ready for the activity. Components of the system were laid out on a printed absorbent mat, as shown in Figure 4.2. Once groups had formed and started to begin the activity, I visited each group and asked each pump operator which pump they were controlling to ensure that the individuals within the groups were clear on their roles. If a group did not complete this identification task correctly, I asked the group to discuss their identification further, and checked in with the group before they started the activity.

Figure 4.2: Photo of T&P system on the printed absorbent mat. Labels were used to identify the components, and the printed absorbent mat had labels to further reinforce the names and relationships of the components.

There were two minor issues with having the system set up for the participants. The first was practical: with the system set up ahead of time, many groups would start playing with the pumps immediately. I would often arrive at a group to check their roles only to find that their system was in the final state analogous to catastrophic climate change and that nobody had read the instructions yet. Although play is an important part of learning, I was concerned that this was not a useful approach for learning.

In the Workbook experiment, I put all of the components into a bag for each group and the groups were required to lay out their system on the absorbent mat and assign labels to all the system components. The pumps and pipes remained connected for ease of use, though this meant that at times the pipes were tangled and groups required assistance. Once groups had correctly labelled the system, they were given water and could commence the simulation exercises. It was intended that the process of identifying the parts of the system would help participants' understanding of the mapping between the source and target domains.

Terminology

The terminology change between the Checklist and Workbook experiments meant that the labels and mat used in the T&P system were changed. The labels were used to identify components. The mats acted as a placeholder for the tubs, as well as a way to soak up wayward water. The natural and anthropogenic emissions pumps were renamed to natural and anthropogenic additions pumps, and the natural absorption pump was renamed to the natural removals pump. The mats used in the activity are shown in Figure 4.3, demonstrating the difference in terminology.

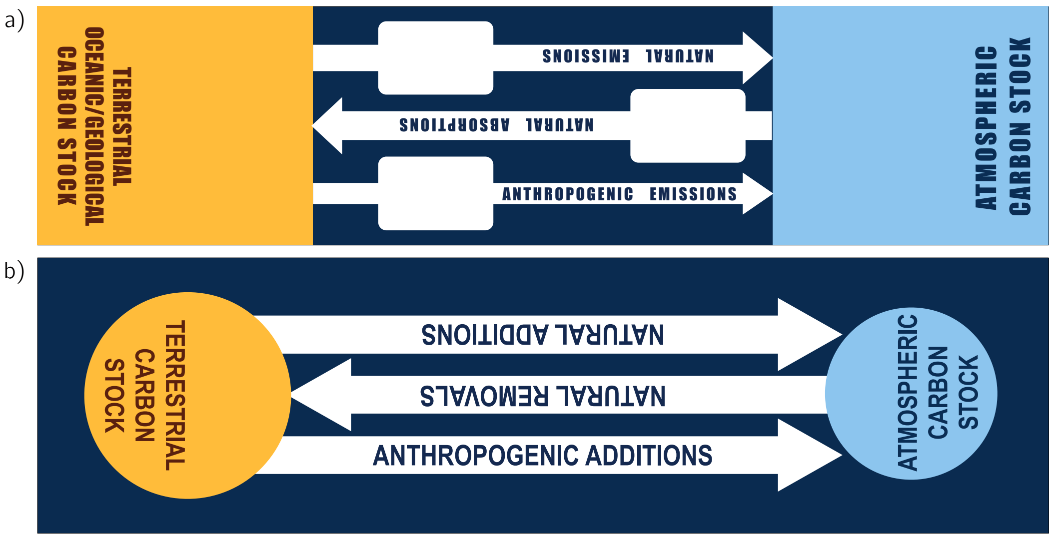


Figure 4.3: Printed absorbent mats used for the T&P activity with different terminology. a) the mat used in the Checklist experiment; b) the mat used in the Workbook experiments

4.1.2 Dynamic hypothesis

As described in §3.1.1, ST practitioners define the boundaries of a problem or system using a model boundary chart. Building on the description of carbon bathtub dynamics given in §2.1.1, Table 4.2 shows the endogenous, exogenous and excluded variables in the physical analogue.

Table 4.2: Model-boundary chart for design of the T&P system. The endogenous and exogenous variables are considered in the T&P system. The excluded variables are not considered further in the model.

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

The stock-and-flow representation of the T&P system shown in Figure 4.4 is made up of two stocks and three flows, as described in the model boundary chart in Table 4.2. Unlike the bathtub stock-and-flow model shown in Figure 3.6, the T&P system is closed; that is, the amount of carbon in the system observes conservation of mass. Terrestrial carbon can move to the atmospheric stock through the natural and anthropogenic additions, and atmospheric carbon can be returned to the terrestrial stock through the natural removals. As discussed in §2.1, there is no anthropogenic removals flow, as this flow is currently negligible, and will remain so in the foreseeable future.

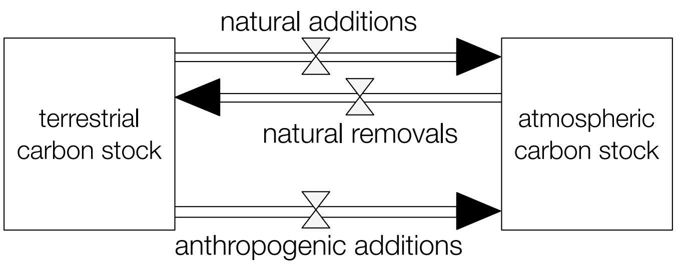


Figure 4.4: Stock-and-flow representation of the T&P system. Carbon (water) moves from the terrestrial carbon stock to the atmospheric carbon stock through the natural and anthropogenic additions flows, and is returned through the natural removals flow.

The stock-and-flow model in Figure 4.4 is shown as a diagrammatic representation of the T&P system in Figure 4.5. Participants manipulate the pumps, which are connected to tubs of water using pipes. The pumps that represent the natural and anthropogenic additions transfer water from the tub that represents terrestrial stock to the tub that represents the atmospheric stock. The pump that represents natural removals transfers water from the tub that represents the atmospheric stock to the tub that represents the terrestrial stock.

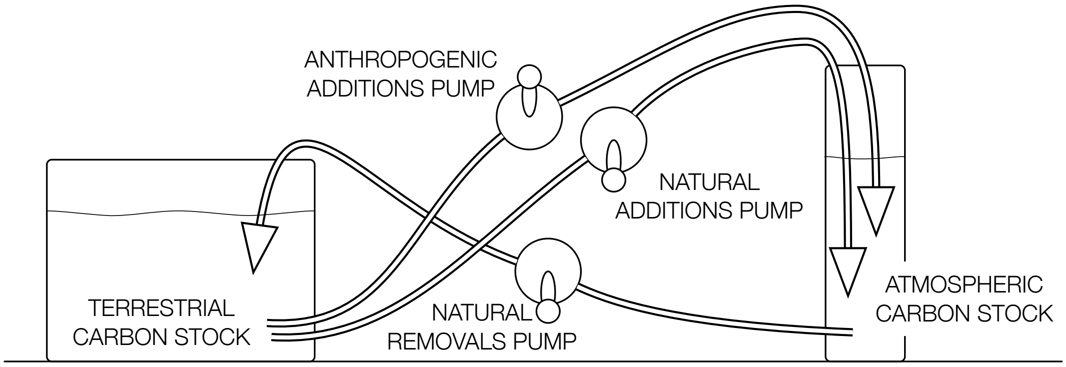


Figure 4.5: Diagram of the T&P system. Participants manipulate pumps which change the state of the stocks. Coordination is required by team members to simulate the system.

In a stock-and-flow simulation model, the behaviour of the model is governed by relationships and formulae. In the T&P activity, the behaviour is governed by the relationships and instructions given to the participants. Unlike a simulation model, the instructions in the T&P activity can be ignored or readily modified by the participants as they experiment with the physical system.

Dynamical behaviour

The behaviour of the T&P system through the instructions follow a simple scaffolded structure, described further in §4.2. The two dynamical relationships of interest in the activity are the dynamic equilibrium of the natural carbon cycle, and then the anthropogenic perturbation to the natural carbon cycle as the anthropogenic additions pump speeds up.

In the dynamic equilibrium situation, the natural additions pump and natural removals pump must remain approximately equal. This behaviour is shown in Figures 4.6a and 4.6b. Small fluctuations occur naturally in the interaction of the participants pumping. Although not part of the formal T&P model, this fluctuation has been described by participants as analogous to seasonal variations due to biomass respiration.

In the anthropogenic perturbation behaviour, the anthropogenic additions pump speed that is called for in the activity instructions is shown in Figure 4.6c, and the resulting observed water level in Figure 4.6d. The y-axis values are not given, as they are relative to the speed of the pump. For example, in Figure 4.6a, the pump speed could be slow or fast, and as long as the average speed is the same, then the level will remain the same.

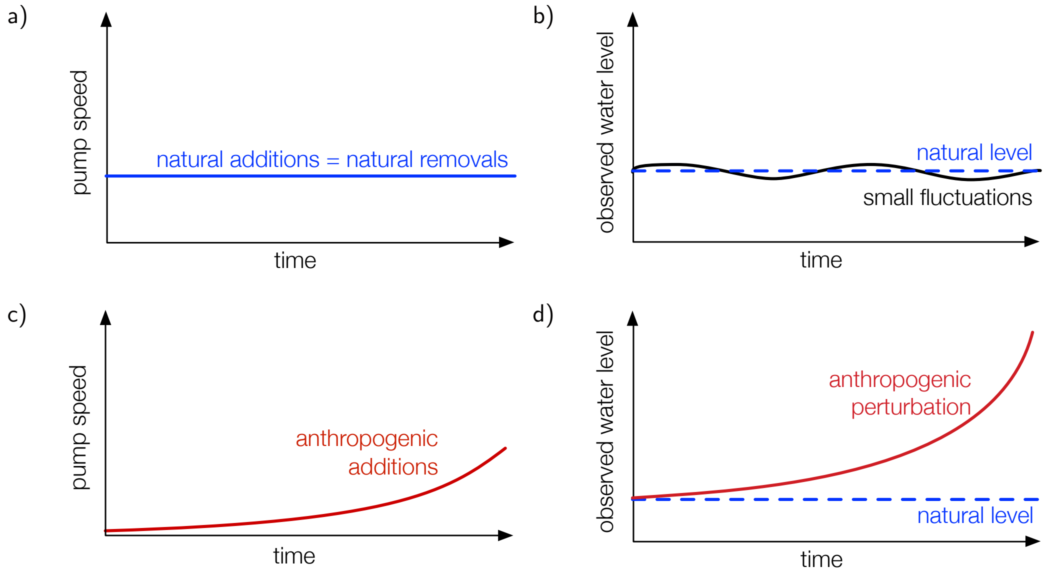


Figure 4.6: Expected behaviour over time graphs for physical simulation. According to the instructions, natural additions must be kept approximately equal to natural removals (a). The reality of the physical system means that small differences between the two natural pumps typically lead to an oscillating behaviour (b). The anthropogenic additions pump disrupts this balance (c), and increases the observed level of water (d).

The final instruction in the T&P activity is to explore (a) the "what-if" scenarios given the situation presented in Figure 4.6d and (b) the challenge of stopping the water level from rising. The group uses the T&P system to explore their strategies. With the constraint that the natural cycle pumps must be pumping on average at the same speed, the only solution is to reduce the anthropogenic additions pump speed to zero to maintain the observed water level.

4.1.3 The T&P conceptual metaphor

The T&P system can be used as a tool for concrete learning because of the metaphorical mapping between the model and the real-world system. Participants can manipulate the T&P system (the conceptual source domain with the water, tubs and pumps), and then make valid inferences in, and construct an understanding of, the target domain (the carbon cycle). The explicit mapping of the concepts in the conceptual source domain to the conceptual target domain is shown in Table 4.3.

Table 4.3: Conceptual mapping of the T&P metaphor. The components in the conceptual source domain are mapped to the conceptual target domain, allowing participants to construct an understanding of one system (the carbon cycle) in terms of another (tubs and pumps). Note: * the arrows "⇒" represent the expressions "corresponds to" or "maps to".

Conceptual source domain: tubs & pumps		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 the burning of 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

As described in §3.2.3, the mapping between the conceptual source and target domains is partial. For participants, this partial mapping may lead to inferences that become areas of confusion. Potential areas of confusion are described in Table 4.4. I do not necessarily believe that the issues listed in Table 4.4 are problems to be fixed in future versions of the activity, but rather are listed for the purpose of further describing areas of potential confusion.

The simplification of the carbon cycle in this representation does, however, make it easier to see the importance of the balance between the key flows of the carbon cycle, which is an important goal of the activity. The model as presented here is not designed to be a simplified representation of the complex carbon system, but rather to demonstrate the effect of unbalanced carbon flows.

Table 4.4: Areas of potential confusion in the conceptual mapping due to both physical limitations and conceptual mapping. Efforts to reduce the effect of these limitations have been made to reduce the opportunity for confusion.

Concept	Description
Physical limitations of the T&P system	
Size of stocks	The terrestrial carbon stock is many orders of magnitude larger than the atmospheric stock. In the T&P system, it is only marginally larger
Feedback	Feedback effects and mechanisms from the conceptual source domain are not present, such as the effect of temperature on carbon storage in the ocean. There is, however, feedback in the system, such as the change of participant behaviour due to the level of the water or the sound of the pumps as they are manipulated at different speeds
Changing rates	The natural carbon cycle fluctuates due to a number of subprocesses and feedback mechanisms; the rate of anthropogenic additions continues to increase. In the T&P system, these changes are not controlled in a regulated way
Homogeneity	The terrestrial carbon stock is not a homogenous entity as it is represented in the T&P system. For example, in the real world, the oceanic stock has a different relationship with the atmospheric stock than with the terrestrial stock.
Conceptual mapping limitations	
Colour of the water	The water is intentionally coloured green to aid visualisation. Some participants remarked that the green was chosen to represent greenhouse gases
Water leaking out of the system	On occasion, water leaves the system through leaks in the connections.
Upper limit of the atmospheric tub	The 'overflow' level of the atmospheric tub. When exceeded, water leaves the system.
Feedback in the system	The natural carbon cycle is influenced by the natural capacity for carbon flux; for example, the more trees available to remove carbon from the atmosphere will also result in higher levels of carbon additions in the future as the organic matter decomposes. There are no such built-in feedback mechanisms in the model.
Reversibility of pumps	Pumps in the T&P system can go in both directions without regard of the feedback processes. To mitigate this possibility, inflow pipes have been placed at the tops of stocks, and outflow pipes have been placed at the bottom of stocks to reduce the likelihood of this behaviour
Water as carbon	Participants could confuse the carbon cycle with the water cycle

The physical limitations of the equipment used means that it is not practical for the model to be scaled accurately. For example, a terrestrial stock that is 50 times larger in volume than the atmospheric stock, or an anthropogenic pump that is 30 times smaller than the natural pumps, is impractical (see Figure 2.1 for relative sizes). This is not a significant limitation, given that the main aim of the T&P activity is to help students understand the impact of unbalanced flows. Removing these elements and considerations from the simple model also means that less prerequisite knowledge about the relationships and behaviour of the system is required to understand the activity, and in turn the model.

4.2 Workshop instructions

In this section, I describe the instructions for the workshop. Activities were typically confined to 45 minutes of a 1-hour tutorial slot. The activity was typically completed in groups of 3-6, and the assessment task was completed individually. Groups were largely autonomous, and the only instruction offered by the facilitator was to advise groups to complete tasks faster or slower to match the progress of other groups and overall time considerations.

Once groups had completed the activity tasks, an assessment sheet was handed out and completed by participants, which was used for the results presented in Chapter 5. After the assessment sheets were completed and collected, a brief presentation was given showing typical answers followed by a general discussion of insights arising from the activity. No personally identifiable information was collected, other than basic demographic profile data.

All the workshops followed a generic structure and schedule, with small variations to accommodate the mechanisms of the different treatment conditions. Due to the limitations of completing this task in a real, scheduled tutorial, the activity had to fit into the available time. The generic structure of all the experimental workshops are shown in Table 4.5.

Table 4.5: Generic structure of workshops

Component (time)	Checklist experiment	Workbook experiments
Introduction *~3 minutes*	Welcome, overview, human ethics explanation and discussion	
Video introduction *~2-8 minutes*	Detailed (7min 45sec)	Physical simulation (1min 20sec) Mental simulation (1min 20sec)
Group formation *~1 minute*	Proximity groups	Random allocation
Assemble model *~1-5 minutes*	System preassembled, allocate roles	Physical: Label parts, allocate roles Mental simulation: Label picture
Activity prompts *~10-15 minutes*	Instructions (dot points)	Workbook prompts modified for all treatment conditions
Assessment task *~5-10 minutes*	Complete individual assessment task and profile survey	
Remaining time	Facilitated discussion about various strategies and consequences	

Note: Workshops were typically confined to the latter part of a two-hour tutorial. Components in bold will be discussed further in this section.

The introductory component was the same across all workshops and treatment conditions, in line with the human ethics clearance process. Students were made aware of the option to withdraw from the activity at any stage without it affecting their standing within the course. The key differences in the video introduction, group formation and activity prompts will be discussed in §4.2.1-§4.2.3. The differences in the assessment task will be discussed in §4.3.

4.2.1 Video introduction

To control for consistency across the workshop sessions, a video introduction was given to provide a small amount of background for the participants. The video was captioned to aid comprehension. Two similar videos were prepared to complement the workbook instructions: one for the physical simulation and one for the mental simulation. A transcript of each video is presented in Appendix C.

The video in the Checklist experiment presented the problem of global climate change, the carbon cycle, a detailed breakdown of the T&P system, the activity, and the challenge within the activity. The video ran for 7min 45sec.

For the Workbook experiments, a revised version of this video was shown, which explained briefly the activity, the simplified carbon cycle and the challenge for the activity. The second video ran for 1min 20sec. The text in the second video mirrored the text used on the first two pages of the workbook to ensure that all participants had been given this information before beginning the activity.

A third video was prepared for the mental simulation treatment condition in the Simulation experiment. The text was as similar as possible between the two treatment conditions. Stills from the videos used in the physical and mental simulation treatment conditions are shown in Figure 4.7.

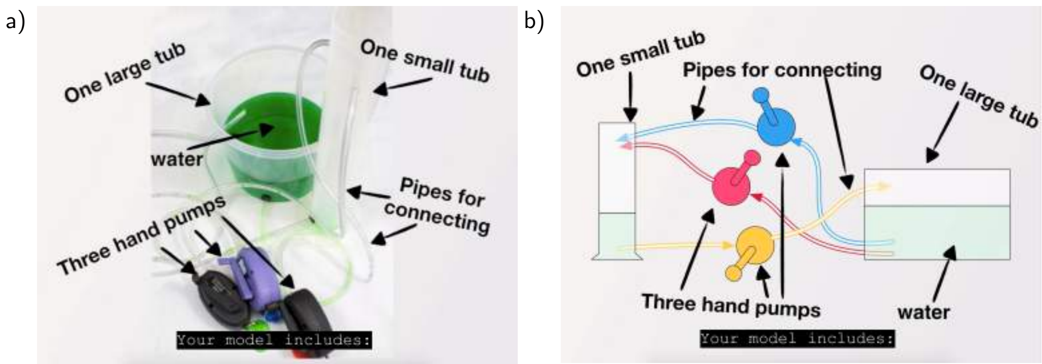


Figure 4.7: Stills from the introductory video in the simulation experiment at 0min 14sec. a) still from the physical simulation treatment video; b) still from the mental simulation treatment video

Upon completion of the video, groups were formed and the activity commenced.

4.2.2 Group formation

The T&P system required at least three sets of hands to operate, with groups of four desirable allowing for a coordinator. A maximum of five groups could be formed due to availability of experimental setups. Smaller groups (i.e. four) were preferred over larger (i.e. six), but this was not considered to be a critical control variable and one that needed to be flexible as the experiment was being conducted in real tutorials. A summary of group sizes is shown later in Table 4.14.

In the Checklist experiment, groups were formed based on proximity in the tutorial room, putting people sitting together in the same group. I observed that this method had three issues worthy of considering. First, it appeared that participants sitting together were more likely to be friends, and this sometimes led to misadventure in the activity. Second, although collaboration was against the instructions given for the assessment task, I observed that apparent collaboration was frequent. This was inevitable in a classroom that was not invigilated; however, I thought that if groups were less likely to be friends, then discussion might reduce. Finally, the collation of the results into groups for data entry required careful sorting and checking when collecting the results, and meant that time was consumed cross-checking for groups during input of data.

In the Workbook experiments, I used a coloured card system to allocate groups and collate responses. Cards displayed a unique alphanumeric value [A-Z, 2-8] on a coloured background in a rotating colour sequence [red, blue, green, yellow, purple][11]. This meant that responses had a unique identifier that could be matched to their group, and were not allocated to a group of people immediately adjacent to them. Groups assembled and completed the task near the system that matched their colour. A diagram demonstrating the two allocation methods is shown in Figure 4.8.

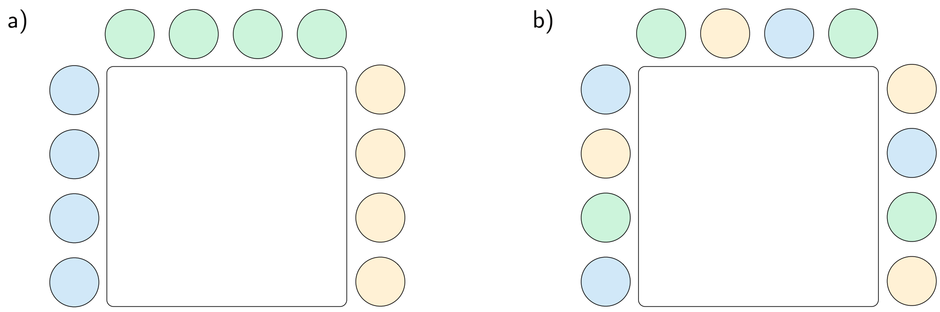


Figure 4.8: Group formation methods. a) groups were formed using proximity groups in the Checklist experiment; b) groups were formed through sequential allocation in the Workbook experiments.

It is not clear whether this change in allocation methods made any difference to a group's performance, nor was the experiment designed to investigate this. Observation suggests that it meant that groups were less likely to know each other well in the latter method. However, it did solve an administrative problem when collecting results, as individuals were more conscious of their identifier, which was used for coding responses.

11. i.e. A=red, B=blue, C=green, D=yellow, E=purple, F=red, G=blue, etc.

4.2.3 Activity instructions

In the Checklist experiment, the instructions were an A4 sheet with nine dot-point tasks. In the Workbook experiments, the instructions was a short workbook with four or five tasks depending on the treatment condition. All the activity instructions are included in Appendix A. The Checklist and Workbook experiments also followed the terminology difference outlined in §4.1.

The change in presentation format of the instructions between experiments allowed for different visual representations. The first two tasks are used as an example below to demonstrate this difference. These tasks were designed to orientate the group to the function and direction of the pumps by having one additions pump operate, then one removals pump operate. The instructions from Checklist experiment were:

> 1. Ask the Natural Emissions to pump to a level of half the capacity of the atmospheric stock, then stop
>
> 2. Ask the Natural Absorptions to drain the atmospheric stock completely, then stop

The workbook format allowed for the use of behaviour over time graphs throughout the activity to further visualise the distinction between stocks and flows. The equivalent tasks in the Workbook experiment are shown in Figure 4.9.

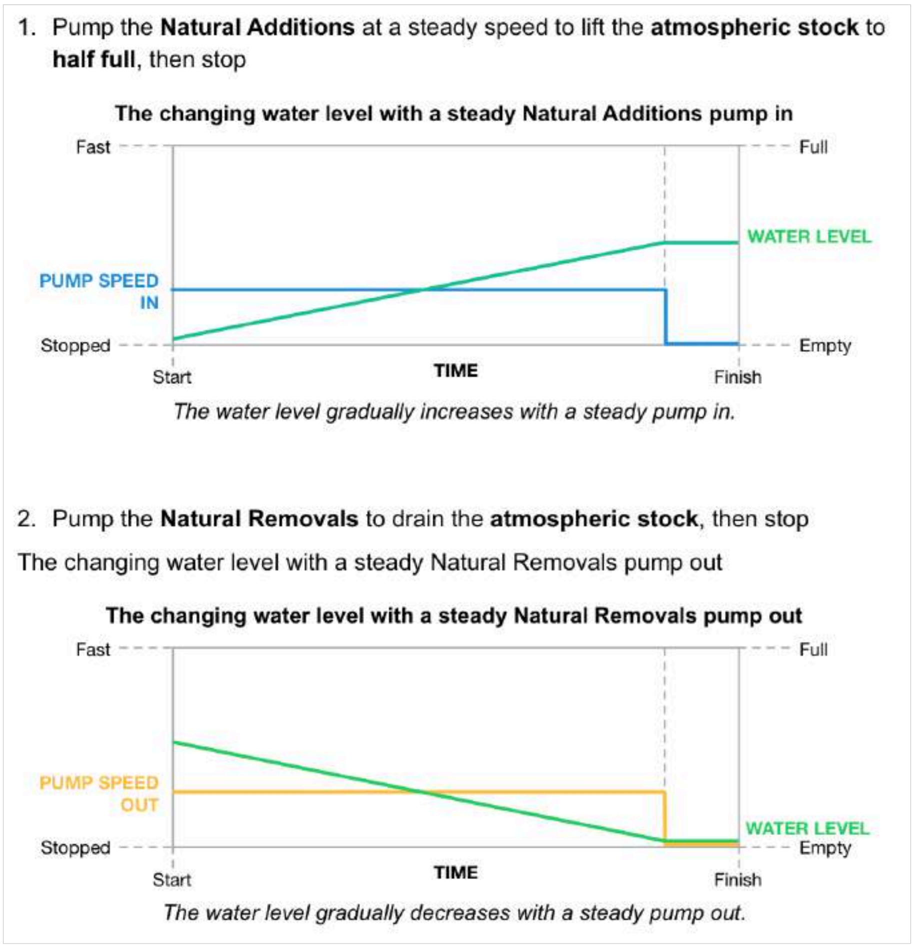


Figure 4.9: Tasks 1 and 2 in the Workbook experiments, showing a written description and graphical representation of the behaviour over time.

The instructions across experiments and treatment conditions followed a similar sequence, mapped out in Table 4.6. In the Dialogue experiment, tasks 4 was separated into two tasks, to accomodate the scenario card and group diagram treatment conditions.

Table 4.6: Sequence of workbook instructions by experiment. All experiments followed the same general scaffolded pattern.

		Workbook	
Task	Checklist	Simulation	Dialogue
Label system, allocate roles	Pre-task	Task 1	Task 1
Natural additions pump test	Task 1	Task 2	Task2
Natural removals pump test	Task 2		
Anthropogenic additions pump test	Task 3		
Natural cycle dynamic equilibrium	Tasks 4-5	Task 3	Task 3
Simulate anthropogenic disruption	Tasks 6-7	Task 4	Task 4
Discussion and conclusions	Tasks 8-9		Task 5

In the Checklist experiment, the instructions were laminated and re-used throughout all the workshops. In the Workbook experiments, the instructions were printed with areas for the groups to complete written and graphical tasks. Each group received a fresh, stapled copy that corresponded with their treatment condition and these were collected for consideration in the results and analysis.

4.3 Assessment tasks

In this section, I describe the assessment tasks that accompanied the activity. The results from the assessment tasks are the primary evidence used for the results in Chapter 5. Table 4.7 outlines the assessment task and description for each of the treatment conditions. Two different assessment sheets were used in the Checklist experiment, and one assessment sheet was used for all of the Workbook experiments.

The assessment tasks were designed to take no longer than 5 minutes to complete. The purpose of the assessment task was to gain insights into the participant's understanding. All assessment tasks involved an open-ended written description and a graphical response. The Checklist experiment used specific numbers in instructions and graph, whereas the Workbook experiments used relative terms in instructions and graph, such as half-full. The Workbook experiments also included a causal-logic task and a group agreement Likert scale. The attributes of the three assessment tasks are summarised in Table 4.7.

Table 4.7: Components of assessment tasks for the three different assessment sheets

Assessment Task	Written instructions and response	Graphical task	Other
ER: Emissions & removals	Detailed	Numerical	NA
AE: Anthropogenic emissions	More detailed		
AA: Anthropogenic additions	Simple	Relational	Causal logic & group agreement

Each of these attributes—written instructions and response, graphical task, causal logic and group agreement—will be discussed and illustrated in this section. The complete assessment sheets are shown in Appendix B.

4.3.1 Written instructions

The written instructions were used to guide the responses in the assessment task. The instructions in emissions & removals (ER) assessment were modified for the anthropogenic emissions (AE) assessment to direct responses towards a reduction in anthropogenic additions, as it was believed that the instructions were contributing to SF failure. The instruction in the anthropogenic additions (AA) assessment were considerably shortened, along with changes to the terminology as described at the beginning of Chapter 4. The written instructions for each assessment task are shown in Table 4.8.

Table 4.8: Written instructions for each assessment task. Modifications were made based on observations of written responses and changes to the terminology used in the activity.

Written instructions
Emissions and Removals (ER)
Q1: In the year 2000, the atmospheric carbon concentration was approximately 370 parts per million (ppm). What needs to happen to the Anthropogenic Emissions to ensure that the atmospheric GHG concentration stabilises at—and does not exceed—400ppm by the year 2100? *[written description]*
Q2: Sketch what you think the trajectory for future Anthropogenic Emissions needs to be for the situation in Q1 to occur. *[graphical response]*
Anthropogenic Emissions (AE)
Q1: In the year 2000, the atmospheric carbon concentration was approximately 370 parts per million (ppm). Consider a situation where the Natural Emissions and Natural Absorptions are equal. What needs to happen to the Anthropogenic Emissions to ensure that the atmospheric GHG concentration stabilises at—and does not exceed—400ppm by the year 2100? *[written description]*
Q2: Sketch what you think the trajectory for future Anthropogenic Emissions needs to be for the situation in Q1 to occur. *[graphical response]*
Anthropogenic Additions (AA)
Q: Describe in your own words what the Anthropogenic Additions pump should do to ensure that the goal level is not exceeded. *[written description]*
Q: Sketch what you think the future trajectory of the Anthropogenic Additions pump should be to not exceed the goal level. *[graphical response]*

4.3.2 Graphical task

The graphical task was modified over the three assessment sheets to be consistent with the terminology in the workshop instructions. The ER and AE assessment sheets used the same graph, which had numerical values labelled on the x- and y-axis. The AA assessment sheet used a graph that had relational values on the x- and y-axis. The term anthropogenic emissions was changed to anthropogenic additions to align with the terminology in the Workbook experiments. Both graphical tasks are shown in Figure 4.10.

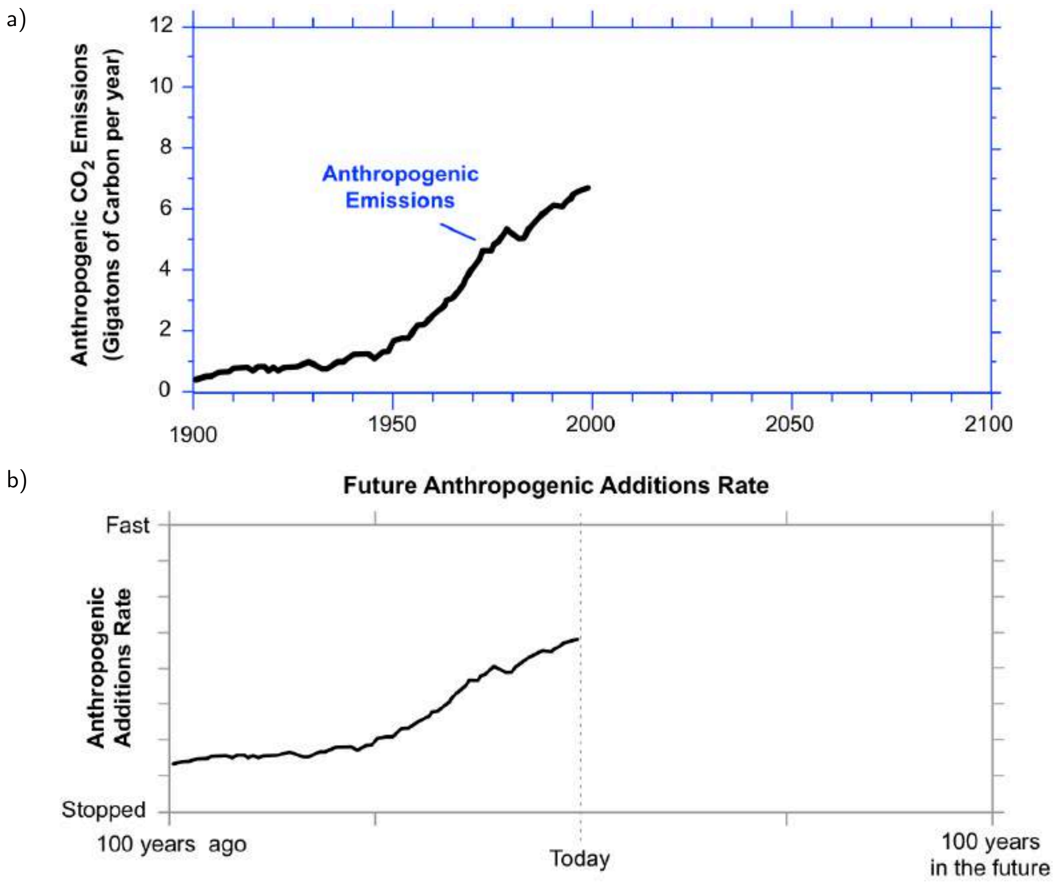
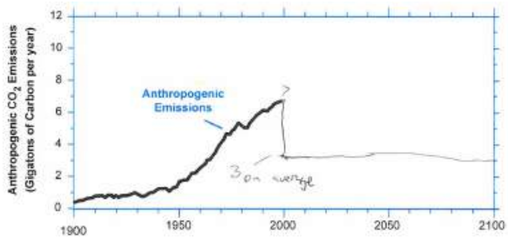


Figure 4.10: Graphical tasks used in the experiments. a) the graphical task used in the ER and AE assessment sheets; b) the graphical task used in the AA assessment sheet.

It should be noted that the graph in Figure 4.10b is not at the same scale as Figure 4.10a, and is an approximation of the representation of the anthropogenic emissions as described in Figure 2.4. The value at 100 years ago should have a y-value slightly closer to ‘stopped’. However, it was placed at just above the first minor tick to allow for the clear differentiation in responses between returning it to the same level and reducing the additions rate to zero.

The rationale for changing both the written and graphical tasks was to direct participants to thinking about the problem using the T&P system, rather than completing an algebraic activity. This type of thinking is shown in response #21034 in Box 4.1, where the participant has taken an algebraic approach to solving the problem, which then hasn’t translated correctly to the graph.

Box 4.1: A response demonstrating an algebraic solution.



Response #21034

30ppm over 100 years ∴ Anthropogenic emissions need to stabilise at 3ppm/10 years to reach 400ppm by year 2010.

A small relevant indication of results are presented below in relation to the terminology used in the graphical task as a justification for this change. The transcripts of written descriptions were analysed, specifically looking for the prevalence of words unique to the questions. The frequency of certain words in the written descriptions, shown in Table 4.9, indicates participants using different problem-solving methods in their thinking. ER and AE responses had much higher use of quantitative explanations in solutions, such as *parts per million* and *zero*, whereas AA responses had higher frequencies of qualitative explanations, such as *slow* and *stop*.

Table 4.9: Selected word frequencies in written responses by assessment sheet

	Frequency	
Word/s	ER and AE	AA
ppm/parts per million	149 (1.96%)	0 (0%)
zero or 0	31 (0.40%)	14 (0.32%)
slow	95 (1.25%)	169 (3.91%)
stop	95 (1.25%)	164 (3.79%)
speed	17 (0.22%)	32 (0.74%)
total number of responses	477 responses (7597 words)	316 responses (4325 words)

Note: the frequency count here may include numerous counts in a single response. Percentages are calculated from the total word counts in each category.

Due to the design of the experiment, it would not be conclusive to connect these results with performance, as many other variables changed. However, the frequency results in Table 4.9 suggest that the ER and AE participants were guided towards solving the graphical task predominantly in the conceptual target domain (the carbon cycle). AA participants were guided to thinking about the conceptual target domain in terms of the conceptual source domain (the T&P system).

4.3.3 Causal logic and group agreement

In the AA assessment sheet, a causal logic test and group agreement Likert scale were introduced into the assessment task. The causal logic test was used to further examine a mismatch between written and graphical responses that was observed in the Checklist experiment. Specifically, to understand whether the mismatch was due to incorrect causal logic or WG confusion. These are shown in Figure 4.11a.

a) **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

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

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

Figure 4.11: Additional tasks on the AA assessment sheet a) a series of questions to test the causal logic of the activity with correct answers indicated; b) The group agreement Likert scale

In the causal logic test, correct responses 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 correct responses had both circled, with an indication that the level would go up and then stay level. Typical examples of the future trajectories used in the causal logic tests are plotted in Figure 4.12; however, participants only had the written questions as shown in Figure 4.11.

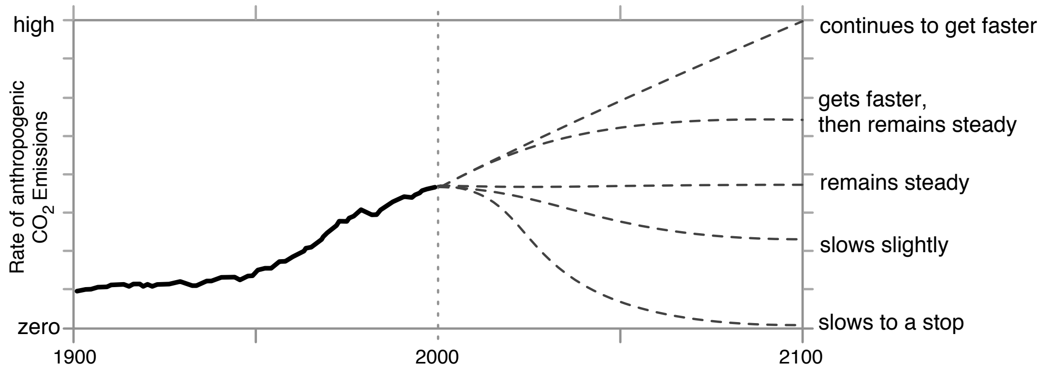


Figure 4.12: Future scenarios represented by the causal logic questions.

The results for the causal logic, and its relationship to the written and graphical responses are described in §5.3.4. For the reporting of results, the number of correct responses in the causal logic for each participant was summed out of 5.

The group agreement Likert scale was used to investigate individual perceptions of the group's response. This was included because of a question raised in the analysis of the ER and AE data, where I observed that graphical responses within the same group could vary wildly. The Likert scale is shown in Figure 4.11b.

Understanding whether participants agreed with each other could also be used as an instrument to better understand whether shared conceptual repertoires had been created. If shared conceptual repertoires had been created, agreement would be high and it would be expected that solutions would be very similar. If agreement was high, but answers were different, it could indicate some level of miscommunication: that the group thought they agreed with each other, but the conceptual repertoires were not shared. If agreement was low, it could indicate that the group was not functional.

4.3.4 Data interpretation and coding

In this section, I describe the coding schemata for the results. Data was collected using a paper-based profile questionnaire and assessment sheet. The assessment sheets are described in §3.4.1-§3.4.3, and are shown in Appendix B. The approach for collection of data had ANU human ethics approval (E2)[12]. No personally identifiable information was collected, and participants were not paid, graded or otherwise incentivised. Participation was encouraged but not required, and workshops were held during regular tutorials.

At the conclusion of the workshop, tests were placed in an envelope, which was labelled with the course code and the tutorial date and time. Once all workshop sessions for an experiment were completed, the responses were sorted and collated. A global identifier was stamped on the sorted tests using a numbering machine. Digital scans were made of the paper tests for data entry, and originals were filed and accessed as required.

The profile survey and coded interpretation of responses were entered into a digital form and spreadsheet. The coding rubrics are explained in this section. The responses then went through a blind, digital card sort, which allowed metadata to be extracted into a spreadsheet from the digital files. This card-sort spreadsheet was compared to the form spreadsheet using a differential script, and inconsistencies were addressed line-by-line.

Profile data & interpretation

The profile data were used to gauge the demographic profile of the participants in the activity. The profile questionnaire is shown in Appendix B, and was included in each assessment sheet. An overview of the questions, answers and reporting categories are shown in Table 4.10. The answer category shows the available answers, except for questions with asterisks, which show the typical responses to free-text questions. The reporting category describes the grouping used in the presentation of results.

12. Protocol 2013/050.

Table 4.10: Reporting categories for profile questions.

Question	Answer Category	Reporting Category
Age category	under 16; 16-17; 18-21	21 and under
	22-25; 26-29; 30 or over	22 and over
Sex	male	male
	female	female
Fields of study*	engineering; science; computer science; mathematics	STEM
	humanities; law; other	other
Degree progress	first year; second year	early year
	later years, postgraduate, other	later year
Language*	English	English
	Chinese; Bahasa; Korean; Portuguese; other	other

*Asterisks indicate questions which had free-text responses, and the answer category shows typical responses. Note: All questions have an NA reporting category for nil responses.

The profile survey and written descriptions were entered as indicated on the survey. In instances where there was no response, the 'NA' marker was used. In instances where there were multiple responses, a judgement call was made. The coding process and ambiguities are discussed briefly for each profile question below.

1. Indicate your age category (in years): participants chose from six age categories that approximately aligned with their study/degree progress; *under 16*, *16-17*, *18-21*, *22-25*, *26-29*, *30 or over*. Once the results were returned, these were later categorised as *under 22* and *22 and over*. One answer for this question was recorded, and interpretation of this question was straightforward.

2. Gender: participants chose from *male* or *female*. This is not, however, an accurate usage of language, as the corresponding answers for gender would be *masculine* and *feminine*. However, this was not considered a potential problem for misunderstanding, as the term 'gender' and 'sex' are commonly interchanged. One response in the later workshops pointed out that the framing of this question is not appropriate, as shown in Figure 4.13.

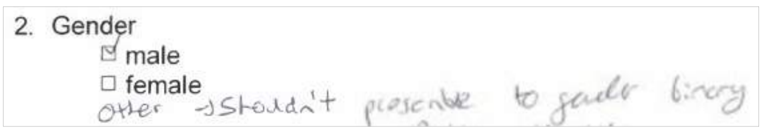
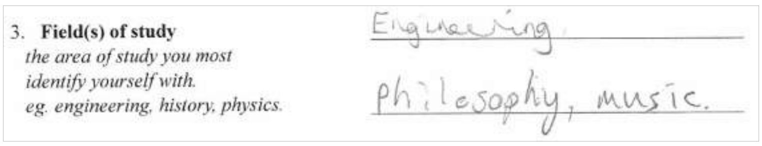

2. Gender
☑ male
☐ female
other → shouldn't prescribe to gender binary

Figure 4.13: Survey response showing gender question.

On reflection, the presentation of this question as a gender binary may have alienated some participants. The Australian Bureau of Statistics (ABS, 2014) is itself conducting a review of its sex standard to consider inclusive indicators for groups that do not fit into the gender binary, and this review is being supported by the Queer community and human rights groups.

3. Field(s) of study: participants had two free-text lines to complete their response. The intention of this question was to have a basic understanding of the types of prerequisite knowledge a student may bring to the activity. However, it would not be possible to draw conclusions based on this response with any certainty. As such, the question was used as a general identity marker. In terms of the universality of the learning activity, it would be desirable to see that this question was not an indicator of performance. For example, that humanities students performed as well as engineering students.

After all the results were entered, these free-text answers to this question were categorised into the following categories: *Engineering* (all specialisations)*;* *Computer Science* (information technology and software engineering)*; Science* (natural and physical sciences); *Other* (humanities, economics, law). The first response only was considered in the analysis, as it was considered to be the participant's dominant field of study. An example response is shown in Figure 4.14 where the respondent included three fields of study—*engineering*, *philosophy* and *music*—with *engineering* being recorded.


3. **Field(s) of study**
the area of study you most identify yourself with.
eg. engineering, history, physics.

Engineering

philosophy, music.

Figure 4.14: Survey response showing field(s) of study with multiple answers. Categorised as 'engineering'. Response #20222

The engineering, computer science and science sub-categories have been combined in the results as STEM for convenience in drawing comparisons between results.

4. Degree progress: participants could choose from five options: *first year, second year, later years, postgraduate, other.* Interpretation of this question was relatively straightforward with one exception. Many later-year students indicated that they were *Other*, and described their circumstance; for example, *third year* or *exchange student.* In these instances, these were recorded as *later years.* The responses to this question were categorised into *early years* (first and second) and *later years* (later and postgraduate).

5. Language you are most comfortable communicating in: participants could choose *English* or *other (please specify).* The intention of this question was to understand whether participants from different linguistic backgrounds performed differently. Only one language was recorded.

Many bilingual students indicated both *English* and their other language. This required interpretation. The other language was recorded in these instances. An example of this scenario is shown in Figure 4.15.

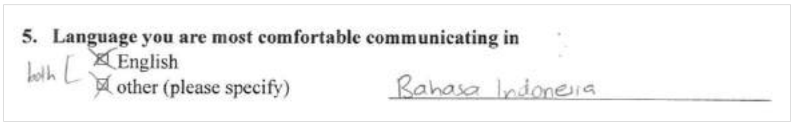
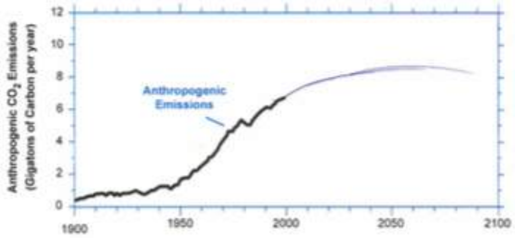

5. Language you are most comfortable communicating in
both
☒ English
☒ other (please specify) Bahasa Indonesia

Figure 4.15: Survey response showing preferred language with multiple answers Recorded as Bahasa Indonesia, and categorised as Other. Response #10424

6. Role in your group: used in workshops that used the physical T&P system. Participants chose their role in the activity from a list including: *Natural additions (emissions); Natural removals (absorptions); Anthropogenic additions (emissions); Observer/Instructor*; *Other (please describe).* The intention of this question was to investigate whether the role in the activity made a difference to their understanding of the dynamical behaviour of the system. In the small number of instances where no response was given, other members of the group were examined. If it was obvious that a role was vacant (for example, three roles had been filled, but one had not), the non-response was mapped to that role. If it was ambiguous, the role was mapped to '*Other*'. In the analysis, the *Observer/ Instructor* category was merged with the *Other* category.

7. Have you done this workshop before, or did you previously know the outcome?: this question was included to see if students had previously done this workshop: for example, if a student was both enrolled in environmental science and engineering. Participants could choose from: *No* or *Yes (please specify).* This question was not recorded in the results, as the number of *Yes* responses as a result of doing the activity before totalled 3, representing a minute portion of results. However, the second part of the question—the 'please specify'—did attract some interesting commentary. An example of this is shown in Box 4.2.

Box 4.2: Survey response showing a positive response to knowing the outcome with an incorrect graph.



Have you done this workshop before, or did you previously know the outcome?

Response #20921
Yes. I am capable of doing Math. It's a set of pumps, come on.

The commentary in Box 4.2 shows an interesting perception that the question is in itself trivial; however, the graphical response shows a clear misunderstanding of the question.

Coding of graphical responses

The graphical responses varied significantly. Initially, a descriptive analysis was undertaken, similar to the categorisation presented in Sterman & Booth Sweeney (2002). However, many inconsistencies began to emerge within the schema in relation to the learning activity. As the circumstances of the studies do not match exactly, it was not considered useful to match their coding schema, so a simpler schema was adopted.

The important distinction with the rubric used is this study, shown in Figure 4.16, is that it considers the trajectory of the graph and also the final magnitude. Figure 4.16 also displays the typical response in each category with a broken line, and a region of acceptable area for each category. Where responses ranged across categories, the dominant category was recorded. If the graphical response was unclear, or ranged across many categories, it was recorded as *Other*.

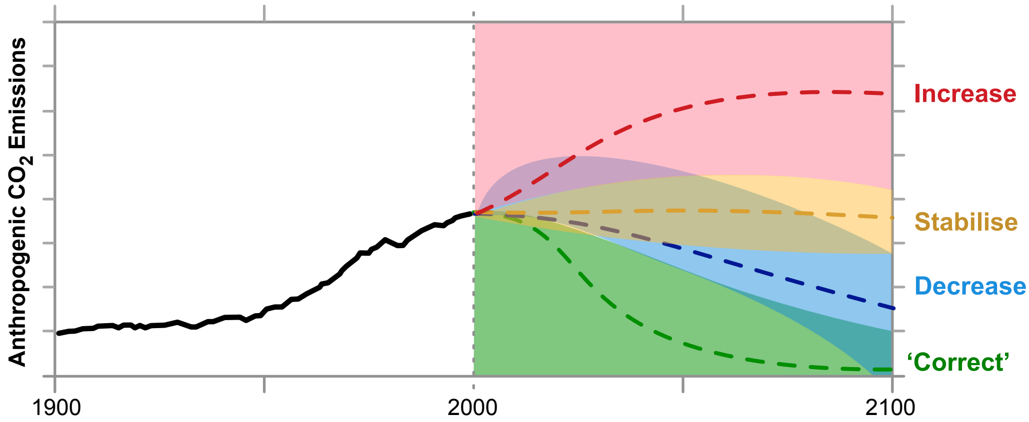
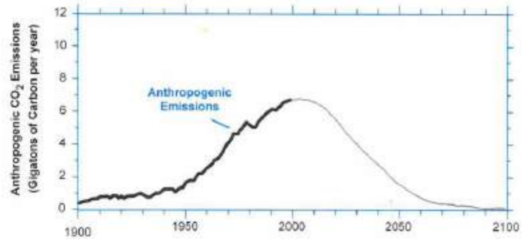


Figure 4.16: Rubric used to categorise graphical responses, with dotted lines showing typical responses. An 'other' category was used where responses did not fit clearly into the rubric. Y-axis values omitted.

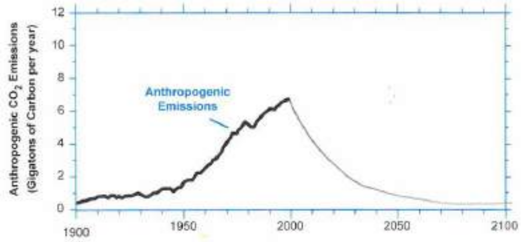
Examples of typical responses for all categories are shown in Box 4.3. The *Decrease* responses show approximately the boundary cases for final magnitude: #20153 is on the boundary of *Correct*, and #20525 is on the boundary of stabilise. The *Other* responses show examples that are clearly difficult to code, but are indeed incorrect.

The typical responses shown in Box 4.3 demonstrate the types of graphs shown in the majority of responses. However, there were a number of responses that were non-typical, and are worth noting here. This is shown in for the purposes of explicitly demonstrating the inconsistencies that arose in coding using the graphical attribute description for the zero emissions task in Sterman & Booth Sweeney (2002, p. 220). This is not at all a criticism of the Sterman & Booth Sweeney approach, but rather a justification for using a different coding schema because of the experiment that I was undertaking. In Box 4.4, the attributes described in Sterman & Booth Sweeney are provided with examples that match their attributes but don't match my coding schema.

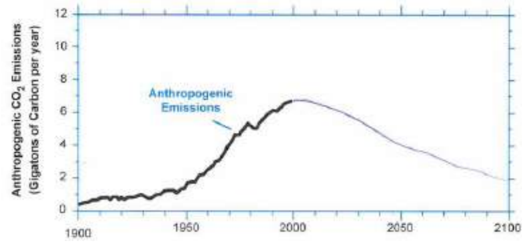
Box 4.3: Typical graphical responses for each of the five graphical response categories.



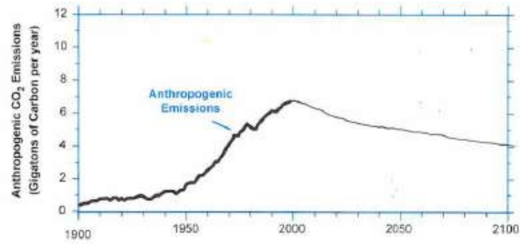
Correct Response #20222



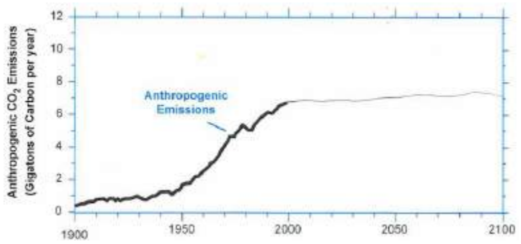
Correct Response #20413



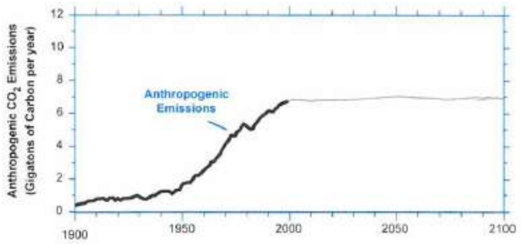
Decrease Response #20153



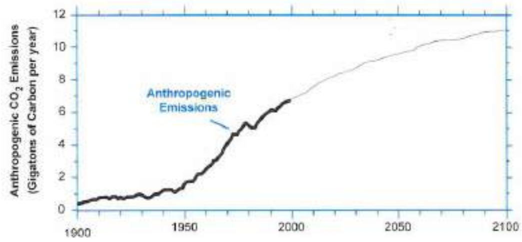
Decrease Response #20525



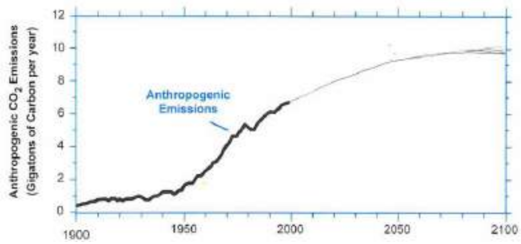
Stabilise Response #20131



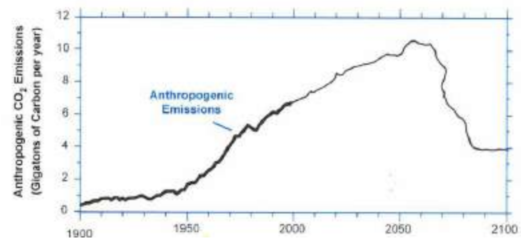
Stabilise Response #20422



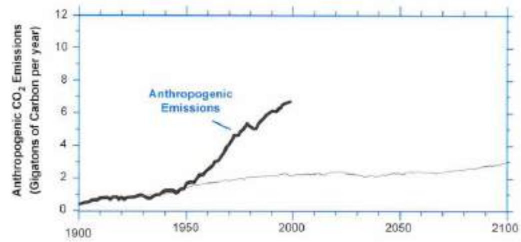
Increase Response #20143



Increase Response #20211



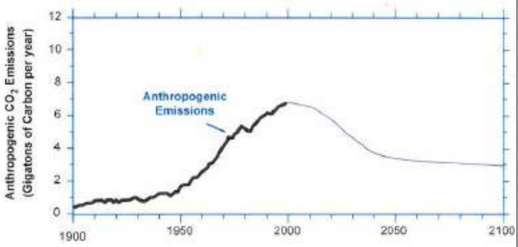
Other Response #11021



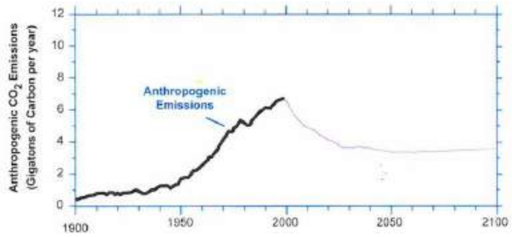
Other Response #21051

Box 4.4: Comparison of graphical responses categories with respect to Sterman & Booth Sweeney's coding schema

Correct: CO2 peaks at or very shortly after the year 2000, then declines at a diminishing rate:

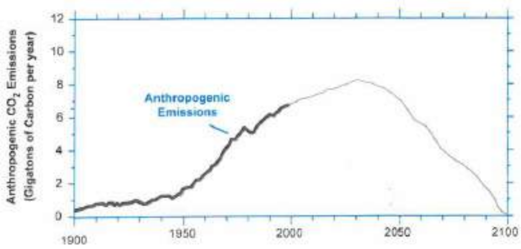


Decrease Response #10343

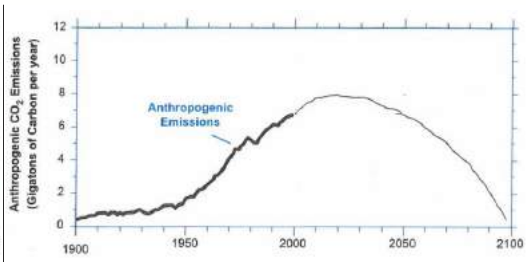


Decrease Response #10613

Incorrect: CO2 increases, then decreases:

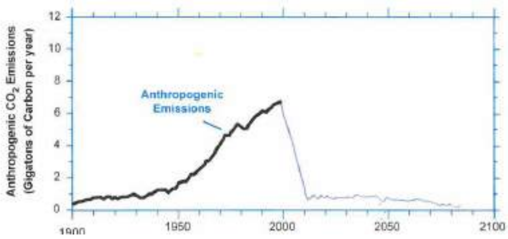


Decrease Response #10523

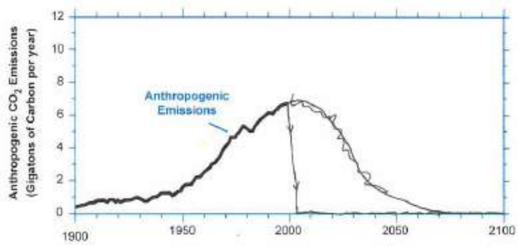


Decrease Response #21531

Incorrect: CO_2 immediately drops and continues to go down
(shows a sudden, discontinuous jump down at or very shortly after 2000):



Correct Response #20444



Correct Response #21111

Using the examples in Box 4.4, it is clear to see that the responses show a degree of correctness, especially in light of the T&P activity. The rationale for the graphical response coding in Box 4.4 is described below.

CO_2 peaks at or very shortly after the year 2000, then declines at a diminishing rate: responses #10343 and #10613 show a trajectory that matches the 'correct' attributes of Sterman & Booth Sweeney; however, the final magnitude would mean that the atmospheric stock in the T&P system would never stop increasing. These have been categorised in this study as *Decrease*.

CO2 increases, then decreases: responses #10523 and #21531 show a partially correct understanding—that the anthropogenic emissions must reach zero—but demonstrate a trajectory that would see the problem get much worse before it gets better. These responses have been categorised in this study as *Decrease*.

CO_2 immediately drops and continues to go down (shows a sudden, discontinuous jump down at or very shortly after 2000): responses #20444 and #21111 show a trajectory that almost immediately drops to zero. Although these graphs violate the feedback mechanisms in the real world and are unimaginable today, the trajectory would achieve a result where the atmospheric carbon goal is not exceeded. Using the logic in the T&P model, it could be argued these participants show a better understanding of the problem than the typical correct answer. These have been categorised as *Correct*.

This schema using in this study and shown in Figure 4.16 has been mapped to the schema used in Sterman & Booth Sweeney based on my interpretation of their results in Table 4.11.

Table 4.11: Perceived mapping of coding schemata between Sterman & Booth Sweeney (2002; 2007) and this study

Sterman & Booth Sweeney	Equivalent category
Correct	
CO_2 peaks at or very shortly after the year 2000, then declines at a diminishing rate.	*Correct* or *decrease*, depending on final magnitude
Incorrect	
CO_2 stabilizes in or after 2000 and never drops	*Increase* or *Stabilise*, depending on final magnitude
CO_2 keeps rising forever	*Increase*
CO_2 immediately drops and continues to go down (shows a sudden, discontinuous jump down at or very shortly after 2000)	*Correct* or *decrease*, depending on final magnitude
CO_2 stabilizes, then decreases	*Stabilise* or *decrease*, depending on final magnitude
CO_2 increases, then decreases	*Increase* or *stabilise*, depending on final magnitude
The CO_2 trajectory is discontinuous (has a sudden jump up or down at some other time than at or very shortly after 2000)	Any category, depending on final magnitude
The CO_2 trajectory follows some other path	Other (likely)

Coding of written descriptions

The written descriptions were both transcribed and coded. Transcription of written descriptions was undertaken by a professional transcription service. Only the relevant written question was provided to the transcription service, alongside the global identifier. Written descriptions were entered into a spreadsheet. The transcription service highlighted any unclear text[13] with 'unreadable' and any non-trivial formulae with 'formula'. Once the transcription service had completed this, I checked 5% of responses for accuracy, and found no significant issues. I examined the returned data and completed any indicated gaps.

For the text analysis, I corrected spelling errors[14] and replaced the following terms for consistency between workshops:

natural emissions/pump => natural additions/pump
natural absorptions/pump => natural removals/pump
anthropogenic emissions/pump => anthropogenic additions/pump
anthropogenic absorption/pump => anthropogenic removals/pump

The written descriptions were also coded. Sterman & Booth Sweeney present a coding system for their CO_2 and Temperature tests that extract concepts from the responses, including: mass balance, energy balance, pattern matching, inertia/delays, CO_2 fertilisation, sink saturation, and technology. Whilst this appears to be a useful coding system, it was immediately apparent that this was not the most useful way to interpret my results, and differentiating from their approach resulted in being a major influence on subsequent workshop designs.

In trying to code the written responses, I found that many participants drew a graph that was not adequately explained by their written description. For example, it was common to see a graph that showed an increasing future anthropogenic additions trajectory together with a written description that stated the anthropogenic additions needed to decrease to zero. This is discussed further in §6.2.

---

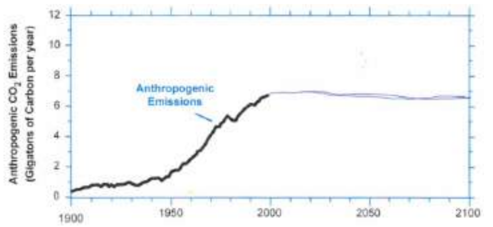
13. Text was either unclear due to handwriting illegibility or due to compression of the digital file. These were cleared up in all instances using the original paper responses

14. Spelling errors were typically transcribed as they appeared in the handwritten description

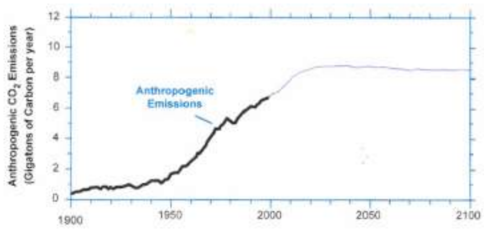
Accordingly, the written descriptions were coded into three categories by the degree to which they agreed with the graphical response: *Match-Strong, Match-Weak,* and *Mismatch.* Responses in the *Match-Strong* category described the behaviour in the graph correctly. Examples are shown in Box 4.5.

Box 4.5: Coding of graphical and written descriptions.

Match-Strong descriptions

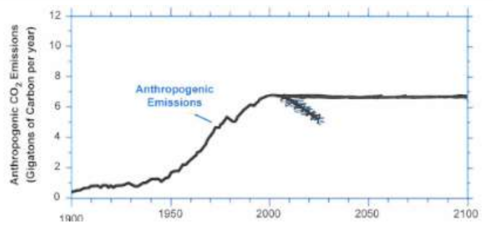


Response #20433- stabilise graph
Anthropogenic emissions need not to be increased at a fast rate over the period of time, it needs to be observed and control to ensure that the atmospheric GHG concentration stabilises & does not exceed.

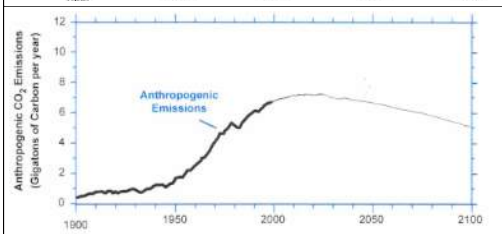


Response #20531- increase graph
That is if the anthropogenic emissions will stop increasing, and remain a constant rate, as it is not very likely to reduce the emissions.

Match-Weak descriptions

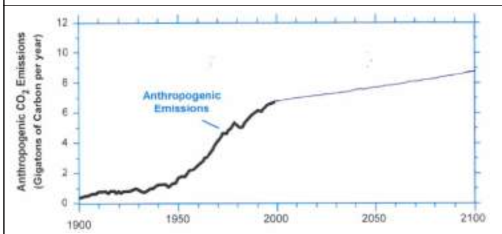


Response #21031 - stabilise graph
Reduce emissions to 1990 levels and maintain it there.

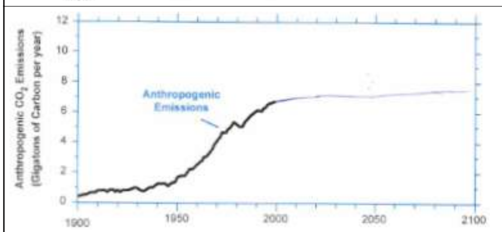


Response #21011 - decrease graph
If natural absorption and emissions are in dynamic equilibrium, anthropogenic emissions need to be cut in such a way that absorption whether through natural means or otherwise can compensate for those emissions.

Mismatch description



Response #20832 - increase graph
The anthropogenic emission needs to be kept at a minimum rate with no more increases.



Response #20121 - stabilise graph
Anthropogenic emissions must slow to a stop or a intervention which results in some anthropogenic absorption to balance out emissions must be performed.

Responses in the *Match-Weak* category described the behaviour in the graph partially, but not incorrectly; these descriptions accompanied a graph that typically had a trajectory in the same direction. Responses in the *Mismatch* category, however, had a trajectory which followed the opposite direction than that described. Hence, 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*.

For the discussion of the results, the *Match-Strong* and *Match-Weak* categories were combined into a *Match* category, as the interpretation between these categories was often difficult to clearly distinguish; whereas the mismatched descriptions clearly contradicted the behaviour in the graph. The *Increase and Stabilise* graphical response categories were also combined to *Incorrect*.

The combined coding of graphical and written responses means that there are six possible combinations in the analysis of these results, which can be grouped into three categories, shown in Table 4.12: Sound accumulation principles (AP), SF failure, and WG confusion.

The latter categories highlight an important distinction between intentional and unintentional SF failure: that an incorrect graph with a matching description (*Incorrect-Match*) represents a response from someone with a flawed mental model about the problem, and; any graph with a mismatching description (*Correct-Mismatch, Decrease-Mismatch, or Incorrect-Mismatch*) demonstrates Written-Graphical confusion, as the written response contradicts their graphical response.

Table 4.12: Graphical-Written response categories with level of graphical-written understanding: *Sound accumulation principles* (AP), *SF failure*, and *WG confusion*.

	Graphical Response		
Written Response	**Correct**	**Decrease**	**Incorrect**
Match	Sound AP		SF failure
Mismatch		WG confusion	

The goal of the T&P activity is to investigate the difference between SF failure and WG confusion, and ultimately help participants to move from these categories to the Sound AP category.

4.4 Treatment conditions

In this section, I present an overview of the treatment conditions in the three experiments. Three experiments were used to assess the research questions outlined at the start of Part II. However, a total of nine workshops were designed, with workshops used to trial ideas before running the larger experiments. A discussion of these workshops is in Appendix D. Treatment conditions were established for the three remaining experiments, with key differences outlined in Table 4.1.

Workshop participants were enrolled in engineering, computer science, or environmental science courses at The Australian National University in 2013-2014. Trial workshops were used to test activity and assessment approaches. Workshops were part of small-group sessions in pre-existing courses that could use the T&P activity as a relevant learning exercise. Degree progress was thus a factor in the experimental design, as many students who participated in their first year in 2013[15] would be in their second year in 2014, and so replicating the activity with these students could lead to biased results. Table 4.13 summarises the participating courses in the T&P activity.

Table 4.13: Summary of workshops and participants in undergraduate courses

Experiment / Semester			Participants / Class (College)	
Checklist	PC-ER	2013	133	ENVS1001 Environment and Society (Science)
			35	COMP3530 Sys. Eng. for Software Engineers (Comp Sci)
Checklist	PC-AE	2013	154	ENGN2225 Systems Engineering Design (Engineering)
			162	ENGN1211 Discovering Engineering (Engineering)
Simulation	PW-AA MW-AA	2014	175	ENGN1211 Discovering Engineering (Engineering)
Dialogue	PW-GD PW-SC	2014	96	ENVS1001 Environment and Society (Science)
			35	VCUG3001 Unravelling Complexity (All colleges)

Note: I was the convener of ENGN2225 in 2013, and VCUG3001 in 2014. Precautions were made through the Human Ethics protocol to minimise influence on these results.

15. Specifically, those enrolled in ENGN1211 in 2013 are likely to enrol in ENGN2225 in 2014

As each T&P activity was run in a real tutorial session in an undergraduate class, the number of responses recorded for each treatment condition depended on the number and size of tutorials. Groups were made up of approximately equal numbers of participants within tutorials. An overview of the number of tutorials, groups and participants are shown by workshop and treatment condition in Table 4.14.

Table 4.14: Summary of group sizes for treatment conditions within workshop sessions

Experiment/treatment condition		Reference	No. of Tutorials	No. of Groups	No. of Participants
Checklist	ER assessment	PC-ER	10	45	168
	AE assessment	PC-AE	16	78	317
Simulation	Physical simulation	PW-AA	9	18	93
	Mental simulation	MW-AA	9	23	82
Dialogue	Physical simulation	PW-AA	8	8	26
	Group Diagram	PW-GD	9	9	33
	Scenario Cards	PW-SC	10	10	35
	Group Diagram & Scenario Cards	PW-GDSC	11	13	51

Note: where treatment references are repeated, the treatment activities are the same (PW-AA)

The differences between the experiments are discussed throughout this chapter. In this section, the differences between the treatment conditions within these experiments are discussed.

4.4.1 Checklist experiment

The Checklist experiment had two versions of the assessment task: the *Emissions & Removals* (ER) and *Anthropogenic Emissions* (AE), described in §4.2. This change was made after observations of confusion in the ER assessment task.

The two assessment tasks in the Checklist experiments are also compared to the base (PW-AA) condition of the Simulation and Dialogue tasks in the results.

4.4.2 Simulation experiment

The Simulation experiment was a between-group experiment comparing physical and mental simulation. In the mental simulation, groups were required to label the parts of a pictorial representation of the analogue, rather than the parts themselves, and then mentally simulate the behaviour over time of the system. The activity sequence between treatments were otherwise identical.

The picture that the group labelled in the workbook task is shown in Figure 4.17, with correct labels shown in the boxes. Once a group had correctly labelled the diagram, they were given an A4-sized copy of the diagram that was correctly labelled to use as a reference.

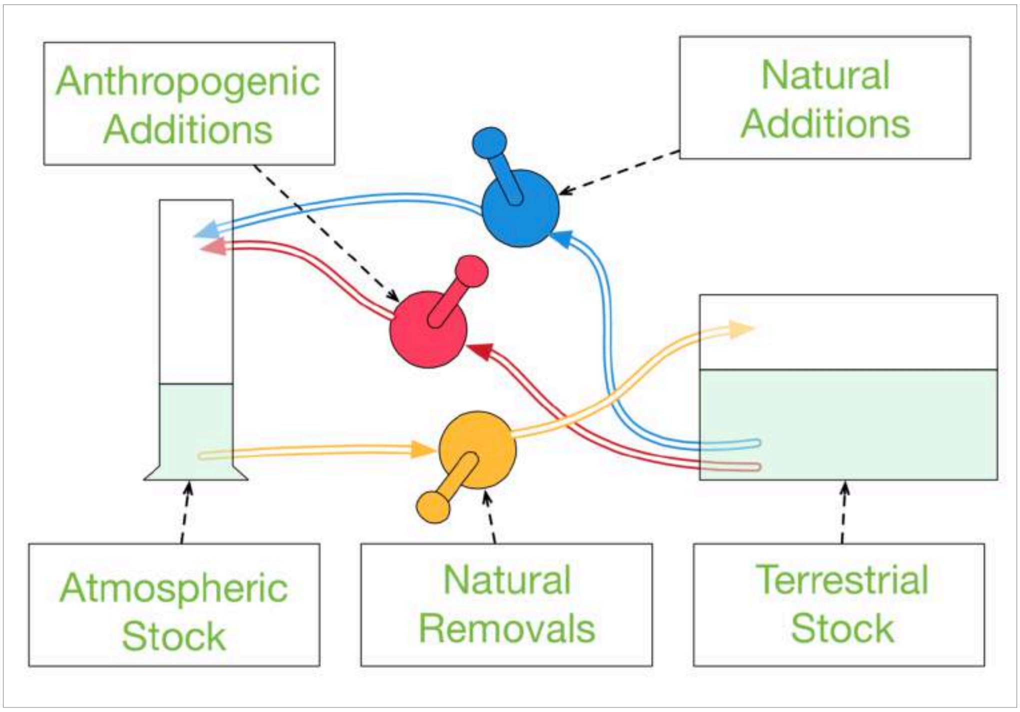


Figure 4.17: Pictorial representation of the T&P analogue used in the MW-AA treatment condition with correct answers filled in. Note that the Additions pumps are interchangeable.

The pictorial representation meant that the experiential nature of the analogue was replaced with a cognitive process rather than a physical process. Participants were prompted to mentally simulate the causal logic of the pumps rather than being able to physically simulate the system.

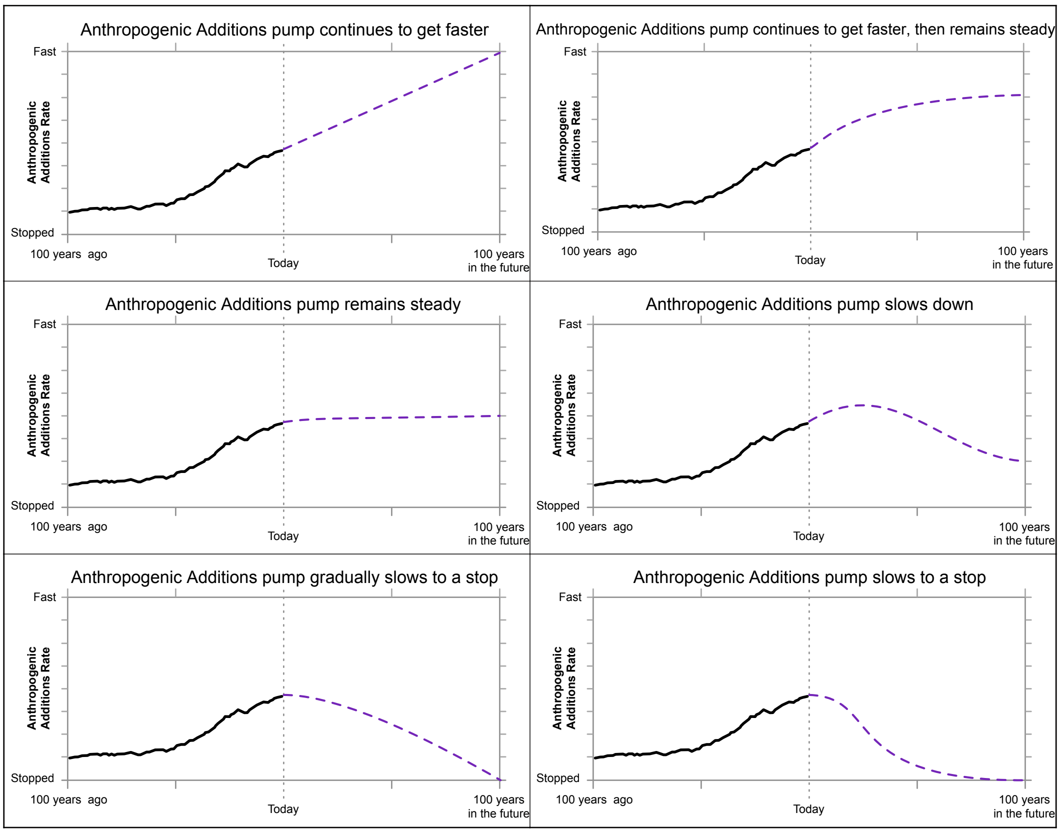
4.4.3 Dialogue experiment

In the Dialogue experiment, a 2×2 between-group factorial experiment was setup to investigate whether targeted prompts could help reduce SF failure, demonstrated by a mismatch between graphical and written responses. All groups used the same assessment task, but had small variations to the instructions within the activity. In this section, the scenario cards (PW-SC, PM-GDSC), group diagram (PW-GD, PW-GDSC) will be described with respect to the base (PW-AA) activity.

Scenario card (SC) treatment conditions

In the SC treatment, Task 4 (see Table 4.6) was modified, asking the group to use scenario cards as a prompt to discuss possible future scenarios for the anthropogenic additions pump. These were six A6-sized 'postcards', that presented future trajectories of anthropogenic additions on the same graph used in the assessment sheet. The cards also included the corresponding written description. The graph and descriptions on the scenario cards are shown in Box 4.6.

Box 4.6: Information presented on the scenario cards, with scenarios ranging from the pump getting faster through to slowing to a stop



The intention behind using the scenario cards was to investigate whether this prompt would reduce the frequency of mismatching written and graphical responses by linking graphs with the correct description.

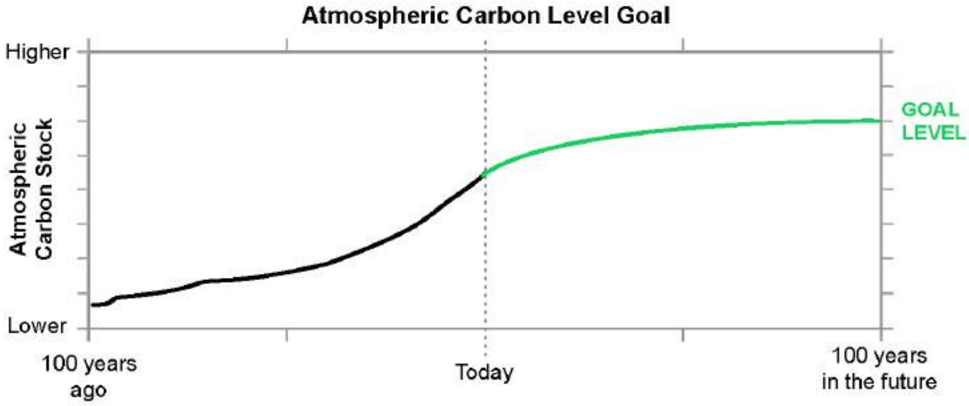
Group diagram (GD) treatment conditions

In the final workbook task, the GD treatment condition required groups to draw what the anthropogenic additions pump was required to do in order to reach the atmospheric carbon goal level, considering that the pumps in the natural cycle should remain approximately equal. In the GD treatment, groups were also required to draw the future trajectory of the anthropogenic additions rate in the final task.

The workbook modification for this task is shown in Figure 4.18. This task replicates the individual assessment task, and presents an opportunity for the group to make answers explicit before moving to the individual task.

Task 5: What now?

In Task 4, your group was asked to stop the atmospheric stock from overflowing. This level was your 'goal' level.



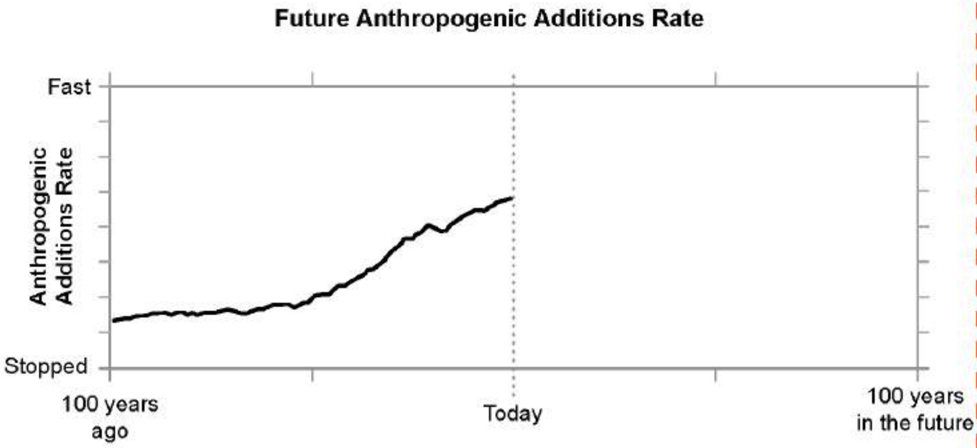
The atmospheric carbon stock should not exceed the goal level.

Describe what needs to happen to the **Anthropogenic Additions** pump to ensure that the goal level is not exceeded (assume that the **Natural Additions** and **Natural Removals** remain approximately the same):

____

____

Sketch what you think the future Anthropogenic Additions pump rate should be to reach the goal level:



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

Figure 4.18: Workbook modification for the GD treatment. The additional task is shown enclosed by a dotted line. This section was blank for groups that were not part of this treatment condition.

5 Workshop results

This chapter is divided into three sections: a summary of all results, results by experiment, and results by task. These results are then discussed further in the Chapter 6 synthesis.

5.1 Summary results

In this section, summary results are presented. This provides a general overview of the results before a more detailed breakdown in the later sections. Summary results are shown by treatment condition, course, assessment sheet, and profile data.

Note that throughout this chapter the graphical-written combinations *S*ound AP, SF failure, and WG confusion described in Table 4.12 are used primarily in the reporting of results. Where p values and odds-ratios (OR) are reported, p is calculated using Fisher's exact test with significance at α=0.05. OR is reported at a 95% confidence interval.

5.1.1 Summary results by treatment condition

Results are shown by treatment condition for each level of graphical-written understanding in Table 5.1. The PC-ER treatment condition produced the worst results at 39% of participants demonstrating *Sound AP*; whereas the PW-GDSC treatment condition produced the best results with 95% of participants demonstrating a *Sound AP*.

Table 5.1: Summary results for each treatment condition by understanding level (frequency and row percentage).

Experiment/Treatment		n	Sound AP		WG confusion		SF Failure	
Checklist	PC-ER	168	65	(39%)	89	(53%)	14	(8%)
	PC-AE	317	162	(51%)	126	(40%)	28	(9%)
Simulation	PW-AA	82	54	(66%)	25	(30%)	3	(4%)
	MW-AA	93	61	(66%)	30	(32%)	2	(2%)
Dialogue	PW-AA	26	18	(69%)	7	(27%)	1	(4%)
	PW-GD	33	24	(73%)	6	(18%)	3	(9%)
	PW-SC	35	28	(80%)	6	(17%)	1	(3%)
	PW-GDSC	51	35	(95%)	2	(5%)	0	(0%)

5.1.2 Summary results by course

The summary results presented in Table 5.2 show two subsequent years using the T&P activity in two courses: first-year engineering (ENGN1211), and first-year environmental science (ENVS1001). In both courses, participants in the Checklist experiment performed worse than the lowest performing treatment groups in the Workbook experiments.

Table 5.2: Summary results for courses by understanding level in each course (frequency and row percentage).

Course/Year/Treatment		n	Sound AP		WG confusion		SF Failure	
ENGN1211	2013 PC-AE	162	86	(53%)	62	(38%)	14	(9%)
	2014 PW-AA	82	54	(66%)	24	(29%)	4	(5%)
	2014 MW-AA	93	61	(66%)	30	(32%)	2	(2%)
ENVS1001	2013 PC-ER	133	53	(40%)	73	(55%)	7	(5%)
	2014 PW-AA	18	11	(61%)	6	(33%)	1	(6%)
	2014 PW-GD	24	16	(67%)	5	(21%)	3	(13%)
	2014 PW-SC	25	18	(72%)	6	(24%)	1	(4%)
	2014 PW-GDSC	29	27	(93%)	2	(7%)	0	(0%)

Each cohort in each course were exposed to a different combination of treatment conditions, but had the PW-AA treatment condition in common. There was no significant difference between the performance of students in ENGN1211 and ENVS1001 in the PW-AA treatment condition (*Fisher exact test:* $p = 0.79$, OR = 1.21 at 95% CI [0.36,3.82]).

However, there is one relevant difference between the first-year students in environmental science and the first-year students in engineering: although students studying in these programs are both typically categorised as STEM students, the Bachelor of Engineering requires a year 12 mathematics prerequisite (ANU 2013 p.130), whereas the degree programs that students in the environmental science course are enrolled in do not have this prerequisite, except where they are studying specific topics[16]. This could lead to an assumption that the engineering students would perform better on the tests, regardless of the treatment condition, which does not appear to be the case.

16. such as maths, physics or chemistry

5.1.3 Summary results by assessment sheet

The summary results presented in Table 5.3 show the aggregate results from the three different assessment sheets: Emissions & Removals (ER), Anthropogenic Emissions (AE) and Anthropogenic Additions (AA)[17]. For the AA assessment, only the PW-AA treatment condition is shown in Table 5.3 as it represents the treatment condition with the most similar tasks to the other sheets.

Table 5.3: Summary results for assessment sheet by understanding level (frequency and row percentage).

Assessment/Treatment		n	Sound AP		WG confusion		SF Failure	
ER	PC-ER	168	65	(39%)	89	(53%)	14	(8%)
AE	PC-AE	316	162	(51%)	125	(40%)	29	(9%)
AA	PW-AA	108	72	(67%)	31	(29%)	5	(5%)

Participants using the ER assessment sheet performed worse (Sound AP: 39%, n=65) than participants using the AE assessment sheet (Sound AP: 51%, n=162). This is a notable result, as the only difference between the two groups was the assessment sheet. Participants using the AA assessment sheet in the PW-AA treatment condition performed better again (Sound AP: 67%, n=72). However, both the assessment sheet and the instructional delivery (from a checklist to a workbook) were modified for this group, so it is not clear which factor influenced the result.

5.1.4 Summary results by profile data

Data from the profile survey show the age category, sex, degree progress, field of study and language most comfortable communicating in. A summary of the participant profiles is shown in Table 5.4 for the Checklist and Workbook experiments. Numbers indicate the number of participants, and percentage columns indicate the percentage within profile categories.

17. see §4.3 for the description of differences

Table 5.4: Summary of profile data across the workshops (category percentages)

Profile category		Checklist		Workbook		
Age category	≤21	406	(83.9%)	268	(83.8%)	p = 0.92
	≥22	75	(15.5%)	51	(15.9%)	
	NA	3	(0.6%)	1	(0.3%)	-
Sex	female	143	(29.5%)	119	(37.2%)	p = 0.03 OR: 0.72 95% CI[0.53,0.98]
	male	333	(68.8%)	199	(62.2%)	
	NA	8	(1.7%)	2	(0.6%)	-
Field of study	STEM	436	(90.1%)	268	(83.8%)	***p < 0.01*** OR: 1.89 95% CI[1.20,3.00]
	other	43	(8.9%)	50	(15.6%)	
	NA	5	(1.0%)	2	(0.6%)	-
Degree progress	early	365	(75.4%)	264	(82.5%)	p = 0.03 OR: 0.66 95% CI[0.52,1.15]
	later	115	(23.8%)	55	(17.2%)	
	NA	4	(0.8%)	1	(0.3%)	-
Language	English	392	(81.0%)	270	(84.4%)	p = 0.21
	other	90	(18.6%)	48	(15.0%)	
	NA	2	(0.4%)	2	(0.6%)	-
	Σ	484		320		

Note: percentages may not total 100% due to rounding.

The Field of study profile category has both significance and an odds ratio that doesn't span the null value at a 95% confidence interval between the experiments. This indicates that there was a significantly higher proportion of STEM students in the Checklist experiment than the Workbook experiment. Profile bias between treatment condition may lead to issues when comparing different workshop designs. The *Sound AP* results were tested for each profile category, shown in Table 5.5. There was no significant difference between performance in these groups for the majority of profile categories.

Table 5.5: Summary results by profile category (frequency and category percentage).

Profile category		n	Sound AP		
Age category	≤21	656	383	(58%)	p = 0.14
	≥22	122	62	(51%)	
Sex	male	512	293	(57%)	p = 0.48
	female	248	149	(60%)	
Field of study	STEM	683	395	(58%)	p = 0.96
	other	86	49	(57%)	
Degree progress	early	610	347	(57%)	p = 0.59
	later	165	98	(59%)	
Language	English	646	383	(59%)	***p < 0.01*** OR: 1.66 95%CI[1.12,2.46]
	other	135	63	(47%)	

Note: NA results have not been included.

The only profile category that is an indicator of performance is the language category. Participants who speak English as their first language performed significantly better (Sound AP: 59%, n=383) than speakers of other languages (Sound AP: 47%, n=63). This could lead to bias in the results if the proportion of English-speaking students is different between treatment groups. However, as shown in Table 5.4, there was no significant difference in chosen language between the Checklist and Workbook experiments.

5.2 Results by experiment

In this section, I present results for the three experiments. In each experiment, I report the specific subset of workshops for each experiment, show the profile data for each treatment condition, and report the level of graphical-written understanding for each treatment condition.

5.2.1 Scaffold experiment

Experiment 1: Scaffolded refinement
Does display of instructions and information in the T&P activity reduce SF failure?

In the Scaffold experiment, the results for the three different assessment sheets will be considered, extending the summary in §5.1.3. The Checklist experiment used the same instructions, with two assessment sheets: Emissions & Removals (ER) and Anthropogenic Emissions (AE). The result from the Workbook experiment reported here was the base condition (PW-AA) for the Simulation and Dialogue experiments.

The treatment conditions are shown in Table 5.6, with instructions, assessment sheet, year, course, and number of participants.

Table 5.6: Scaffold experiment treatment conditions by course and assessment sheet.

Treatment	Instructions	Assessment	Year	Course	n
PC-ER	Checklist	ER	2013	COMP3530	35
				ENVS1001	133
PC-AE	Checklist	AE	2013	ENGN1211	162
				ENGN2225	154
PW-AA	Workbook	AA	2014	ENVS1001	18
				VCUG3001	8
				ENGN1211	82

The profile data is shown in Table 5.7 for each treatment condition. This shows that the samples were biased in the sex, field, degree progress and language categories, largely due to the different cohorts used in the experiment. As shown in Table 5.5, language is an indicator of performance, with English speakers 1.66 95%CI [1.12,2.46] more likely to demonstrate *Sound AP*. If there is no difference between the treatment conditions, we would expect PC-ER to have the best results, as it is the treatment group with the highest proportion of English speakers at 88%, above PW-AA at 81% and PC-AE at 78%.

Table 5.7: Scaffold experiment treatment conditions by profile data (frequency and cell percentage of population).

Profile category		PC-ER		PC-AE		PW-AA		
Age	≤21	138	(83%)	268	(85%)	90	(84%)	$p = 0.83$
	≥22	28	(17%)	47	(15%)	17	(16%)	
Sex	male	77	(47%)	256	(82%)	71	(66%)	$p < 0.01$
	female	87	(53%)	56	(18%)	36	(34%)	
Field of study	STEM	125	(76%)	311	(99%)	94	(88%)	$p < 0.01$
	other	40	(24%)	3	(1%)	13	(12%)	
Degree progress	early	124	(75%)	241	(77%)	94	(88%)	$p = 0.02$
	later	42	(25%)	73	(23%)	13	(12%)	
Language	English	146	(88%)	246	(78%)	87	(81%)	$p = 0.02$
	other	20	(12%)	70	(22%)	20	(19%)	
	Σ	168		316		108		

Note: percentages may not total 100% due to rounding, and counts may not equal total due to NA responses. p-values calculated using Fisher's exact test.

However, this is not the result when considering the treatment groups. In fact, the PC-ER treatment condition performs the worst in the graphical, written and combined tasks, with the PW-AA performing the best in all tasks. Approximately two-thirds of participants in the PW-AA treatment condition demonstrated Sound AP, whereas this was 51% in the PC-AE treatment condition and 39% in the PC-ER treatment condition, shown in Table 5.8.

Table 5.8: Scaffold experiment results by treatment category for graphical and written tasks (frequency and cell percentage of population).

Treatment	n	Graphical Correct		Graphical Task Correct/Decrease		Written Task (Match)		Sound AP	
PC-ER	168	22	(13%)	80	(48%)	79	(47%)	65	(39%)
PC-AE	316	114	(36%)	180	(57%)	190	(60%)	162	(51%)
PW-AA	108	44	(41%)	74	(69%)	76	(70%)	72	(67%)
Σ	*592*	*180*	-	*334*	-	*345*	-	*299*	-
Exact test (p)		***p < 0.01***		***p < 0.01***		***p < 0.01***		***p < 0.01***	

The treatment condition is a significant factor for both the written and graphical tasks ($p < 0.01$ on all tasks). The performance improvement between the PC-ER and PW-AA tasks are 28% in demonstrating a correct graph, 21% in demonstrating a correct or decreasing graph, 23% in writing a matching response, and 28% in achieving Sound AP.

These results are positive, and indicate that the display of instructions and information in the activity has a significant effect on the results. This is discussed further in Chapter 6.

5.2.2 Simulation experiment

Experiment 2: Physical simulation
Does representational form of the T&P activity reduce SF failure?

The simulation experiment investigated whether participants performed better using the T&P system or a picture of the T&P system. Participants using the picture of the T&P system were unable to simulate the behaviour of the system, and so had to mentally simulate how they thought the system would behave.

The treatment conditions are shown in Table 5.9, with instructions, simulation type, year, course, and number of participants. In this instance, the treatment conditions are contained within a single first-year engineering class.

Table 5.9: Simulation experiment treatment conditions by course and simulation type.

Treatment	Instructions	Simulation	Year	Course	n
PW-AA	Workbook	Physical	2014	ENGN1211	82
MW-AA	Workbook	Mental	2014	ENGN1211	93

There were no significant profile data differences between participants in the treatment conditions, shown in Table 5.10, likely because the experiment here is presented with results from a single course. Specifically, language was shown to be a significant factor for whether participants could reach the correct answer in Table 5.5, and does not appear to bias treatment conditions in this experiment in a significant way. Hence, the groups are considered to be random, and the participants' profiles are not considered to bias results.

Table 5.10: Simulation experiment treatment conditions by profile data (frequency and cell percentage of population).

Profile category		PW-AA (ENGN1211)		MW-AA (ENGN1211)		
Age	≤21	75	(93%)	87	(94%)	*p* = 1
	≥22	6	(7%)	6	(6%)	
Sex	male	62	(77%)	75	(82%)	*p* = 0.46
	female	19	(23%)	17	(18%)	
Field of study	STEM	81	(100%)	90	(97%)	*p* = 0.25
	other	0	(0%)	3	(3%)	
Degree progress	early	78	(96%)	90	(97%)	*p* = 1
	later	3	(4%)	3	(3%)	
Language	English	63	(78%)	80	(86%)	*p* = 0.17
	other	18	(22%)	13	(14%)	
	Σ	82		93		

Note: category percentages may not total 100% due to rounding, and counts may not equal total due to NA responses.

In this experiment, the physical simulation treatment condition (PW-AA), performed marginally better than the mental simulation treatment condition: 4% in the correct graphical task, 2% in the written task, and less than 1% in the Sound AP category. No results in this experiment are significant. Approximately two-thirds of participants were able to demonstrate Sound AP regardless of simulation method. These results are shown in Table 5.11.

Table 5.11: Simulation experiment results by treatment category for graphical and written tasks (frequency and cell percentage of population).

Treatment	n	Graphical				Written Task (Match)		Sound AP	
		Correct		Correct/Decrease					
PW-AA	82	33	(40%)	56	(68%)	57	(70%)	54	(66%)
MW-AA	93	32	(34%)	62	(67%)	63	(68%)	61	(66%)
Σ	*175*	*65*	-	*118*	-	*120*	-	*115*	-
Exact test (p)		p = 0.88		p = 0.87		p = 0.87		p = 1	

These results show that there is no significant difference in the results, regardless of whether participants physically manipulated the pumps. This has the implication that physical simulation did not improve participants' mental models. This finding is discussed further in Chapter 6.

5.2.3 Dialogue experiment

Experiment 3: Focussed dialogue
Do opportunities for focussed dialogue during the T&P activity reduce SF failure?

The Dialogue experiment investigated whether participants performed better when they were exposed to specific prompts during the activity. The prompts were designed to encourage focussed dialogue: scenario cards were used to aid exploration of ideas, and a group diagram was used to focus thinking about a solution. The experiment was set up with a 2×2 factorial design, which allowed exploration of the interaction between these two factors. The two treatment conditions were therefore in four treatment groups: without a group diagram or scenario cards (PW-AA); with either a group diagram (PW-GD) or scenario cards (PW-SC), and; with both group diagram and scenario cards (PW-GDSC).

The treatment groups are shown in Table 5.12, with instructions, dialogue prompts, year, course, and number of participants. The experiment was contained within a single first-year environmental science class.

Table 5.12: Dialogue experiment treatment conditions by course and dialogue prompts.

Treatment	Instructions	Dialogue Prompts	Year	Course	n
PW-AA	Workbook	None	2014	ENVS1001	18
PW-GD	Workbook	Group Diagram	2014	ENVS1001	24
PW-SC	Workbook	Scenario Card	2014	ENVS1001	25
PW-GDSC	Workbook	Group Diagram & Scenario Cards	2014	ENVS1001	29

There were no significant profile data differences between participants in the treatment conditions, shown in Table 5.13, likely because the experiment here is presented with results from a single course. However, the PW-AA does appear to have a more balanced age and sex ratio than other treatment conditions. Given that these are not significant indicators of performance, as shown in Table 5.5, these are not considered to influence results. Hence, the groups are considered to be random, and the participants' profiles are not considered to bias results.

Table 5.13: Dialogue experiment treatment conditions by profile data (frequency and cell percentage of population).

Profile category		PW-AA (ENVS1001)		PW-GD (ENVS1001)		PW-SC (ENVS1001)		PW-GDSC (ENVS1001)		
age	≤21	10	(56%)	20	(83%)	20	(80%)	23	(79%)	*p = 0.20*
	≥22	8	(44%)	4	(17%)	5	(20%)	6	(21%)	
sex	female	9	(50%)	9	(38%)	9	(36%)	14	(48%)	*p = 0.71*
	male	9	(50%)	15	(63%)	16	(64%)	15	(52%)	
field	STEM	12	(67%)	15	(63%)	19	(79%)	23	(79%)	*p = 0.45*
	other	6	(33%)	9	(38%)	5	(21%)	6	(21%)	
degree progress	early	16	(89%)	20	(83%)	20	(80%)	26	(90%)	*p = 0.78*
	later	2	(11%)	4	(17%)	5	(20%)	3	(10%)	
language	English	16	(89%)	18	(75%)	22	(88%)	24	(83%)	*p = 0.62*
	other	2	(11%)	6	(25%)	3	(12%)	5	(17%)	
	Σ	18		24		25		29		

Note: percentages may not total 100% due to rounding, and counts may not equal total due to NA responses. *p*-values calculated using Fisher's exact test.

The results for the Dialogue experiment are considered for the with- and without-GD and SC conditions, shown in Table 5.14. Using scenario cards to achieve a Correct/Decrease answer on the graphical task was shown to be significant ($p < 0.01$) OR = 4.37 at 95% CI: [1.4,15.4] indicator for performance.

Table 5.14: Dialogue experiment results by treatment category for graphical and written tasks (frequency and cell percentage of population).

Treatment	n	Graphical Task				Written Task Match		Sound AP	
		Correct		Correct/Decrease					
With GD	53	27	(51%)	43	(81%)	46	(87%)	43	(81%)
Without GD	43	17	(40%)	32	(74%)	31	(72%)	29	(67%)
Exact test (p)		*p = 0.307*		*p = 0.465*		*p = 0.121*		*p = 0.157*	
With SC	54	30	(56%)	48	(89%)	46	(85%)	45	(83%)
Without SC	42	14	(33%)	27	(64%)	31	(74%)	27	(64%)
Exact test (p)		*p = 0.039*		*p < 0.01*		*p = 0.201*		*p = 0.056*	
With GD and SC	29	22	(76%)	27	(93%)	27	(93%)	27	(93%)
With one or none	67	22	(33%)	48	(72%)	50	(75%)	45	(67%)
Exact test (p)		*p < 0.01*		*p = 0.029*		*p = 0.049*		*p = 0.010*	

The combination of the GD and SC tasks produce significantly better results for the graphical task (93%, p = 0.029; OR = 5.27 [1.12, 50.12]) and subsequently the Sound AP (93%, p = 0.010; OR = 6.50 [1.41, 61.42]) than using none or just one of the treatments. The written task has an odds-ratio that spans the null value, and thus is not considered significant (93%, p = 0.049; OR = 4.53 [0.96, 43.37]).

This result is an important finding for the double-loop learning activity. In the Simulation experiment, there was no difference between the physical and mental simulation treatment conditions. However, in this Dialogue experiment, the use of prompts appear to have both a significant and large effect. The SC treatment had a significant effect by itself, but the GD treatment by itself did not; however, the interaction between these treatment conditions (GDSC) produced the best results.

The causal logic test was used to further investigate whether the different treatment factors led to a better understanding. The two-factor analysis of variance showed no significant main effect on the causal logic task for the GD groups, F(1,141) = .04, p = 0.84, or for the SC groups, F(1,141) = 1.91, p = 0.17, or for the interaction between GD and SC groups, F(1,141) = 0.31, p = 0.58. However, a one-way analysis of variance revealed significant differences in causal logic scores between the different graphical response groups F(2,93) = 4.78, p = 0.01. Respondents with *Correct* graphs (M = 4.52, SD = 1.07) performed better in the causal logic task than respondents with *Decrease* graphs (M = 3.74, SD = 1.44) and *Incorrect* (M = 3.62, SD = 1.60). There was no significant differences in causal logic scores between the written match and mismatch response groups $F(1,94) = 0.67$, $p = 0.42$. The overall agreement rate with the group's response was 96%. There was no significant difference between the treatment conditions for the causal logic test.

5.3 Results by task

In this section, I present the results for the various tasks in the following order: graphical results, written results, graphical–written matches, causal logic results, group results, and a text analysis.

5.3.1 Graphical responses

The graphical responses for all the treatment conditions are presented in Table 5.15. The goal of the learning activity was to improve the mental models of the participants, demonstrated through improved and coherent assessment results. The correct graph and incorrect graph columns are the important indicators in this table, where fewer incorrect and more correct responses are desirable.

Table 5.15: Graphical response category for each treatment condition (frequency and row percentages).

Experiment/Treatment		Graphical Response						Σ
		Correct		Decrease		Incorrect		
Checklist	PC-ER	22	(13.1%)	58	(34.5%)	88	(52.4%)	168
	PC-AE	114	(36.1%)	66	(20.9%)	136	(43.0%)	316
Workbook	PW-AA	44	(40.7%)	30	(27.8%)	34	(31.5%)	108
	MW-AA	32	(34.8%)	30	(32.3%)	31	(33.3%)	93
	PW-GD	13	(39.4%)	11	(33.3%)	9	(27.3%)	33
	PW-SC	18	(51.4%)	13	(37.1%)	4	(11.4%)	35
	PW-GDSC	35	(68.6%)	11	(21.6%)	5	(9.8%)	51
	Σ	278	(34.6%)	219	(27.2%)	307	(38.2%)	804

There are some broad observations that can be made from these data. The most striking is that the SC treatment conditions (PW-SC and PW-GDSC) outperformed all others. The Workbook experiments results are better than the Checklist experiment, as can be seen with a small improvement between the Checklist (PC-AE: 36.1% correct, 43.0% incorrect) and Workbook (PW-AA: 40.7% correct, 31.5% incorrect) treatment conditions. The Physical simulation treatment condition (PW-AA: 40.7% correct, 31.5% incorrect) performed better than the Mental simulation treatment condition (MW-AA: 34.8% correct, 27.3% incorrect). The best treatment condition to achieve a Correct graph were the groups with Scenario cards (PW-SC: 51.4% correct, 11.4% incorrect; PW-GDSC: 68.6% correct, 9.8% incorrect).

5.3.2 Written descriptions

The Scaffold experiment identified that the written descriptions demonstrated a level of confusion when compared to the graphical responses. Subsequent workshops aimed to increase both the frequency of correct graphical responses and reduce the frequency of mismatching written descriptions. The frequency of matching and mismatching responses are shown in Table 5.16. A matching response does not mean a 'correct' understanding, rather a written description that supports (or for mismatch, contradicts) the graphical response.

Table 5.16: Written description matches for each treatment condition (frequency and row percentages).

Experiment/Treatment		Written Description				Σ
		Match		Mismatch		
Checklist	PC-ER	79	47.0%	89	53.0%	168
	PC-AE	190	60.1%	126	39.9%	316
Workbook	PW-AA	76	70.4%	32	29.6%	108
	MW-AA	63	67.7%	30	32.3%	93
	PW-GD	29	82.9%	6	17.1%	35
	PW-SC	45	88.2%	6	11.8%	51
	PW-GDSC	27	81.8%	6	18.2%	33

Similar observations to the graphical task results can be made for the written description data, shown in Table 5.16. The changes between Checklist (PC-AE: 39.9% mismatch) and Workbook (PW-AA: 29.6% mismatch) treatment conditions appears to have reduced the frequency of confusion in the test. Treatment conditions using scenario cards and/or a group diagram outperform all other treatment conditions with over 80% match responses, suggesting that these treatment conditions help the participants to describe their graphical response more clearly.

5.3.3 Graphical and written descriptions

In this section, the written–graphical responses are compared, as shown in the detail categories of Table 4.12. The two written description categories and three graphical response categories are shown in Table 5.17 for each treatment condition. The percentage column shows the percentage of responses for each category within the treatment condition. Written *Match* responses suggest good , regardless of graphical response.

Table 5.17: Graphical-Written response categories by treatment condition (frequency and treatment percentages).

Treatment/Written Description			Graphical Response						Σ
			Correct		Decrease		Incorrect		
Checklist PC-ER	Match		22	13.1%	43	25.6%	14	8.3%	79
	Mismatch		0	0.0%	15	8.9%	74	44.0%	89
		Σ	22	-	58	-	88	-	168
Checklist PC-AE	Match		113	35.8%	49	15.5%	28	8.9%	190
	Mismatch		1	0.3%	17	5.4%	108	34.2%	126
		Σ	114	-	66	-	136	-	316
Workbook PW-AA	Match		43	39.8%	29	26.9%	4	3.7%	76
	Mismatch		1	0.9%	1	0.9%	30	27.8%	32
		Σ	44	-	30	-	34	-	108
Workbook MW-AA	Match		32	34.4%	29	31.2%	2	2.2%	63
	Mismatch		0	0.0%	1	1.1%	29	31.2%	30
		Σ	32	-	30	-	31	-	93
Workbook PW-GD	Match		13	39.4%	11	33.3%	3	9.1%	27
	Mismatch		0	0.0%	0	0.0%	6	18.2%	6
		Σ	13	-	11	-	9	-	33
Workbook PW-SC	Match		18	51.4%	10	28.6%	1	2.9%	29
	Mismatch		0	0.0%	3	8.6%	3	8.6%	6
		Σ	18	-	13	-	4	-	35
Workbook PW-GDSC	Match		35	68.6%	9	17.6%	1	2.0%	45
	Mismatch		0	0.0%	2	3.9%	4	7.8%	6
		Σ	35	-	11	-	5	-	51

Note: Percentages may not equal 100% due to rounding.

The SC treatment conditions show clearly fewer instances in the *Mismatch-Incorrect* category (GDSC: 7.8%; SC: 8.6%) than other treatment conditions. This is an important result, and indicates that the SC treatment conditions either helped to make the graphical response explicit for the participant (fewer incorrect graphs), or connected the graphical response with the written description for the participant, clarifying the description (fewer mismatch descriptions).

To investigate the graphical literacy further, the two important categories in Table 5.17 are the *Correct-Match* and the *Incorrect-Mismatch* categories. The responses in the *Correct-Match* category show a sound understanding of the problem, whereas the responses in the *Incorrect-Mismatch* category typically show a sound written understanding described with a mismatching graph: a typical demonstration of *WG confusion*. Results from these two categories are summarised in Table 5.18.

Table 5.18: Summary of Graphical-Written responses by treatment condition (frequency and cell percentage).

Experiment/Treatment		Graphical-Written Description				Σ
		Correct-Match		Incorrect-Mismatch		
Checklist	PC-ER	22	(13%)	74	(44%)	168
	PC-AE	113	(36%)	108	(34%)	316
Workbook	PW-AA	43	(40%)	30	(28%)	108
	MW-AA	32	(34%)	29	(31%)	93
	PW-GD	13	(39%)	6	(18%)	33
	PW-SC	18	(51%)	3	(9%)	35
	PW-GDSC	30	(81%)	2	(5%)	37

The PC-ER treatment condition was the only treatment that more *Incorrect-Mismatch* category responses than *Correct-Match* category responses. The reduction of *Incorrect-Mismatch* responses in the PW-SC and PW-GDSC treatment conditions indicate a clear reduction of SF failure due to poor graphical literacy.

5.3.4 Causal logic

A causal logic test was included in the AA assessment sheet to further investigate the SF failure observed in the graphical-written responses for the Checklist experiment. The causal logic test results were used to help understand whether a misunderstanding was occurring through incomplete mental models or *WG confusion*.

The individual causal logic questions were analysed to understand the variability in responses, shown in Table 5.19. The questions that have an increasing flow and increasing stock were the most frequently correct responses (Q1: 96%; Q5: 86%) The questions with a steady or decreasing flow rate had the least frequent correct responses (Q2: 68%; Q3: 71%; Q4: 67%).

Table 5.19: Correct causal logic responses by treatment condition (frequency and column percentages).

Exp./Treat.	Q1		Q2		Q3		Q4		Q5		Ave/5	n
PW-AA	102	(94%)	73	(68%)	75	(69%)	67	(62%)	90	(83%)	3.77	108
MW-AA	91	(98%)	50	(54%)	56	(60%)	55	(59%)	83	(89%)	3.60	93
PW-GD	32	(97%)	27	(82%)	23	(70%)	24	(73%)	28	(85%)	4.06	33
PW-SC	35	(100%)	29	(83%)	32	(91%)	27	(77%)	31	(89%)	4.40	35
PW-GDSC	35	(95%)	29	(78%)	30	(81%)	31	(84%)	32	(86%)	4.24	37
Σ	295	(96%)	208	(68%)	216	(71%)	204	(67%)	264	(86%)	3.88	306

Note: the causal logic tasks were not included in the Checklist experiment assessment sheets; only Workbook experiments are shown.

The average causal logic sums for each treatment condition are also shown in Table 5.19. Participants in the SC treatment groups (PW-SC: 4.40; PW-GDSC: 4.24) outperformed other groups, with the mental simulation treatment (MW-AA: 3.60) performing the worst on the causal logic task.

The average causal logic sum for each treatment condition by level of graphical-written understanding are shown in Table 5.20. Participants demonstrating Sound AP performed better on average in the causal logic task (4.02) than participants with WG confusion (3.60) and SF failure (3.00). This result would suggest that participants drawing a correct graph with a matching written description have a more complete mental model of the problem at the time of testing than students with incorrect graphs.

Table 5.20: Causal logic scores by treatment condition and understanding level (frequency and average score out of 5).

Treatment	n	Sound AP		WG confusion		SF failure	
		Ave/5	n	Ave/5	n	Ave/5	n
PW-AA	108	3.86	72	3.58	31	3.60	5
MW-AA	93	3.62	61	3.67	30	2.00	2
PW-GD	33	4.46	24	3.17	6	2.67	3
PW-SC	35	4.46	28	4.33	6	3.00	1
PW-GDSC	37	4.37	35	2.00	2	0.00	0
Σ	306	4.02	220	3.60	75	3.00	11

5.3.5 Group effects

The T&P activity in all workshops relied on group work. Although the activities were designed to be uniform between groups, the experience of each group would have been different owing to the insights, conversations and internal group dynamics. These interactions were not recorded in the experimental data.

However, we can examine the results by group to better understand whether participants were influenced by their group. In this section, I explore the group effects by examining the extent group members agreed with their group, the likelihood of all members of the group producing a correct graph and the likelihood of drawing a correct graph based on the role within the group.

Group agreement

On the AA assessment sheet, individuals were asked to what extent they agreed with their group's solution in the T&P activity. For the group diagram treatment conditions, the group was explicitly asked to draw a diagram in the activity workbook together before the individual diagrams. Like all of the workbook tasks, these were not compulsory.

Almost all participants (M: 96%) agreed with their group regardless of treatment condition, as shown in Table 5.21. This reveals insights into group dynamics. As all participants completed their questionnaire individually, there is no peer pressure for students to feel compelled to agree with their group.

Table 5.21: Agreement rates by treatment condition (frequency and row percentage).

Treatment	n	Agree		Half agree		Disagree	
PW-AA	108	104	(96%)	1	(1%)	2	(2%)
MW-AA	93	90	(97%)	2	(2%)	1	(1%)
PW-GD	33	30	(91%)	2	(6%)	1	(3%)
PW-SC	35	33	(94%)	1	(3%)	1	(3%)
PW-GDSC	37	36	(97%)	1	(3%)	0	(0%)
Σ	306	293	(96%)	7	(2%)	5	(2%)

Group diagram

The group diagram is a clear point of reference for group and individual agreement. The group diagram was exactly the same task as in the individual assessment, used in the PW-GD and PW-GDSC treatment conditions. The group diagram result by the two treatment conditions is shown in Table 5.22. In both treatment conditions, the Correct graph was most frequently drawn; however, no group in the PW-GDSC treatment condition drew an incorrect graph, and subsequently no individual from the PW-GDSC treatment condition went on to draw an incorrect graph in the individual assessment.

Table 5.22: Group graphical response category for GD treatment conditions.

Experiment/Treatment		Graphical Response						Σ
		Correct		Decrease		Incorrect		
Workbook	PW-GD	4	(44%)	3	(33%)	2	(22%)	9
	PW-GDSC	8	(80%)	2	(20%)	0	(0%)	10
	Σ	12	(63%)	5	(26%)	2	(11%)	19

Examining the GD treatment conditions in Table 5.21, 66 (94%) of participants stated that they agreed with their group; however, comparing the group graph to the individual graph, 59 (84%) of participants drew the same individual graph as their group graph. Of the 11 (16%) of participants that drew a different graph than their group, 4 drew a graph that improved on their group's answer, and 7 drew a graph that was worse than their group's answer.

Likelihood of all group members drawing the same graph

This section investigates the frequency of all members in the group drawing a correct graph, all members in the group drawing an incorrect graph, or the frequency of no members in a group drawing a correct graph by treatment condition. The PC-ER, PC-AE, PW-AA and MW-AA treatment conditions had fewer than 5% of groups with participants who all drew a correct graph. The PC-ER, PW-AA and PW-GD treatment conditions had more than 10% of groups with participants who all drew an incorrect graph, shown in Table 5.23. The PC-ER (64%) and PW-GD (44%) treatment conditions had no member within the group who drew a correct graph.

Table 5.23: Groups with all correct or all incorrect graphical responses (group frequency and row percentage)

Experiment/Treatment		Graphical Response: All correct		All incorrect		No correct in group		Σ groups
Checklist	PC-ER	1	(2%)	8	(18%)	28	(64%)	44
	PC-AE	2	(3%)	4	(5%)	17	(22%)	78
Workbook	PW-AA	0	(0%)	3	(12%)	5	(19%)	26
	MW-AA	1	(4%)	1	(4%)	6	(26%)	23
	PW-GD	1	(11%)	1	(11%)	4	(44%)	9
	PW-SC	3	(30%)	0	(0%)	3	(30%)	10
	PW-GDSC	7	(70%)	0	(0%)	1	(10%)	10
	Σ	15	(8%)	17	(9%)	64	(32%)	200

Group roles

In all physical treatment conditions, participants took on a role within the group: either in charge of a pump or as an instructor/observer. The results by role are shown in Table 5.24. Participants controlling the Natural Additions pump performed better on average (Sound AP: 61%, n=104) than other roles. There are no significant results to indicate that the role determines performance across all treatment conditions.

Table 5.24: Sound AP by simulation role and treatment condition (frequency and Sound AP percentage).

Treatment	n	Anthropogenic Additions		Natural Additions		Natural Removals		Instruction/ Other	
PC-ER	168	16	(41%)	17	(40%)	10	(25%)	22	(47%)
PC-AE	316	34	(45%)	50	(65%)	35	(45%)	43	(51%)
PW-AA	108	21	(78%)	16	(64%)	17	(68%)	18	(60%)
PW-GD	24	5	(63%)	6	(75%)	7	(70%)	6	(86%)
PW-SC	28	9	(82%)	6	(75%)	8	(73%)	5	(100%)
PW-GDSC	36	9	(100%)	9	(90%)	8	(100%)	8	(89%)
Σ	306	94	(55%)	104	(61%)	85	(49%)	102	(56%)
		$p = 1$		$p = 0.09$		$p = 0.08$		$p = 0.93$	

Note: there were no roles for MW-AA treatment condition.

5.3.6 Text analysis

The text from the written description was analysed for frequency of terms using the process described in §4.3. The written descriptions that described correct and incorrect graphical responses were separated to investigate whether the language used in the written description varied. Table 5.25 shows the fifteen most frequently used words, alongside the percentage of responses that the word was used in. Grey cells indicate words that were included in either the ER and AE assessment sheets in the Checklist experiments or the AA assessment sheet in the Workbook experiments. These grey cells have been discarded from further analysis because they could be considered as textual prompts in the activity, but are shown here for context. Asterisks represent a wildcard, so *increas** would match increase, increasing, increases and increased.

Table 5.25: Frequently used words in the written descriptions by experiment and graph category (frequency of word count divided by number of responses)

Checklist experiment				Workbook experiment			
correct graph		incorrect graph		correct graph		incorrect graph	
emission*	86%	emission*	64%	stop	57%	slow	44%
anthropogenic	76%	anthropogenic	58%	pump	46%	pump	35%
need	52%	need	40%	slow	44%	stop	33%
absorption*	39%	absorption*	35%	addition*	35%	down	29%
natural	37%	natural	34%	anthropogenic	33%	addition*	27%
ppm	34%	increas*	23%	natural	25%	anthropogenic	20%
stop	32%	ppm	21%	removal*	22%	rate	19%
increas*	28%	reduc*	21%	down	22%	removal*	17%
reduc*	27%	slow	18%	level	22%	level	17%
slow	24%	rate	18%	need	21%	natural	15%
rate	22%	stop	17%	rate	18%	decreas*	12%
down	22%	decreas*	17%	exceed	16%	reduc*	12%
concentration	19%	down	14%	goal	16%	need	11%
must	19%	concentration	12%	decreas*	13%	goal	10%
equilibrium	18%	carbon	12%	complete*	11%	exceed	7%

Note: * represents a wildcard. Grey cells display words that are included in either the ER, AE or AA assessment sheets.

The remaining words appearing in the written descriptions were then further analysed. These were, in alphabetical order: *complete**, *decreas**, *down*, *increas**, *must*, *reduc**, *slow* and *stop*. The frequency of written descriptions using these words is shown in Table 5.26. The most notable observation from these data is the high proportion of correct graphical responses in the Workbook experiment that used the words *slow* (44%) and *stop* (57%) in the written description. *Stop* is also the most frequent word in responses in the Checklist experiment with a correct graph (32%).

Table 5.26: Frequently used words not used in assessment sheet by experiment (frequency of word count divided by number of responses).

	Checklist experiment				Workbook experiments			
Word	Correct graph		Incorrect graph		Correct graph		Incorrect graph	
complete*	17	(13%)	6	(2%)	16	(11%)	7	(4%)
decreas*	21	(15%)	41	(12%)	18	(13%)	13	(7%)
down	30	(22%)	38	(11%)	31	(22%)	21	(12%)
increas*	38	(28%)	90	(26%)	13	(9%)	6	(3%)
must	26	(19%)	35	(10%)	8	(6%)	3	(2%)
reduc*	37	(27%)	72	(21%)	12	(8%)	4	(2%)
slow	32	(24%)	52	(15%)	62	(44%)	34	(19%)
stop	43	(32%)	40	(11%)	81	(57%)	28	(16%)

*Complete** and *stop* represent two concepts central to the learning outcome of the T&P activity: in order for the atmospheric carbon concentration to stabilise, anthropogenic emissions must stop completely. Both words were used more frequently in conjunction with a correct graph than with an incorrect graph. This is an important observation and offers a clear and obvious opportunity for communicating this scenario to the broader public:

> Anthropogenic emissions need to come to a complete halt (which is not likely) or if not, the rate of emissions need to be reduced significantly.
>
> *-Response #21655, PC-AE*

> Be slowly reduced until it is stopped completely.
>
> *-Response #500610, PW-GD*

> Eventually stop completely. Assuming natural removals/ additions remain balanced (as per model) and no other removal (as per model).
>
> *-Response #601216, MW-AA*

The words *increas** and *reduc** were more likely to be used in the Checklist experiment, whereas the word *slow* was used more frequently in the Workbook experiment:

> Reduce anthropogenic emissions. Increase natural absorptions.
>
> *-Response #11044, PC-ER*

> The anthropogenic emissions must to reduce or even stop to stabilise at this level.
>
> *-Response #20742, PC-AE*

> Natural absorption must also increase so as to balance out a rise in anthropogenic emissions. And anthropogenic emissions must also go down. Must be in equilibrium.
>
> *-Response #20824, PC-AE*

> Slow very quickly to a stop, to keep it as it currently is, additions by the anthropogenic pump has to stop immediately.
>
> *-Response #500101, PW-SC*

> It should gradually slow down to a rate that causes equilibrium with the other pumps in terms of the carbon cycle.
>
> *-Response #601909, MW-AA*

This observation demonstrates the different language that participants were using in response to the problem presented in the test. The descriptions using *increas** and *reduc** typically describe the actions required of components in the conceptual target domain, such as *anthropogenic emissions* and *natural absorptions.* The descriptions using *slow* more directly describe the actions required of components in the conceptual source domain, such as *pumps.*

Further, the word *pump* appears in 1% of correct responses in the Checklist experiment, and in 46% of correct responses in the Workbook experiment. This suggests that participants were framing their written description based on their physical experience of using the pumps, and understand the conceptual target domain (the carbon cycle) through the physical analogue (the T&P system).

PART III SYNTHESIS

Preface

In Part I, I introduced the background for this thesis, examining stock-and-flow thinking and the carbon cycle, and philosophies of how knowledge is created. In Part II, I detailed my workshop methodology and presented my results. Here, in Part III, I conclude this thesis by discussing the implications of my findings in the context of STEM education.

In Chapter 6, I explore the results of the experiments in relation to the double-loop learning approach presented in Chapter 3. I answer the research questions outlined in Chapter 4, and discuss the outcomes of the experiments.

In Chapter 7, I conclude this thesis by drawing out challenges for educators using constructionist approaches in undergraduate STEM courses, and detailing opportunities for further work.

6 Analysis and Discussion

You can't think about thinking without thinking about thinking about something.
- Seymour Papert (2005)

I have organised this chapter around the implications arising from the experimental results. First, I discuss the research questions proposed in Chapter 4. This leads to a discussion of the difference between stock-flow failure and written-graphical confusion, and to a reflection on the difference between hands-on learning and constructionism. Finally, I conclude the analysis by proposing double-loop learning as a powerful idea.

6.1 Analysis of research questions

In this section, the three research questions are discussed, building on the results presented for the three experiments in Chapter 5.

6.1.1 Scaffolded refinement

Research Question: *Does display of instructions and information in the T&P activity reduce SF failure?*

Display of instructions and information within the T&P activity had a clear effect on participants' ability to draw the correct graph, and provide a matching written description.

There were three treatment conditions considered to answer this question: PC-ER, PC-AE and PW-AA. Each treatment condition used a different assessment sheet. The PC and PW treatments used different instructions: PC treatments used checklist instructions and the PW treatment used workbook instructions.

This research question investigates whether this display of information reduced SF failure. SF failure was measured using the levels of understanding described in Table 4.12. Results in the Sound AP category demonstrate sound understanding of accumulation principles in the given context: the graphical response shows a future anthropogenic emissions that fall significantly from current levels, and the written response aligns with the graphical response. Results in the WG confusion category demonstrate a contradiction between the graphical and written responses; for example, an increasing graph is coupled with a description that describes the solution as the anthropogenic additions coming to a stop. SF failure is shown where the graph is incorrect (stabilising or increasing), but is coupled with a description that supports this trajectory.

Therefore, the first research question is specifically investigating whether changes in the display of information reduces the frequency of responses in the SF failure category. Table 6.1 shows the treatment conditions and graphical-written categories.

Table 6.1: Scaffold experiment results by treatment condition for graphical-written categories (frequency and cell percentage of population).

Treatment	n	Sound AP		WG confusion		SF failure	
PC-ER	168	65	(39%)	89	(53%)	14	(8%)
PC-AE	316	162	(51%)	126	(40%)	28	(9%)
PW-AA	108	72	(67%)	32	(30%)	4	(4%)
Σ	592	299	-	247		46	
Exact test (p)		$p < 0.01$		$p < 0.01$		$p < 0.01$	

The treatment conditions were refined over three iterations. There was no significant change of SF failure between the first two treatments (+1%, PC-ER: 8%; PC-AE: 9%; $p = 1$). There was, however, a significant reduction of WG confusion between the PC-ER and PC-AE treatment conditions (–13%, PC-ER: 53%; PC-AE: 40%; $p < 0.01$ OR:1.70 95%CI [1.14,2.52]). This is notable, as the only difference between the two treatment conditions was the language used on the assessment sheet. This indicates that this was an important factor in reducing WG confusion, but had no effect on SF failure.

It is worth recognising that both WG confusion and SF failure demonstrate deficient mental models in different ways. SF failure demonstrates inadequate mental models in the problem space: the mental model violates accumulation principles. WG confusion demonstrates inadequate mental models around written and graphical literacy: most often a confusion between the rate of change and the absolute rate.

Recognising this, the results may indicate that single-loop learning was prevalent in these activities: participants were assimilating the T&P activity into their existing mental models, both of the problem (where SF failure was demonstrated) and of the graphing task (where WG confusion was demonstrated).

The difference between the PC-AE and PW-AA treatment conditions shows a not-significant reduction of both WG confusion (–10%, PC-AE: 40%; PW-AA: 30%; $p = 0.07$) and SF failure (–5%, PC-AE: 9%; PW-AA: 4%; $p = 0.09$). However, when combined into the graphical-written categories, there is a significant increase to the Sound AP category (+18%, PC-AE: 51%; PW-AA: 67%; $p < 0.01$ OR: 1.90 95%CI [1.18,3.10]).

This second result suggests that mental models were being refined about both the problem and the assessment task, but that it might be more accurate to describe this process as the mental models were being orientated, rather than challenged. However, it is clear that the display of instructions and information can have a significant effect on a participant's ability to demonstrate sound accumulation principles.

6.1.2 Physical simulation

Does representational form of the T&P activity reduce SF failure?

The two representational forms—a physical system, and a picture of a physical system—did not have a clear effect on participants' ability to draw the correct graph and provide a matching written description. Both the PW-AA and MW-AA treatment conditions used workbook instructions, and the AA assessment sheet.

Table 6.2 shows the treatment conditions and graphical-written categories. Note that Table 6.2 includes the results from all experiments that used the PW-AA treatment condition, whereas in Table 5.11, only the results within the Simulation experiment were included[18]. There was no significant difference for Sound AP within the PW-AA cohorts ($p = 1$).

18. Table 4.13 shows all experimental treatment groups

Table 6.2: Simulation experiment results by treatment condition for graphical-written categories (frequency and cell percentage of population).

Treatment	n	Sound AP		WG confusion		SF failure	
PW-AA ENVS1001	82	54	(66%)	25	(30%)	3	(4%)
PW-AA ENGN1211	18	61	(66%)	6	(33%)	1	(6%)
PW-AA VCUG3001	8	7	(88%)	1	(13%)	0	(0%)
PW-AA Total	108	72	(67%)	32	(30%)	4	(4%)
MW-AA	93	61	(66%)	30	(32%)	2	(2%)
Σ	201	133	-	62		6	
Exact test (p)		p = 0.88		p = 0.76		p = 0.69	

There were fewer instances of SF failure in the MW-AA condition (–2%, PW-AA: 4%; MW-AA: 2%) and fewer instances of WG confusion in the PW-AA (–2%, PW-AA: 30%; MW-AA: 32%). However, both these results are small and not significant.

Just like the modeller who constructs a model and then is surprised by its behaviour, I was surprised (and frustrated) to see that the physical simulation activity did not perform better than the mental simulation.

After careful consideration of the motives behind this research question and experiment, my interpretation is that this result was not the failure of the experimental conditions, but rather it demonstrates that the T&P activity was an easy-to-access and strong conceptual metaphor. Participants used the T&P system to think through the problem without interacting directly with the system. The behaviour of the system could be inferred, based on the participant's experience with container-schema logic.

This observation led to another between-group experiment, included in Appendix D. Second-year engineering students were given the bathtub test (similar to that described in Box 2.1, and shown in Box D.5), with competing cover stories: one using a pump system similar to the T&P system, one using a bathtub, and one using mathematical additions and subtractions. A summary of the results are shown in Table 6.3 for this linear dynamics task.

Table 6.3: Linear dynamics results by graphical response (column percentages)

Graphical response	Abstract		Bathtub		Pumps	
Correct	32	(43%)	49	(63%)	44	(60%)
Incorrect	42	(57%)	29	(37%)	29	(40%)
Σ	74	-	78	-	73	-
Exact test (p)	p = 0.01		p = 0.12		p = 0.39	

The Pumps treatment performed slightly better than the Bathtub treatment; however, there was no significant difference between the Bathtub and Pumps treatments ($p = 0.87$). The Abstract treatment performed significantly worse than other two, with participants only half as likely to draw a correct graph ($p = 0.01$, OR: 0.48 95%CI[0.26,0.87]).

This result supports the argument that the T&P system helps students to reason through basic stock-and-flow concepts. In this linear dynamics test, students were required only to mentally simulate the behaviour of the system. Students with the Pumps treatment were easily able to assimilate their mental models, and performed just as well as those with the Bathtub treatment.

The answer to this research question at this stage is that representational form does not appear to reduce SF failure. Other representational forms could be experimented with, such as the use of flight simulators and other computer simulation interfaces, or in the application to other problems and contexts. However, with almost a third of participants still demonstrating WG confusion, representational form alone does not appear to be an indicator of performance.

6.1.3 Focussed dialogue

Do opportunities for focussed dialogue during the T&P activity reduce SF failure?

The third and final research question examines the focussed dialogue loop of the double-loop learning model. The focussed dialogue loop is designed to encourage participants to discuss focussed questions during the activity. The motivation for this research direction came from Newell's (2012) argument about the importance of focussed dialogue in creating a shared understanding.

Two prompts were used at different stages of the activity: scenario cards to prompt thinking about future behaviour of the system, and a group diagram to encourage members to make their thinking explicit and shared. Table 6.4 shows the treatment conditions and graphical-written categories. Note that Table 6.4 includes the results from all experiments that used the treatment conditions, whereas Table 5.14 only included results within the Dialogue experiment[19].

Table 6.4: Dialogue experiment results by treatment category for graphical and written tasks (frequency and cell percentage of population).

Treatment	n	Sound AP		WG confusion		SF failure	
With GD	70	59	(84%)	8	(11%)	3	(4%)
Without GD	143	100	(70%)	38	(27%)	5	(3%)
Exact test (p)		$p = 0.03$		$p = 0.01$		$p = 0.71$	
With SC	72	63	(88%)	8	(11%)	1	(1%)
Without SC	141	96	(68%)	38	(27%)	7	(5%)
Exact test (p)		$p < 0.01$		$p < 0.01$		$p = 0.27$	
With GD and SC	37	35	(95%)	2	(5%)	0	(0%)
With one or none	176	124	(70%)	44	(25%)	8	(5%)
Exact test (p)		$p < 0.01$		$p < 0.01$		$p = 0.36$	
Σ	213	159	(75%)	46	(22%)	8	(4%)

Note: percentages may not equal 100% due to rounding.

The treatment condition tasks both show a significant improvement for Sound AP, with the effect of the scenario cards (+20%; with: 88%; without: 68%; $p < 0.01$ OR: 3.27 95%CI[1.45,8.14]) larger than the effect of the group diagram

19. Table 4.13 shows all experimental treatment groups

(+14%; with: 84%; without: 70%; $p = 0.03$ OR: 2.30 95%CI[1.06,5.33]). However, the interaction effect is greater still (+25%; GDSC: 95%; with one or none: 70%; $p < 0.01$ OR: 7.29 95%CI[1.76,64.9]).

Results are also positive for the reduction of WG confusion. The frequency of SF failure is also reduced with the treatments; however, due to the small group sizes the results are not significant. The fact that there were zero responses in the GDSC SF failure category, is a clear indication that the opportunities for focussed dialogue both increased frequency of Sound AP and decreased frequency of SF failure.

There is one further important observation to do with the participants demonstrating sound accumulation principles. The Sound AP category has been broken into the *Correct-Match* and *Decrease-Match* categories, shown in Table 6.5.

Table 6.5: Dialogue experiment results for Sound AP by graph type (frequency and cell percentage).

Treatment	n (S.AP)	Sound AP			
		Correct graph		Decrease graph	
With GD	59	43	(73%)	16	(27%)
Without GD	100	61	(61%)	39	(39%)
Exact test (p)		$p = 0.17$			
With SC	63	48	(76%)	15	(24%)
Without SC	96	56	(58%)	40	(42%)
Exact test (p)		$p = 0.03$			
With GD or SC	52	31	(60%)	21	(40%)
With GD and SC	35	30	(86%)	5	(14%)
With neither	72	43	(60%)	29	(40%)
Exact test (p)		$p < 0.01$			
Σ	159	104	(65%)	55	(35%)

When the Sound AP group is divided into correct and decrease graph categories, the GD treatment is not considered to be a significant indicator of whether participants could draw a correct graph (+12%, with: 73%; without: 61%; $p = 0.17$); however, the SC treatment is (+18%, with: 76%; without 59%; $p = 0.03$ OR: 4.02 95%CI [1.41,14.19]). Of participants with Sound AP, the GDSC treatment was an indicator of drawing a correct graph (+26%, PW-GDSC: 86%;

PW-AA: 60%; $p < 0.01$ OR: 4.00 95%CI[1.32,14.6]). The prompts in combination were useful for both helping participants move to the Sound AP category, and for drawing a correct graph.

The answer to the third research question is clearly in the affirmative. These small opportunities for focussed dialogue through the activity improved the likelihood of participants achieving an answer with a sound understanding of accumulation. Further work could examine how different prompts could influence the results in different ways; for example, using flight simulators to explore the cause-and-effect of different policy decisions, or scenario planning to explore different possible futures.

6.2 SF failure or graphical confusion?

From the early experiments, it was clear that understanding how SF failure occurs was not straight-forward. It was clear, however, that participants were frequently providing confused responses: for example, a graph that demonstrated the correlation heuristic was often described in the written response as "decreasing to zero" (PC-ER: *Incorrect–Mismatch*: 44%).

I propose two broad explanations for this mismatch. An SF failure explanation assumes that the graphical response is a true representation of the participant's mental model. This would support the high rates of pattern-matching that previous studies have shown, and would suggest that SF failure was indeed prevalent. The alternative explanation is that the written description is a better representation of their mental model and there is confusion with the graph: written-graphical confusion. This would support an argument that suggests that the participants carry mental models that have the correct cause-and-effect logic, but do not have the graphing skills to represent this correctly.

This WG confusion explanation has not been considered in any detail in previous studies, which assume that many of the students involved were highly educated STEM students, particularly graduate students at MIT and Harvard. The cohort in my study, too, have high levels of mathematics training. Engineering students are required to have advanced maths (including calculus) as a prerequisite for the degree, and study two courses in calculus and linear algebra in their first year.

Further, there was little evidence that students in environmental science, where there is not this requirement, performed better or worse. Perhaps describing WG confusion as SF failure could be seen as a collective blind spot, as described in §3.2.4.

The WG confusion explanation is a reasonable line of inquiry for three reasons:

1. calculus teaching: students learn differentiation and integration largely in terms of manipulation of rules and formulae rather than in the visual representations presented in behaviour over time graphs
2. the dissipation process: described in Figure 6.1, where the stock of memorised facts for an exam or quiz are forgotten over time, especially as a surface learner
3. teaching approaches and standards: in many situations, a student can pass a course with a mark of 50%. When courses are predominantly test-based, a student could perceivably pass the course with perfect knowledge of half the material, or half the knowledge of all the material.

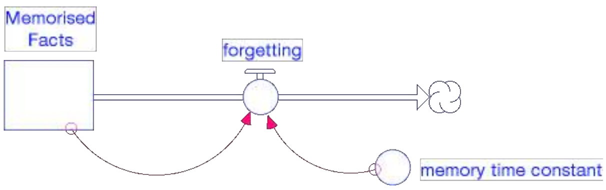
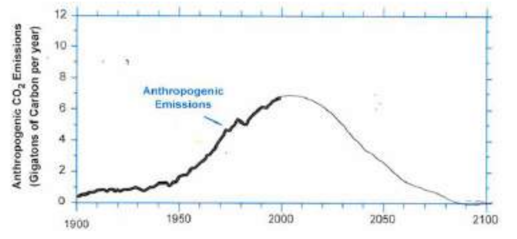
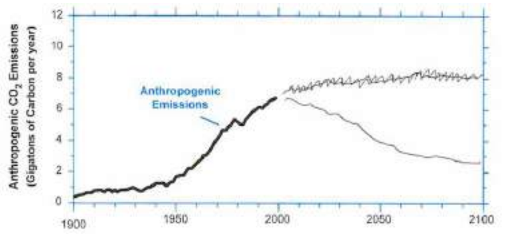
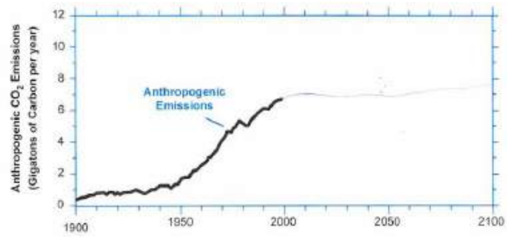
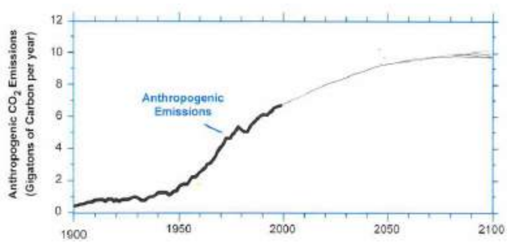


Figure 6.1: Generic structure of the dissipation process, in this instance forgetting memorised facts (from Richmond, 2010).

These areas of uncertainty make room for the argument that suggest the graphical literacy skills may be inadequate to answer the test as intended. I argue that the written description is a better representation of the participant's mental model than the graphical response. Students, even of STEM topics, are likely to have an adequate working written literacy ahead of an adequate working graphical literacy.

The WG confusion can be seen quite clearly in the examples in Box 6.1, where all responses discuss the anthropogenic emissions slowing down and then stopping in order to reach the goal level. All graphs, however, show different trajectories.

Box 6.1: Different graphical responses with similar written description. Either the graphs are what the participant intended, and the written statements confirm increasingly incorrect mental models, or the written statements are correct, and the different graphs show different levels of graphical literacy.

	*Response #20122 - correct graph* *[Correct-Match]* Anthropogenic emissions need to slow down and eventually stop to stabilise atmospheric GHG concentrations.
	*Response #20841- decrease graph* *[Decrease-Match]* Anthropogenic emissions have to be reduced significantly if not completely. The other way is to find an alternative absorptions (engineered solutions) in addition of natural absorptions.
	*Response #20354 - stabilise graph* *[Incorrect-Mismatch]* The anthropogenic emissions rate needs to be reduced and optimally stopped.
	*Response #20211 - increase graph* *[Incorrect-Mismatch]* Needs to slow, then stop completely. (if it doesn't stop completely, it will exceed 400 ppm, regardless of how slow it goes. Natural absorption could increase.)

In all the workshop activities, no group deliberately had water exceed the top of the atmospheric carbon tub. Further, no group engaged in a strategy that significantly raised the atmospheric carbon level before reducing it, so it is strange to see this as a result in the graphical responses.

The SF failure and WG confusion explanations may indeed co-exist in the responses. Some incorrect responses may be due to flawed mental models, and some incorrect responses may be due to poor graphical literacy. A *Correct* graph coupled with a description that matched the graph (see #20211 in Box 6.1) demonstrates a high likelihood that the participant understands the dynamics and is graphically literate. An *increase* graph coupled with a description that mismatched the graph could demonstrate either a misunderstanding of the activity or written-graphical confusion.

To examine the difference between the graphical and written responses further, the results for the SF failure and WG confusion are shown in Table 6.6. If only examining the graph, an incorrect graph could be described as SF failure. However, in most cases the incorrect graph is accompanied with a description that does not match the graph, shown in the 'Both' column of Table 6.6.

Table 6.6: Incorrect responses for graphical and written tasks by treatment condition (frequency and percentages as a proportion of the sample population).

Experiment/Treatment		Just SF Failure (Incorrect graph)		Just WG confusion (Mismatch description)		Both (Incorrect-Mismatch)		Σ
Checklist	PC-ER	88	(52%)	89	(53%)	74	(44%)	168
	PC-AE	136	(43%)	126	(40%)	108	(34%)	316
Workbook	PW-AA	34	(32%)	32	(30%)	30	(28%)	108
	MW-AA	31	(33%)	30	(32%)	29	(31%)	93
	PW-GD	9	(27%)	6	(18%)	6	(18%)	35
	PW-SC	4	(11%)	6	(17%)	3	(9%)	51
	PW-GDSC	5	(10%)	2	(5%)	2	(5%)	33

The observations in these results challenge the notion that SF failure is due to inadequate mental models of the problem, and may be due largely to inadequate graphing ability, even with well-educated participants. This result has even broader implications to education systems than Sterman's call for innovative teaching of systems principles: if the larger problem is graphical literacy, educators need to re-examine the way that students learn about fundamental mathematics.

The T&P activity should be viewed as a way forward for this problem. In an activity that took no longer than an hour, different conditions saw SF failure reduce from approximately half of all responses to less than 10 per cent of responses. This demonstrates the power of carefully designed activities to facilitate learning.

6.3 Hands-on learning is not the same as constructionism

The distinction between hand-on learning and constructionism became extremely apparent to me after the first trial workshop with students, where I struggled to get students to understand what was happening in this fantastic hands-on activity that *I* had created *for* them.

During this initial failure, I was reminded of my undergraduate engineering education experience: apparently pointless problem sets, cut-and-paste-from-the-internet-until-the-code-works programming exercises, and cramming for endless and meaningless exams. Now sitting on the other side, I was getting frustrated at my students for apparently experiencing the same thing.

The difference in performance occurred when the activity switched from being just a hands-on activity to a double-loop learning activity. That is, when the learning in the activity shifted from an experience I shared with students (the checklist experiments) to being an experience that the students created for themselves based on my prompts (the workbook experiments). This, of course, failed at first, and is described in §D.4, but by changing my own mental models about the problem, I have been able to leverage constructionist pedagogy with modest success.

This observation has had a pervasive effect on observations of my own teaching and teaching in my discipline. Engineering has always had opportunities to engage with experiential, hands-on learning, particularly because of its application of natural sciences to the construction of real-world objects. The hardware lab is a classic example of this. However, many laboratories are created to mirror the checklist process of workbook exercises, rather than being designed to provide opportunities for students to engage in constructionist learning.

> It seems most efficient to provide students with step-by-step assistance, tools, and tricks to organize their thoughts and get to a finished product as quickly as possible. This well-intentioned support may in fact have the effect of stifling creativity and forcing students to create products that simply mirror the checklist they have been given. Students, especially the "good" students who have been well-trained to follow directions, will march through the steps with little thought at all.
>
> - Martinez & Stager (2013)

A recent conversation with a laboratory technician comes to mind. He was (politely) complaining about having to set up a later-year undergraduate electronics hardware lab to the point where the lab work was essentially a plug-and-play activity. He would much prefer to see the students having to design and assemble the circuitry themselves. He said that there was not enough time for the students to do this within the laboratory session, as they didn't have the required hands-on experience with the equipment. This situation is a damning example of the state of engineering education, and is no doubt replicated (and not recorded in the education literature) across the country and the globe.

It could be argued that making things accessible and technically easy for students in practicals is a useful approach for educators, as students focus on the theoretical grounding required to pass examinations in preparation for the workforce. However, Trevelyan argues that the pedagogy in engineering education does not even align with actual engineering practice. The engineering education literature "reveals an overwhelming belief that engineering is all about solitary technical work: technical problem-solving and design" (Trevelyan, 2010, p. 384). The important skills required in professional practice are not learnt at university, such as life-long learning (Trevelyan, 2014). There is no reason why these skills could not be learnt at university, or indeed be the primary focus of a university program in engineering.

The founder of the field of system dynamics, Jay W. Forrester, recognised the disconnect between education practices and the need to develop creative problem-solvers in a 1967 symposium to the USA National Academy of Engineering:

> [The engineer] must identify the significant and critical problems, but in his education, problems have been predetermined and assigned. He must develop the judgment to know what solutions to problems are possible, but in school the problems encountered are known to have answers. He should be excited by new and unsolved challenges, but for 20 years he has lived in an educational system where he knows he is repeating the work of last year's students.
>
> - Jay W. Forrester
> (in National Academy of Engineering 1970)

The ‘maker’ movement and its proponents (makers) may offer direction in this area. Makers explore engineering disciplines, such as electronics, mechatronics, materials science and design through play and tinkering with physical objects. Making, hacking, tinkering, failing and improving are everyday activities for makers.

Martinez and Stager (2013) describe constructionism as the best way to ‘do’ constructivism. In Seymour Papert’s Constructionist Learning Lab, he outlines the philosophy behind this particular maker space with eight big ideas, paraphrased from Stager (2005) below:

1. learn by doing
2. technology as a building material
3. through ‘hard’ fun
4. by learning to learn
5. by taking the proper amount of time for the job
6. you can’t get it right without getting it wrong
7. do unto ourselves as we do unto our students
8. knowing about digital technology is as important as reading and writing

Making and tinkering puts the student in a social environment where the role of ‘instructor’ is removed and replaced with ‘facilitator’. Open-ended challenges are effective prompts, which the students solve through constructing objects that satisfy the question. For example, “Make something that will start here and end there in one minute” (Martinez & Stager, 2013). Students are encouraged to solve problems that are important to their task, and the role of the facilitator is to provide prompts that extend the students' inquiry. For example, when a working prototype has achieved the first goal, the goal can be stretched: *can you make something that will start here and stop there in 10 seconds?*

Building such a learning environment in a university classroom is not straightforward. It is in stark contrast to the prevailing pedagogy of lectures, problem sets, and theory exams. These approaches ignore opportunities for the student to reconstruct the theoretical knowledge that a *start-here, stop-there* design challenge could offer. The challenge could be used to explore a variety of concepts, approaches and intangible benefits that might be required to complete the challenge, such as the social interactions required for team work, communication to seek new perspectives, collaboration to troubleshoot ideas, and returning to the drawing board when mental models are proved wrong; essentially, learning how to learn.

6.4 Double-loop learning as a powerful idea

Throughout the design of the T&P activity and through the results, it became clear that careful design of the activity could allow students to use the T&P system to create their own knowledge about the carbon cycle. As students worked their way through the workbook, there were countless shared 'aha' moments and deep, meaningful discussion among students as they began to realise that zero emissions is drastically different to zero growth in emissions.

The most memorable of these took place in a local primary school, where I was running an activity alongside Paul Compston, Barry Newell, and a classroom teacher with a group of about 20 year 3 to year 6 students. Towards the end of the session, I asked the students to think about what sorts of things they could do in their daily life to help reduce the atmospheric carbon stock. Students proposed plans to plant trees, walk to school, recycle at home, et cetera. One student talked about a future where everyone would drive electric cars. But, unlike the previous ideas, the student next to him admonished him. He had had his 'aha' moment. He exclaimed to the class that, "This would not be enough!" It could only ever stop the water level from going up. Not only should we stop the water from rising but we need to do things that would bring the level down.

The understanding of this primary school student has had a lasting effect on me as an educator. It was clear that this student, without any formal training in system dynamics, let alone textbook calculus, was able to clearly understand the causal logic that undergraduate STEM students were getting confused about. For me, this was a first-hand experience of the type of intuition that I had heard Jay W. Forrester talk about at his fireside chat with Khalid Saeed at the 2012 *International System Dynamics Conference* in Washington. Forrester had described young children intuitively understanding complex dynamics, such as that which govern the accumulation of friends in the playground, and how good or bad behaviour can change this accumulation.

> It's much easier to teach system dynamics in grades one through eight than it is any time later, because they have a lot less to unlearn. All the fundamental ideas can be grasped in the first few years. There have been people who quite successfully explained to kindergarteners the importance of stocks and flows. The water in the bathtub is a stock. The flows are obvious. Your reputation is a stock. The good and bad things you do are the flows. And they can go through their environment and find the stocks and the flows. By fifth and sixth grade they can be doing computer simulation models that, for the most part, are now being taught in graduate school.
>
> - Jay W. Forrester in conversation with Khalid Saeed (2013)

My experience with this primary school student showed me first-hand that there was a powerful idea at play. However, it wasn't until the last experiment that I realised that the powerful idea was not the T&P activity itself, but rather the process of double-loop learning that prompted students to have their own 'aha' moments.

7 Conclusion

The best way to construct knowledge or understanding is through
the construction of something shareable, outside of a student's head. [...]
such artifacts are evidence of learning.

- Martinez & Stager (2013)

The aim of this thesis was to improve how engineering educators understand the use of concrete experiences in the learning process.

I investigated this aim by creating a physical simulation activity to help students discover for themselves the dynamics of the carbon cycle, and map the future trajectory of anthropogenic additions required to stabilise the level of atmospheric carbon. I found that alongside the physical simulation itself, informed, scaffolded dialogue is essential for students to achieve a sound understanding of the problem.

7.1 Lessons for educators

The double-loop learning model has been the backbone of this thesis. Along with the T&P activity itself, the model is also the major outcome of this thesis.

The variables used in Sterman's double-loop learning model were used to examine mental models in the Chapter 3 literature review. The research methodology in Chapter 4 was framed around the instance of the double-loop learning model for physical simulation. Whilst the lessons from this thesis could be applied broadly to many educational fields, here I generalise my experience from the T&P activity to opportunities and challenges for constructionist approaches in engineering education. I do this by examining considerations for practical classes in engineering through the links in the double-loop learning influence diagram, annotated in Figure 7.1.

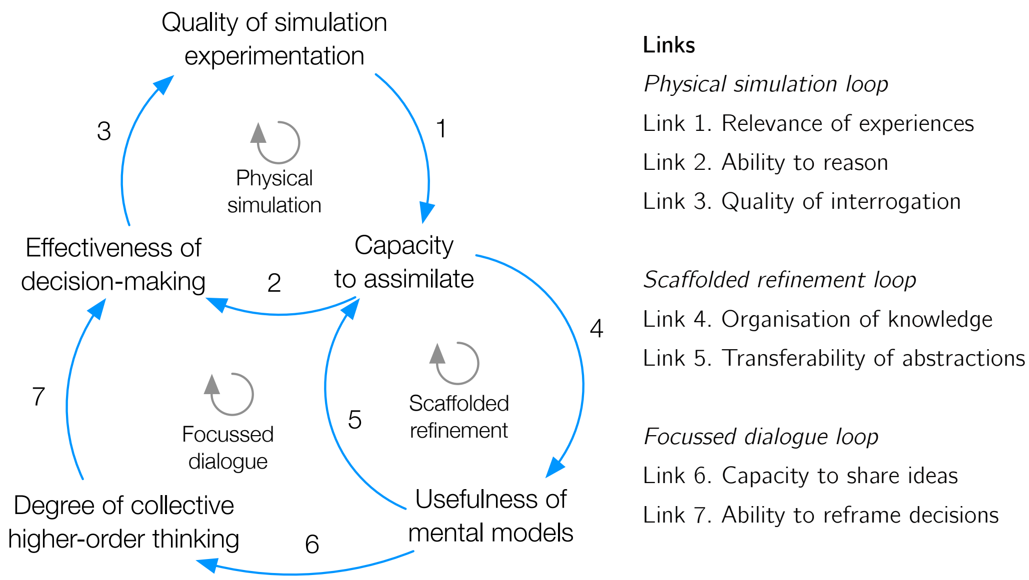


Figure 7.1: Annotated influence diagram of double-loop learning. The links are numbered and referred to in-text below.

Physical simulation loop

Link 1. Relevance of experiences

> Practical activities should be relevant to the learner's experience with the real world. Building learning activities based on conceptual metaphors can help the learner assimilate new knowledge. Keeping the conceptual metaphor simple and flexible makes the experience generalisable to other situations.

Practical activities should be relevant to the learner's experience with the real world. Building learning activities based on conceptual metaphors can help the learner assimilate new knowledge. Keeping the conceptual metaphor simple and flexible makes the experience generalisable to other situations.

Insights from the T&P activity

The T&P system itself did not change throughout the workshops; however, the way that students experienced the system did. When I asked students to assemble the system in Trial workshop 2, confusion was rife, and no groups were able to independently build the system as I had intended.

In the Checklist experiment, students picked up an assembled system and were asked to identify which pump they were controlling (Sound AP: PC-AE 51%; n=162). In the Workbook experiments, students were given an assembled system, but were required to label the components in the system before the activity began (Sound AP: PW-AA 67%, n=72).

Further, the T&P system was easily understood, to the extent that the students could simulate the behaviour of the system without interacting at all with the physical system (Sound AP: MW-AA 66%, n=61).

See the discussion about Trial workshop 2 in §D.3, The rationale about the decision to use a pre-assembled system in §4.1.1, and the summary result between treatment conditions in §5.1.1.

Link 2. Ability to reason

> The practical should orientate the learner to the interactions within the problem space, and help the learner reason about the problem clearly. Obstacles that are impeding the learner's ability to reason should be removed.

Insights from the T&P activity

The activity originally described the carbon cycle using the terms 'emissions' and 'absorptions'. Observations during the Checklist experiment indicated that students were having trouble negotiating the terminology, and that this was inhibiting their ability to reason. In the Workbook experiments, the terms 'additions' and 'removals' were used instead, and this confusion appeared to reduce.

The written descriptions between the Checklist and Workbook experiments somewhat demonstrate this change. Considering responses that drew a correct graph, *emission** was the most frequently used word in the Checklist experiment, used in 86% of responses; in the Workbook experiment, *additions* was used in 35% of responses. Whereas, the most frequently used word in the Workbook experiment was *stop*, used in 57% of responses; Checklist experiment: 32%. The students in the Workbook experiment appear to be reasoning with the problem more clearly than the students in the Checklist experiment, who were negotiating the terminology.

See the discussion about the change in language in §4.1.1, and the text analysis results in §5.3.6.

Link 3. Quality of interrogation

> The practical should enable students to conduct high-quality, open-ended interrogation of the ideas.

Insights from the T&P activity

The simulation environment should be consistent with the behaviour of the problem that is being modelled. In early trial versions of the activity within my research group, the pipes between the tubs were not fixed, and instead relied on someone in the group to hold them in place. This meant that effort was being spent dealing with the mechanics of the activity, rather than spending that time learning. Further, with the pipes being placed into the tubs, students could easily change their level in relation to the water, meaning that what was an additions pump could easily turn into a removals pump, violating the intended structure of the system.

Eventually, the pipes were cut to exact sizes, and holes were drilled into the tubs at appropriate levels and made watertight. Although there were other small mechanical problem with this, such as small amounts of water leaking out of pumps, tubs near toppling due to the vigour of pumpers, and pumps occasionally breaking, the majority of the effort was engaging with the system and interrogating the behaviour of the system.

Scaffolded refinement loop

Link 4. Organisation of knowledge

> Practical activities should be readily organised into learners' existing stock of mental models. Careful attention is required to prompt learners to check-in with and question their mental models continuously through their learning, based on observations from the practical.

Insights from the T&P activity

Simulation is an ideal activity to interrogate a problem. However, simulation itself does not ensure that the interrogation is challenging or useful; opportunities to reflect on the simulation environment and its meaning are important to include in the instructions.

Reflecting back on the Checklist experiment, the nine dot-point instructions describe the desired simulation behaviour I wanted students to experience. Student could conceivably go through the activity with little understanding of what they are doing or why. In the Checklist experiment, written-graphical confusion was common (WG confusion: PC-ER: 54%). This was a missed opportunity for their learning.

In the Workbook instructions, students were orientated and prompted to think about the activity. This was done a number of ways. The instructions included behaviour over time graphs, helping to show the intended behaviour over time. Headings in the instruction tasks were used, such as 'Task 3: Simulate the Natural Carbon Cycle', which may have contributed to the orientation of ideas through the experiment. Further, targeted questions were used through the experiment:

- What real-world thing does the water in your system represent?
- [D]iscuss what kinds of real-world things are represented by the Natural Additions and Natural Removals pumps. What happens when one of the pumps is running faster or slower than the other?
- [D]iscuss with your group how to stop the atmospheric stock from overflowing.

- Describe what needs to happen to the Anthropogenic Additions pump to ensure that the goal level is not exceeded (assume that the Natural Additions and Natural Removals remain approximately the same).

These small questions prompted students to think both about the meaning of the activity and their actions within it. On reflection, there could have been many more opportunities within the instructions to do this; however, written-graphical confusion reduced over the experiments (WG confusion: PC-AE: 40%; PW-AA: 30%; PW-GDSC: 5%).

See the instruction sheets in Appendix A for the full instructions used in the experiment, and the discussion about the development of all the workshop instructions in §4.2.

Link 5. Transferability of abstractions

> Learners should use the practical to develop mental models that improves their capacity to understand the world around them, and reason with the theory in useful ways.

Insights from the T&P activity

The learning environment should be simple and flexible, as to allow creative and imaginative thought about the problem at hand. This observation aligns with the higher-order thinking skills, such as abstraction, generalisation and transferring to new domains.

The debriefing discussions about the activity were rich for demonstrations of generalisable knowledge. These discussions were not formally part of the experiment. The discussion was based on a series of semi-structured questions that had prepared answers. These evolved and changed based on the direction of discussions, but centred on:

- how would you improve the T&P system to better represent other important aspects of the carbon cycle?
- what could you (your family/community/city/country) do in your daily life to reduce your anthropogenic additions to zero?
- what would happen to global emissions if you (your family/community/city/country) decided to drop to zero?

The answers to these questions, which were asked after assessment, demonstrate how students were able to assimilate their experience of the T&P into their mental models, and relate these to other situations.

For example, students typically improved the system by suggesting adding an anthropogenic removals pump. Then showing students the relatively insignificant amount of carbon that is currently removed anthropogenically from the atmosphere compared to the burning of fossil fuels (the size of the economy) helped them to see the importance of, for example, decarbonising the economy. Increasing the speed of the removals pump was another common strategy, through planting more trees. This was an opportunity to demonstrate the effect of deforestation, and highlight the distinction between different types of stocks, such as that trees are buffers, and both absorb and respire carbon and decay on short timeframes, returning carbon into the atmosphere and to the soil.

These examples show how students were using the T&P activity as the source domain for their thinking about the problem, and could imagine how the system would behave with a new structure, new rules or new information. See the discussion in §6.4 for further examples of this creative thought.

Focussed dialogue loop

Link 6. Capacity to share ideas

> The practical should prompt learners to make their thinking explicit to others. This both prompts learners to organise their knowledge more carefully, and provides an opportunity for learners to collaborate and learn from each other.

Insights from the T&P activity

The Dialogue experiment shows how thinking was made explicit through the creation of a graph. Creation is a fundamental part of constructionist pedagogy, although admittedly a better task may have been to get students to build new models of their own, collaborate on interesting projects, or write a poem.

The group simulation nature of the activity helped to get students interacting with each other's ideas. The scaffolded questions through the workbook helped to create a shared environment for knowledge creation. The construction of a graph as a group had an effect on the performance of individuals in the assessment task, shown by the improvement in Sound AP between the with GD groups (81%) and without GD groups (67%).

However, of the nine groups in this treatment condition, only four drew a 'correct' group diagram. One group in this treatment condition went on to have solutions that were all incorrect, and four groups went on to have no individual graphs correct. This could be an example of group think, where the focussed dialogue was useful for reaching a shared understanding, but not a correct shared understanding.

See the results in §5.2.3 for the dialogue experiment, and the discussion in §6.1.3 for further examples.

Link 7. Ability to reframe decisions

> The practical should enable individuals to collectively reframe their decisions and allow them to explore ideas that are outside their individual capacity to generate.

Insights from the T&P activity

The combination of the scenario cards and the group diagram was the most effective treatment condition to help students demonstrate a sound understanding of accumulation principles. The scenario cards were a significant indicator of drawing a correct graph (with SC: 56%; without SC: 33%; $p = 0.039$).

However, it was the combination with the group diagram that achieved the best results. The GDSC treatment condition (Sound AP: PW-GDSC: 93%) outperformed the base treatment (Sound AP: PW-AA: 67%), and the original treatment condition (Sound AP: PC-ER: 39%). The GDSC treatment was also an indicator of drawing a correct graph (with both: 76%; with SC, GD or neither: 33%), and groups using scenario cards and group diagram outperformed groups just using the group diagram task (PW-GDSC: 80%; PW-GD: 44%). No groups in the PW-GDSC group drew an incorrect graph, and subsequently no individuals drew an incorrect graph.

This highlights the importance of not only focussed dialogue, but informed focussed dialogue. See the results in §5.2.3 for the Dialogue experiment, and the discussion in §6.1.3 for further examples.

7.2 Further research

There is, of course, more to be done. To explore the further work, I will examine the links from the double-loop learning model again, in the frame of opportunities to improve in the T&P system as a tool for teaching accumulation principles.

Physical simulation loop

Link 1. Relevance of experiences

Further work is required to explore how the system can be applied to other contexts. The Creative Learning Exchange (CLE) house a large range of materials designed for educators to use, such as graphing the accumulation of money in a bank account, or the processes of trees growing in a forest. These are ready opportunities for physical simulation using a system like the Tubs & Pumps.

Link 2. Ability to reason

There are opportunities to explore how students interact with the T&P system and their in-situ ability to reason. Future work is required to see whether the physical simulation can be transferred to other media to enhance the experience, such as systems built on electronic sensors with in-built feedback, interacting with real-time graphs, providing formative feedback through the activity.

Link 3. Quality of interrogation

The process of physical simulation is a sensory experience. However, the physical system can be easily manipulated in ways that don't observe feedback processes, and can't be easily changed into other situations. Future work is required to understand whether the quality of interrogation in physical simulation could be improved, perhaps with the introduction of flight-simulator approaches, and flexible scenarios where students could be stretched once they have mastered one situation to examine new contexts and different policy environments.

Scaffolded refinement loop

Link 4. Organisation of knowledge

Further work is required to examine whether there are better ways to present tasks and instructions. The growth of grass-roots maker movements, online user created content, and purposeful use of technology may be an opportunity to engage in better modes of delivering instructions and prompts, and better aligned with the way that students engage with content.

This has relevance to the display of information in many different activities in engineering education, such as the display of content, worksheets, laboratory instructions, assignment guides and exams.

Link 5. Transferability of abstractions

The T&P activity was designed to be incorporated into existing tutorials as a short introduction to systems thinking. However, these courses were not system dynamics courses, and theory was not developed further. The T&P system could serve as a powerful base metaphor for a course on system dynamics or that develops systems thinking skills further.

Research is required to investigate whether the T&P activity could serve as a powerful, or at least useful, idea for students beginning to study system dynamics.

Focussed dialogue loop

Link 6. Capacity to share ideas

Students need the opportunity to construct their own ideas about the problem, and then test them in a safe, encouraging environment. However, the results suggest that the group diagram orientated some groups, and disorientated others. Timely, formative feedback may be a useful technique to help groups of students orientate their thinking about the problem. Further research could examine the effect of formative feedback on the group decision-making process.

Link 7. Ability to reframe decisions

The scenario cards were useful to help students draw the correct graph. However, it is not clear whether students used the graphs and text on the scenario cards to generate their own mental models about the problem, or whether they adopted the solution and made it fit their mental model. Further work could examine the use of other techniques for reframing thinking as a group, such as scenario planning between groups, to prompt broader discussion.

As a conclusion, it is worthwhile reflecting on the future of the T&P activity in relation to helping students construct their own understanding of accumulation principles. This thesis has made clear that there are directions forward in the field of SF failure, most notably through building the capacity of students to distinguish between stocks and flows. Future work is required to transfer the lessons of this experiment in physical simulation to other contexts and theory domains, to help students understand other concepts. And, perhaps most importantly, nurturing the powerful idea of using a constructionist approach to double-loop learning to teach fundamental accumulation principles. The lessons that come out of this thesis have relevance for engineering educators and other disciplines where accumulation principles are paramount, both at university and other teaching scenarios.

This page is intentionally blank

8 APPENDIX

Appendix A Workshop instructions

A.1 Checklist workshop instructions

GREENHOUSE GAS (GHG) WORKSHOP INSTRUCTIONS

In your groups of 4, allocate the roles of:

- Anthropogenic emissions
- Natural emissions
- Natural absorptions
- Instructor/Observer

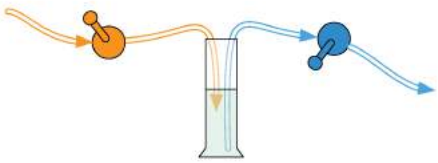
Please read these instructions out aloud before starting the activity. During the activity, the Instructor/Observer should prompt the group.

1. Ask the **Natural Emissions** to pump to a level of half the capacity of the atmospheric stock, then stop
2. Ask the **Natural Absorptions** to drain the atmospheric stock completely, then stop
3. Ask the **Anthropogenic Emissions** to pump to a level of half the capacity of the atmospheric stock, then stop
4. Now ask the **Natural Emissions** and **Natural Absorptions** to pump in a relationship to maintain the water level at half the capacity of the atmospheric stock. Prompt the emissions to speed up and then slow down and observe the water level.
5. Ask the **Natural Emissions** and **Natural Absorptions** to concentrate on matching their flow rates, rather than looking at the atmospheric stock. This represents **Dynamic Equilibrium**.
6. Now ask the **Anthropogenic Emissions** to pump at a very slow rate, and prompted to increase that rate over a period of time
7. When the atmospheric stock reaches a level about 2cm from the top of the atmospheric stock, the **Observer** will ask the group to troubleshoot how to stop it from overflowing. Keep in mind that the **Natural Emissions** and **Natural Absorptions** should roughly stay the same.
8. You can repeat step 7 to see if other strategies work.
9. Participants discuss the observed dynamics

A.2 Dialogue experiment workbook

The Carbon Cycle

Tubs & Pumps



Chris Browne, Barry Newell & Paul Compston
Australian National University

Contact: chris.browne@anu.edu.au

Title

Overview

In this activity you are going to build a model of the natural carbon cycle, and then look at the role that burning fossil fuels has on the level of carbon in the atmosphere.

You should have:

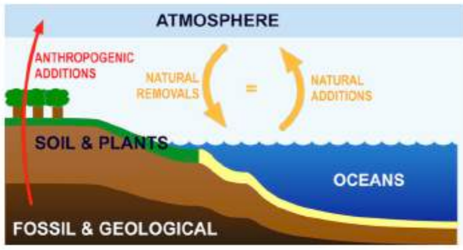
- one larger tub
- one smaller tub
- three hand pumps
- pipes to connect the pumps to the tubs
- water

Some rules:

- explore the model by playing with it
- while playing, there are no 'wrong' answers
- remember to have fun!

Introduction to the Carbon Cycle

Carbon moves naturally between stocks, such as the oceans, soil, plants and the atmosphere. The atmosphere holds many times less carbon than the combined stocks in the oceans, soil and plants, but changes in the atmospheric carbon level can change the Earth's climate. In the natural cycle, the carbon additions to and removals from the atmosphere are approximately the same. The burning of fossil fuels adds carbon to the atmosphere, but the amounts are relatively small compared to those involved in the natural cycle.

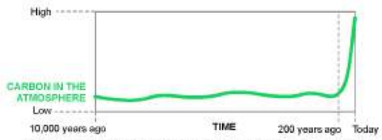


p1

Behaviour Over Time

The amount of carbon in the atmosphere has remained approximately the same for the last 10,000 years. However, the level has increased significantly over the last 200 years due to human activity, such as the burning of fossil fuels.

Carbon in the atmosphere over the last 10,000 years



The carbon level in the atmosphere has risen significantly in the last 200 years.

The burning of fossil fuels has increased over this time with a rise in material consumption, and continues to do so. At the same time, natural sources of carbon removals such as forests are being depleted. As the level of carbon in the atmosphere increases, heat becomes trapped and the surface temperature gradually increases.

Task 1: Assemble Your System

As a group, take the labels that you have and stick them to the part of the system that is represented by each label.

Once you have assembled your group's system, get your facilitator to check and give you water before you start Task 2.

While you are waiting, assign a person to be in charge of each pump, and someone to read the instructions. If you have extra members in your group, they can observe the level of water in the tubs.

Tips for assembly:

- the **terrestrial stock** (plants, soil, oceans) of carbon is many times larger than the **atmospheric stock**
- when attaching the pumps to pipes, they should each draw water from the bottom of a tub

Question: What real-world thing does the water in your system represent?

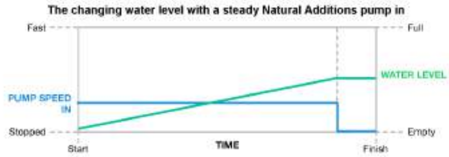
p2

Task 2: Getting Started

To get started, make sure that all of the carbon (water) is in the terrestrial stock.

1. Pump the **Natural Additions** at a steady speed to lift the **atmospheric stock** to **half full**, then stop

The changing water level with a steady Natural Additions pump in

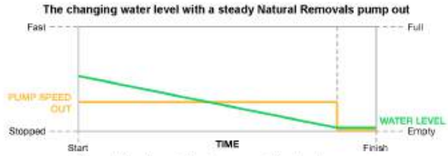


The water level gradually increases with a steady pump in.

2. Pump the **Natural Removals** to drain the **atmospheric stock**, then stop

The changing water level with a steady Natural Removals pump out

The changing water level with a steady Natural Removals pump out



The water level gradually decreases with a steady pump out.

3. Pump the **Anthropogenic Additions** at a steady speed to lift the **atmospheric stock** to **one-third full**, then stop

You can now move on to Task 3.

p3

Activity instructions from Dialogue experiment workbook: Title page - page 3

With prompts | Without prompts

Task 3: Simulate the Natural Carbon Cycle

To simulate the natural carbon cycle, the **Natural Additions** and **Natural Removals** need to pump at the same average speed.

1. In your group, simulate the natural carbon cycle with the **atmospheric stock** held at approximately **one-third full**.
2. Then discuss what kinds of real-world things are represented by the **Natural Additions** and **Natural Removals** pumps. What happens when one of the pumps is running faster or slower than the other?

The water level with equal Natural Additions and Natural Removals

The water level should stay approximately the same

Task 4: Simulate the Burning of Fossil Fuels

To simulate the anthropogenic disruption to the natural carbon cycle, the **Anthropogenic Additions** pump will need to start slow, and get gradually faster. In this task, the behaviour of the **atmospheric stock** should look similar to the graph on page 2.

Remember, the **Natural Additions** and **Natural Removals** pumps represent the natural carbon cycle, and so need to keep running at the same average speed.

1. In your group, simulate the anthropogenic disruption of natural carbon cycle.
2. Stop when the **atmospheric stock** is almost full.
3. Then, open the **Scenario Cards** and use them to guide your discussion of how your group could stop the atmospheric stock from overflowing.
4. See if your selected strategy works by playing with the pumps.
5. If you have time, see what happens in other strategies.

p4 with Scenario Card treatment

Task 3: Simulate the Natural Carbon Cycle

To simulate the natural carbon cycle, the **Natural Additions** and **Natural Removals** need to pump at the same average speed.

1. In your group, simulate the natural carbon cycle with the **atmospheric stock** held at approximately **one-third full**.
2. Then discuss what kinds of real-world things are represented by the **Natural Additions** and **Natural Removals** pumps. What happens when one of the pumps is running faster or slower than the other?

The water level with equal Natural Additions and Natural Removals

The water level should stay approximately the same

Task 4: Simulate the Burning of Fossil Fuels

To simulate the anthropogenic disruption to the natural carbon cycle, the **Anthropogenic Additions** pump will need to start slow, and get gradually faster. In this task, the behaviour of the **atmospheric stock** should look similar to the graph on page 2.

Remember, the **Natural Additions** and **Natural Removals** pumps represent the natural carbon cycle, and so need to keep running at the same average speed.

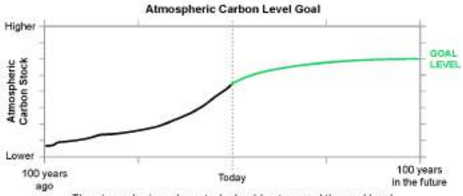
1. In your group, simulate the anthropogenic disruption of natural carbon cycle.
2. Stop when the **atmospheric stock** is almost full.
3. Then, discuss with your group how to stop the atmospheric stock from overflowing.
4. See if your strategy works by playing with the pumps.
5. If you have time, see what happens in other strategies.

p4 without Scenario Card treatment

Task 5: What now?

In Task 4, your group was asked to stop the atmospheric stock from overflowing. This level was your 'goal' level.

Atmospheric Carbon Level Goal



The atmospheric carbon stock should not exceed the goal level.

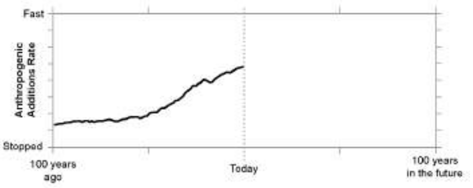
Describe what needs to happen to the **Anthropogenic Additions** pump to ensure that the goal level is not exceeded (assume that the **Natural Additions** and **Natural Removals** remain approximately the same):

____

____

Sketch what you think the future Anthropogenic Additions pump rate should be to reach the goal level:

Future Anthropogenic Additions Rate



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

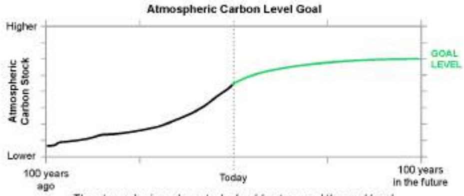
p5

with Group Diagram treatment

Task 5: What now?

In Task 4, your group was asked to stop the atmospheric stock from overflowing. This level was your 'goal' level.

Atmospheric Carbon Level Goal



The atmospheric carbon stock should not exceed the goal level.

In your group, discuss what needs to happen to the **Anthropogenic Additions** pump to ensure that the goal level is not exceeded (assume that the **Natural Additions** and **Natural Removals** remain approximately the same).

p5

without Group Diagram treatment

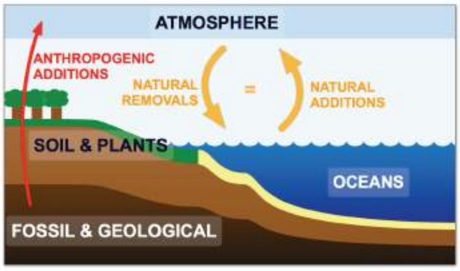
Activity instructions from Dialogue experiment workbook p4-5

A.3 Simulation experiment workbook

Introduction to the Carbon Cycle

In the Earth system, carbon moves naturally between major stocks, such as the oceans, soil, plants and the atmosphere. The atmosphere holds many times less carbon than the combined stocks in the oceans, soil and plants, but changes in the atmospheric carbon level can change the Earth's climate.

In the natural cycle, the carbon additions to and removals from the atmosphere occur at approximately the same rate. The burning of fossil fuels adds carbon to the atmosphere, but the rates are relatively low compared to those involved in the natural cycle.



A simplified model of carbon flows between major carbon stocks

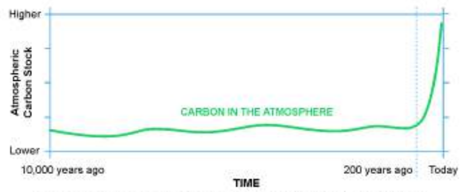
- 1 -

p1

Behaviour Over Time

The amount of carbon in the atmosphere has remained approximately constant for the last 10,000 years. However, the carbon stocks have increased significantly over the last 200 years due to human activities, such as the burning of fossil fuels.

Atmospheric carbon stock over the last 10,000 years



The carbon stock in the atmosphere has risen significantly in the last 200 years.

The rate at which fossil fuels are burnt has increased over the last 200 years with a rise in material consumption, and continues to do so. At the same time, the rates of natural removal of carbon from the atmosphere, such as those provided by forests, are being reduced. As the amount of carbon in the atmosphere increases, heat becomes trapped and the surface temperature gradually increases.

- 2 -

p2

Physical simulation treatment

Activity Overview

In this activity you will build a model of the natural carbon cycle, and then look at the role that burning fossil fuels has on the level of carbon in the atmosphere.

You should have:

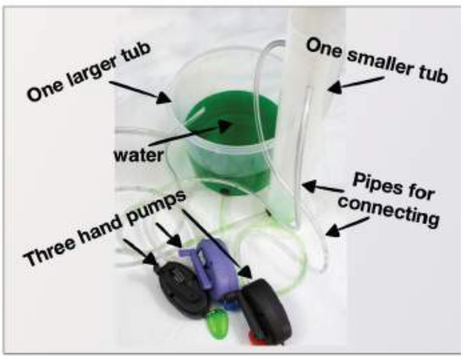
- one large tub and one small tub
- three hand pumps
- pipes to connect the pumps to the tubs
- water
- a mat in case you spill water

Some rules:

- explore the model by playing with it
- while playing, there are no 'wrong' answers
- remember to have fun!

Tips for labelling:

- the **terrestrial stock** (plants, soil, oceans) of carbon is many times larger than the **atmospheric stock**
- when attaching the pumps to pipes, they should each draw water from the bottom of a tub



- 3 -

p3 in Physical simulation treatment condition

Mental simulation treatment

Exercise Overview

In this exercise you are going to refer to a model of the natural carbon cycle, and then look at the role that burning fossil fuels has on the level of carbon in the atmosphere.

Consider a model that was made up of:

- one large tub and one small tub
- three hand pumps
- pipes to connect the pumps to the tubs
- water

Some rules:

- explore the problems by using the model
- while working as a group, there are no 'wrong' answers
- remember to have fun!

Task 1: Label Your System

As a group, write in the box a label for each part of the system. The labels you can choose from are:

- the atmospheric carbon stock
- the terrestrial carbon stock
- natural additions (to the atmosphere)
- natural removals (from the atmosphere)
- anthropogenic additions (to the atmosphere)

Once you have labelled your group's system, get your facilitator to check before you move on to Task 2.

Tips for labelling:

- the **terrestrial stock** (plants, soil, oceans) of carbon is many times larger than the **atmospheric stock**
- the pipes should each draw water from the bottom of a tub

- 3 -

p3 in Mental simulation treatment condition

Activity instructions from Simulation experiment workbook p1-3.
Note: pages 1 and 2 are common to both treatment conditions.

Physical simulation treatment	Mental simulation treatment

Task 1: Label Your System

As a group, take the labels that you have and stick them to the part of the system that is represented by each label.

You should have labels for:

- the atmospheric carbon stock
- the terrestrial carbon stock
- natural additions (to the atmosphere)
- natural removals (from the atmosphere)
- anthropogenic additions (to the atmosphere)

Once you have assembled your group's system, get your facilitator to check before you move on to Task 2. While you are waiting, assign a person to be in charge of each pump, and someone to read the instructions. If you have extra members in your group, they can observe the level of water in the tubs.

Based on what your group knows about the carbon cycle, provide an example of what real-world things these parts of the model represent:

Water in the large tub:..

Water in the small tub: ....

The Natural Additions pump: ....

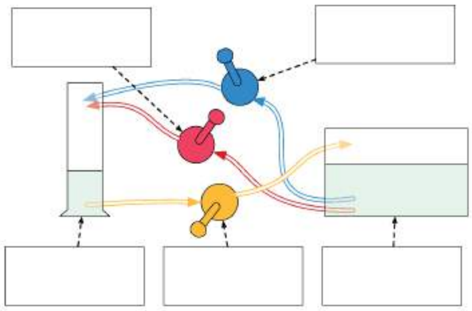
The Natural Removals pump:...

The Anthropogenic Additions pump: ....

- 4 -

p4 in Physical simulation treatment

Write the labels in the boxes for each part of the model.



Based on what your group knows about the carbon cycle, provide an example of what real-world things these parts of the model represent:

Water in the large tub:..

Water in the small tub: ....

The Natural Additions pump: ....

The Natural Removals pump:...

The Anthropogenic Additions pump: ....

- 4 -

p4 in Mental simulation treatment

Task 2: Getting Started

To get started, make sure that all of the carbon (water) is in the **terrestrial stock**. Pump the **Natural Additions** at a steady speed to lift the **atmospheric stock** to **half full**, then stop. Observe the changing **atmospheric stock** over time.

The Natural Additions pump at a steady speed, then stopped

Graph the corresponding level of carbon in the **atmospheric stock** in the graph below. Assume that the tub is empty at the start.

Describe what happens to the water level: ....

....

- 5 -

p5 in Physical simulation treatment

Task 2: Getting Started

Consider a situation where all of the carbon (water) is in the **terrestrial stock**. Now consider pumping the **Natural Additions** pump at a steady speed to lift the **atmospheric stock** to **half full,** and then stop.

The Natural Additions pump at a steady speed, then stopped

Graph the corresponding level of carbon in the **atmospheric stock** in the space below. Assume that the tub is empty at the start.

Describe what happens to the water level: ....

....

- 5 -

p5 in Mental simulation treatment

Activity instructions from Simulation experiment workbook p4-5 (cont.)

Physical simulation treatment

With the water level at half the **atmospheric stock**, pump the **Natural Removals** to drain the **atmospheric stock**, then stop. Observe the changing **atmospheric stock** over time.

The Natural Removals pump at a steady speed, then stopped

Graph the corresponding level of carbon in the **atmospheric stock** in the graph below. Assume that the tub is half full at the start.

Describe what happens to the water level: ....

....

Pump the **Anthropogenic Additions** at a steady speed to lift the **atmospheric stock** to **one-third full**, then stop. You can now move on to Task 3.

- 6 -

p6 in Physical simulation treatment

Mental simulation treatment

Consider the water level at **half** the **atmospheric stock.** Now consider pumping the **Natural Removals** pump to drain the **atmospheric stock**, and then stop. How would the **atmospheric stock** change over time?

The Natural Removals pump at a steady speed, then stopped

Graph the corresponding level of carbon in the **atmospheric stock** in the graph below. Assume that the tub is half full at the start.

Describe what happens to the water level: ....

....

- 6 -

p6 in Mental simulation treatment

Task 3: Simulate the Natural Carbon Cycle

To simulate the natural carbon cycle, the **Natural Additions** and **Natural Removals** need to pump at the same average speed.

In your group, simulate the natural carbon cycle with the level of the **atmospheric stock** held at approximately **one-third full**.

The water level with equal Natural Additions and Natural Removals

The water level should stay at approximately one-third full

What happens to the **atmospheric stock** if the **Natural Removals** pump is operating faster than the **Natural Additions** pump?

....

....

....

....

- 7 -

p7 in Physical simulation treatment

Task 3: Simulate the Natural Carbon Cycle

To simulate the natural carbon cycle, the **Natural Additions** and **Natural Removals** need to pump at the same average speed.

In your group, consider the natural carbon cycle with the level of the **atmospheric stock** held at approximately **one-third full**.

The water level with equal Natural Additions and Natural Removals

The water level should stay at approximately one-third full

What happens to the **atmospheric stock** if the **Natural Removals** pump is operating faster than the **Natural Additions** pump?

....

....

....

....

- 7 -

p7 in Mental simulation treatment

Activity instructions from Simulation experiment workbook p6-7 (cont.)

Physical simulation treatment

Task 4: Simulate the Burning of Fossil Fuels

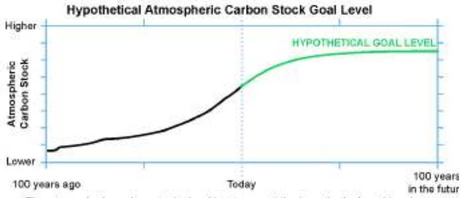
To simulate the anthropogenic disruption to the natural carbon cycle, the **Anthropogenic Additions** pump will begin slow, and gradually get faster. Remember, the **Natural Additions** and **Natural Removals** pumps represent the natural carbon cycle, and so need to keep running at the same average speed. The behaviour of the **atmospheric stock** should look similar to the graph on page 2.

1. In your group, simulate the anthropogenic disruption of natural carbon cycle and stop when the **atmospheric stock** is almost full.
2. Discuss with your group how to stop the **atmospheric stock** from overflowing and see if your strategy works by playing with the pumps. Discuss with your group how to stop the **atmospheric stock** from overflowing (your hypothetical goal level). If you have time, see what happens in other strategies.
3. Once your group has agreed on a strategy, describe what needs to happen to the **Anthropogenic Additions** pump to not exceed the hypothetical goal level of the **atmospheric stock.**

....

....

Hypothetical Atmospheric Carbon Stock Goal Level



The atmospheric carbon stock should not exceed the hypothetical goal level.

When you have finished your discussion, let your facilitator know that you are finished the activity.

- 8 -

p8 in Physical simulation treatment

Mental simulation treatment

Task 4: Simulate the Burning of Fossil Fuels

To simulate the anthropogenic disruption to the natural carbon cycle, the **Anthropogenic Additions** pump will begin slow, and gradually get faster. Remember, the **Natural Additions** and **Natural Removals** pumps represent the natural carbon cycle, and so need to keep running at the same average speed. The behaviour of the **atmospheric stock** should look similar to the graph on page 2.

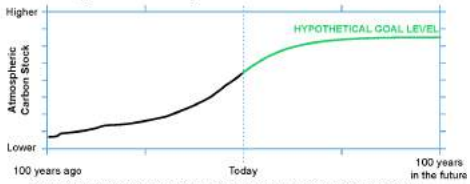
Consider that the **atmospheric stock** is now almost full. Discuss with your group how to stop the **atmospheric stock** from overflowing (your hypothetical goal level). Describe what needs to happen to the **Anthropogenic Additions** pump to not exceed the hypothetical goal level of the **atmospheric stock.**

....

....

....

Hypothetical Atmospheric Carbon Stock Goal Level



The atmospheric carbon stock should not exceed the hypothetical goal level.

When you have finished your discussion, let your facilitator know that you are finished the exercise.

- 8 -

p8 in Mental simulation treatment

Activity instructions from Simulation experiment workbook p8 (final)

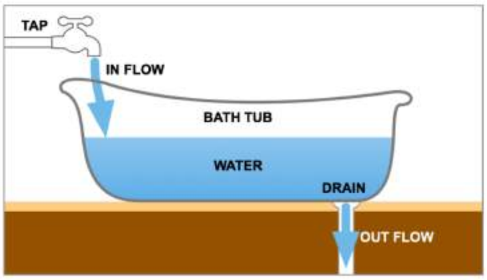
A.4 Trial Workshop 2 Instructions

Task 1: Build a Bathtub Model

Use the tubs, pumps and pipes that you have to build a model of a bathtub.

- imagine that the tap is always on
- imagine that there is no plug
- use both tubs (and make one the bath tub)

Design your model to keep the water level in the bath at a constant level.



Above: Bathtub, where the tap is always on and the plug is never in.

If the level of water was graphed over time, it would look like this:

p2

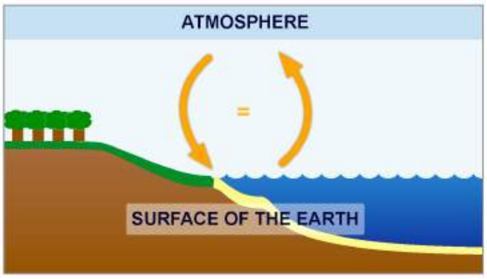
Once everyone agrees, draw your group's system in the box below.

(Use labels as much as you can and describe the speeds of the pumps.)

p3

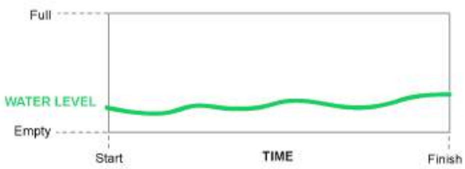
Task 2: The Natural Carbon Cycle

Use the tubs, pumps and pipes that you have to build a model of a the natural carbon cycle. Remember that all pumps you use need to always be moving.



Above: the Natural Carbon Cycle.

If the amount of carbon in the atmosphere was graphed over time, it would look like this:



p4

Once you have built your system and played with the pumps, draw your group's system in the box below.

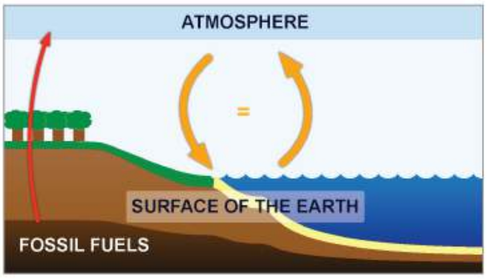
(Use labels as much as you can and describe the speeds of the pumps.)

p5

Activity instructions from Trial Workshop 2 - p2-5. Title page omitted.

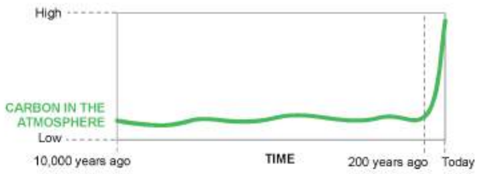
Task 3: The Impact of Burning Fossil Fuels

Use the tubs, pumps and pipes that you have to add the burning of fossil fuels to your model of the natural carbon cycle. Remember that all pumps you use need to always be moving.



Above: the Natural Carbon Cycle and the Burning of Fossil Fuels.

If the amount of carbon in the atmosphere was graphed over time, it would look like this:



p6

Once you have built your system and played with the pumps, draw your group's system in the box below.

(Use labels as much as you can and describe the speeds of the pumps.)

p7

Activity instructions from Trial Workshop 2 - p6-7. (cont.)

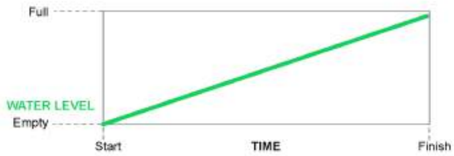
A.5 Trial Workshop 3 Instructions

Task 1: IN Pumps

1. Hook up one pump (the IN pump) so that you can move water from the tub to the cylinder.
2. Pump the water at a slow and steady rate from the tub to the cylinder until the water level reaches the top of the cylinder.

If the pump goes like this...

Then the level of the water in the cylinder should go like this...



Check this by playing with the pumps.

1

Notes/Prompts

Help the participants figure out how the pumps work by showing or instructing how to put the pipes on the pumps. Take your time so that everyone can see how they are assembled together.

Prompt discussion as required..

What is a steady state?

How long it would take to fill up the cylinder?

p1

Task 2: OUT pumps

1. Hook up an OUT pump into your system, with the cylinder still full.
2. Pump the water OUT at a slow and steady rate back into the tub.

If the pump goes like this...

Draw what you think will happen to the water level in the cylinder...

Check your answers by playing with the pumps, and talking to your group.

Q. What would happen if the OUT pump went twice as fast?
Q. What would happen if the task went for twice as long?

2

Notes/Prompts

Remember to get the group to draw on their book what should happen to the water level before doing the activity.

You could prompt discussion with the questions, and get different people to have a go. Move on once the group 'gets it'.

p2

Task 3: IN and OUT pumps

1. Start with an empty cylinder.
2. Pump the IN pump to move water into the cylinder.
3. When the water level reaches half way up the cylinder, start pumping the OUT pump at the same speed.

If the pumps go like this...

Draw what you think will happen to the water level in the cylinder...

Q. What happens when the IN and OUT pumps are the same?

3

Notes/Prompts

Take the opportunity to discuss Dynamic Equilibrium.

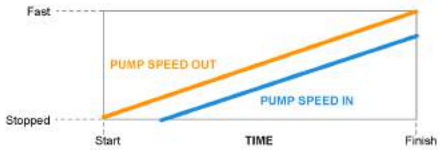
Ask the participants to pump as fast/slow as they can (as long as it's the same).

p3

Task 4: IN and OUT II

1. Start with a cylinder half full of water.
2. Pump the OUT pump at a steadily increasing speed.
3. After a short time, pump the IN pump at a steadily increasing speed, but never quite as fast as the OUT pump.

If the pumps go like this...



Draw what you think will happen to the water level in the cylinder...

Q. What would happen if the IN pump and the OUT pump were reversed?

4

Notes/Prompts

The water should go down slowly and steadily (if the difference is always the same).

This might be tricky for some groups, as the direction of the lines is opposite to the direction of the water level.

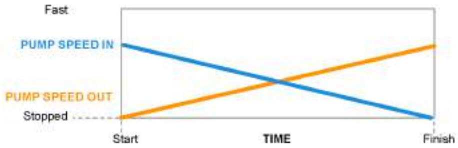
p4

Activity instructions from Trial Workshop 2 - p1-4.
Title page omitted *Note: Notes/Prompts are for facilitators only*

Task 5: Increasing & Decreasing

1. Start with an empty cylinder.
2. The OUT pump starts very slow, and gradually gets faster and faster.
   The IN pump starts very fast, and gradually gets slower and slower.

If the pumps go like this...



Draw what you think will happen to the water level in the cylinder...

Q. What would happen if the IN pump and the OUT pump were reversed?

5

Notes/Prompts

This will definitely start to test the visual calculus. The correct water-level graph looks like a semi-circle. Point out some interesting landmarks, such as:

"When is it increasing the most?"
"When is it decreasing the most?"
"When will it stay level (or in dynamic equilibrium)?"

p5

Task 5: Increasing & Decreasing II - Beach Case Study

Imagine the Tubs and Pumps system represents the number of people on a beach.

Bondi Beach, Sydney, Australia

Identify how the different parts in the model match to the different things on the beach.

Tubs & Pumps parts	People on the Beach
Water level in the cylinder	
Water level in the tub	
Water entering the cylinder	
Water leaving the cylinder	

6

Notes/Prompts

Water level in the cylinder = People on the beach
Water level in the tub = Rest of the population (etc)
Water entering the cylinder = People arriving at the beach
Water leaving the cylinder = People leaving the beach

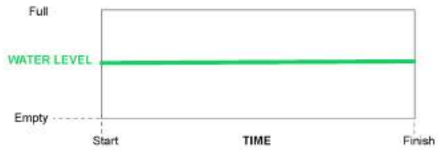
Discuss if there are any other 'things' missing from the system, eg. people arriving by bus/car/walking, which tub are the people in the water in?, etc.

p6

Task 6: Pump Speeds

Now think about what the pumps are doing to change the water level. Start with the cylinder half full of water, and both pumps should be going.

If the water level stays like this...



Draw what you think needs to happen to the IN and OUT pump speeds...

Q. Is there more than one answer?
Q. What needs to happen for the water level to remain the same?

7

Notes/Prompts

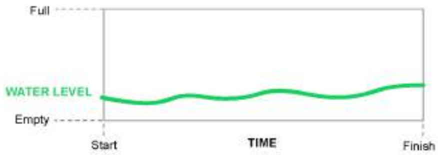
Should be easy if the group understood dynamic equilibrium.
Answer is correct as long as in-flow = out-flow

p7

Task 7: Delays

Start with the cylinder a quarter full of water, and both pumps should be moving at all times.

If the water level changes like this...



Q. Write what the IN and OUT pumps do to make the water levels go up and down like the graph?

8

Notes/Prompts

The small oscillations are caused by the two pumps working at slightly different and changing rates.

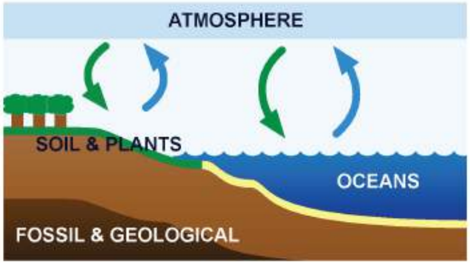
p8

Activity instructions from Trial Workshop 2 - p5-8. (cont.)

Task 8: Natural Carbon Cycle Case Study I

Imagine that the Tubs and Pumps represent carbon in the atmosphere.

Over the last 10,000 years, the level of carbon in the atmosphere has stayed relatively stable in natural cycles.



Simplified Natural Carbon Cycle

Q. What sorts of things remove carbon from the atmosphere?

Q. What sorts of things add carbon to the atmosphere?

Q. What does carbon in the atmosphere do?

9

Notes/Prompts

Questions should prompt an interesting discussion.

A. Removals could be: planting trees (afforestation), plants, oceans

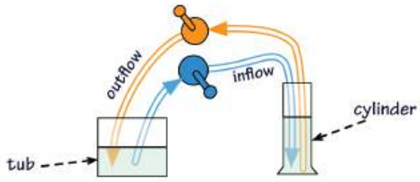
A. Additions could be: cutting down trees (deforestation), plants, animals, oceans

A. Greenhouse effect (regulating temperature on Earth). Possibly discuss that more carbon increases this effect and raises temperatures.

p9

Task 8: Natural Carbon Cycle Case Study II

Using the system that you have already been using, which might look something like this...



..map parts of the natural cycle of carbon in the atmosphere to the system.

Carbon in the atmosphere (natural cycle)	Tubs & Pumps parts
Carbon in the atmosphere	
Natural removals	
Natural additions	
Carbon stored in ocean/geology/land/soil	

10

Notes/Prompts

Carbon in the atmosphere = big tub

Natural removals = OUT pump

Natural additions = IN pump

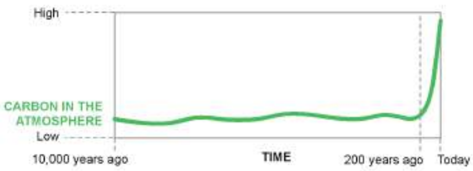
Carbon stored in ocean/geology/land/soil = cylinder

p10

Task 9: Fossil Fuel GHG Emissions

Now think about the role of humans burning Fossil Fuels over the last 200 or so years. Remember that the natural cycle should stay about the same,
like in Task 8.

Change your Tubs & Pumps system to try and match the historical trend for the last 200 years. You can add parts to the system if you like.



Draw your new system below:

11

Notes/Prompts

Try not to prompt at all.

If necessary, point participants to the other equipment available on their table.

p11

Task 10 - The Future

Q. If your system keeps going as it is, what do you think is going to happen? (You can play with the pumps if you like.)

Q. What could you do with system to stabilise the water level?

Q. What could you do with the system to decrease the water level?

12

Notes/Prompts

Have a small discussion, and then capture the discussion by getting a participant to write down their solution.

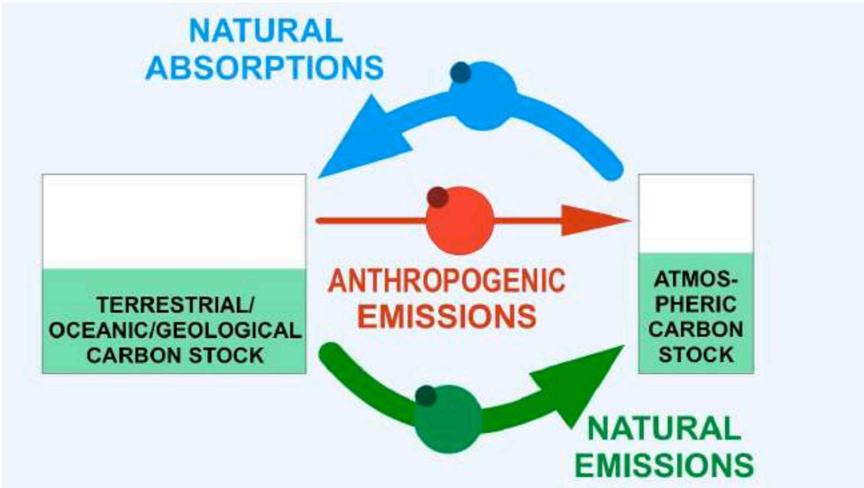
p12

Activity instructions from Trial Workshop 3 p9-12. (cont.)

Appendix B Assessment tasks used in the workshops

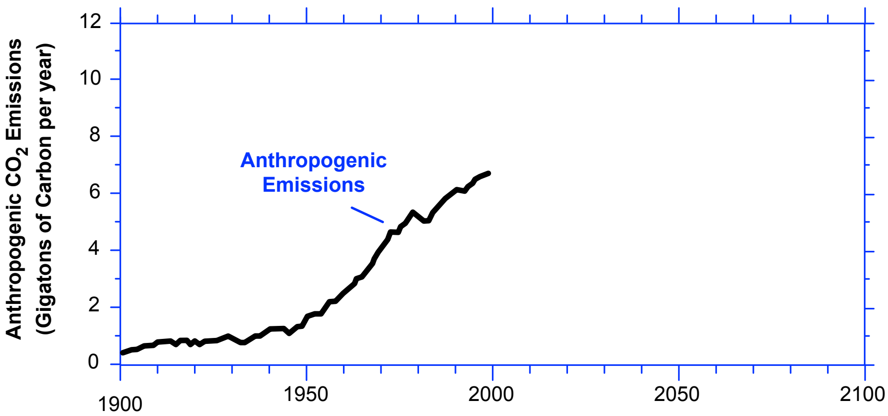
B.1 Emissions & Removals (ER) assessment sheet

Questionnaire



Q1: In the year 2000, the atmospheric carbon concentration was approximately 370 parts per million (ppm). What needs to happen to the emissions to ensure that the atmospheric GHG concentration stabilise at — and does not exceed — 400ppm by the year 2100?

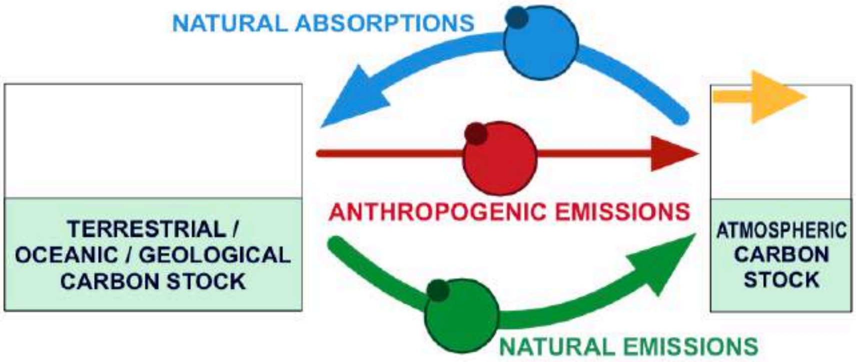
Q2: Sketch what you think the trajectory for future GHG emissions needs to be for the situation in Q1 to occur.



Graph: Anthropogenic CO_2 Emissions (Gigatons of Carbon per year)

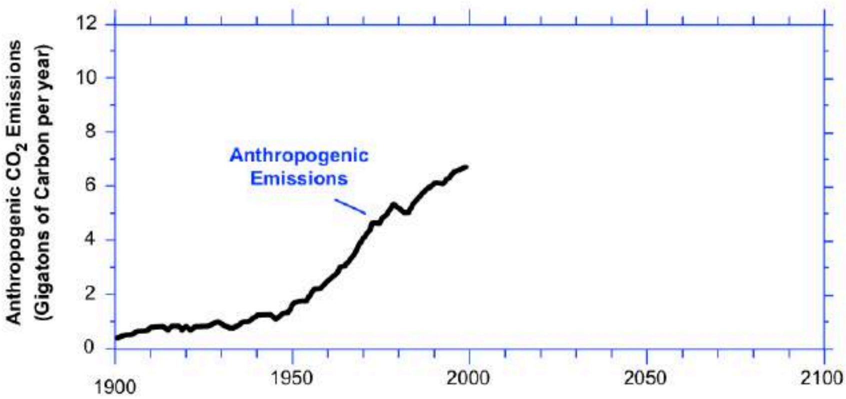
B.2 Anthropogenic Emissions (AE) assessment sheet

Questionnaire



Q1: In the year 2000, the atmospheric carbon concentration was approximately 370 parts per million (ppm). Consider a situation where the Natural Emissions and Natural Absorptions are in dynamic equilibrium. What needs to happen to the Anthropogenic Emissions to ensure that the atmospheric GHG concentration stabilises at — and does not exceed — 400ppm by the year 2100?

Q2: Sketch what you think the trajectory for future Anthropogenic Emissions needs to be for the situation in Q1 to occur.



Graph: Anthropogenic CO_2 Emissions (Gigatons of Carbon per year)

Note: the 400ppm goal is described in the activity.

B.3 Anthropogenic Additions (AA) assessment sheet

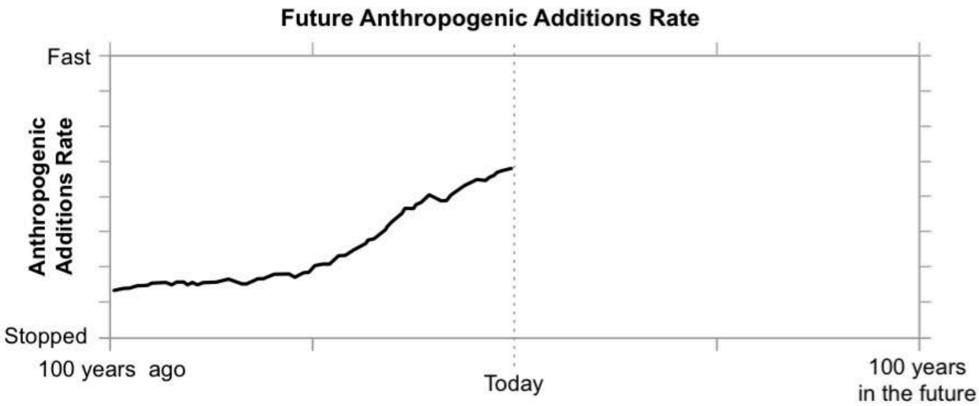
Individual Questionnaire

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 atmospheric goal level is not exceeded:

Q3. Sketch what you think the future trajectory of the **Anthropogenic Additions** pump should be to not exceed the goal level.



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	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

Assessment sheet used in the Workbook experiments

Note: the goal level referred to here is carried over from the activity

B.4 Profile questionnaire

The answers from these questions will be used to categorise and report results.

Please complete the following questions:

1. Indicate your age category (in years):
   - ☐ under 16
   - ☐ 16-17
   - ☐ 18-21
   - ☐ 22-25
   - ☐ 26-30
   - ☐ 30 or over
2. Gender
   - ☐ male
   - ☐ female
3. Field(s) of study ______________________________
   the area of study you most identify yourself with.
   eg. engineering, history, physics. ______________________________
4. Degree progress
   - ☐ first year
   - ☐ second year
   - ☐ later years
   - ☐ postgraduate
   - ☐ other (please specify) ______________________________
5. Language you are most comfortable communicating in
   - ☐ English
   - ☐ other (please specify)______________________________
6. Role in your group
   - ☐ Anthropogenic Additions
   - ☐ Natural Additions
   - ☐ Natural Removals
   - ☐ Observer
   - ☐ other (please specify) ______________________________
7. Have you done this workshop before, or did you previously know the outcome?
   - ☐ No
   - ☐ Yes (please specify) ______________________________

The profile questionnaire used across workshops *Note: in the Mental simulation treatment condition, Q6 was omitted. The age categories under 16 and 16-17 was added for Trial Workshop 2, which included high school students.*

Appendix C Video transcripts

C.1 Checklist experiment video transcript

Consider the problem of greenhouse gas emissions. In 2001 the Intergovernmental Panel on Climate Change (IPCC), a scientific panel organised by the United Nations, concluded that the Earth's climate system is unequivocally warming. They attribute this to human activities, which cause warming through processes that increase concentrations of greenhouse gases in the atmosphere, such as deforestation and burning fossil fuels. In recent years, many countries have been looking for a solution to reduce greenhouse gas emissions, such as through taxes or emissions trading schemes. We are going to do a quick activity to see what the likely effect of such interventions could be. The climate system is an extremely complex system, with many interrelated parts and components.

One way to consider a complex problem is to reduce it to its key drivers. In the carbon cycle, the four largest 'stocks' of carbon are the Oceans, Terrestrial and Geological structures, and the Atmosphere. Carbon is a part of all living things, and passes through the carbon cycle in varying ways and at varying rates. The ocean is the largest carbon stock, with much of the carbon being stored in cooler, deeper regions. Fossil fuels and geological structures are the next largest store, followed by terrestrial stores, such as soils and plants. The Atmosphere is the smallest store, containing a much smaller amount than the Terrestrial, Geological and Oceanic stores combined. Remember that we are building a simple model, so there are other stocks and relationships not shown here, but these are considered the key stocks. The units used here are Gigatons of Carbon.

The next aspect of the carbon cycle is how much carbon moves between these stocks. The natural cycles can be considered as fairly consistent, with about the same amount of carbon being absorbed as being emitted. Though, these processes are increasingly becoming out of sync, as, for example, deforestation makes it more difficult for terrestrial sources to absorb carbon, leading to more carbon in the atmosphere, resulting in warmer surface temperatures. The major change to the cycle, however, is the burning of fossil fuels. It is interesting to note how small this flow is compared to the much larger numbers exchanged naturally, but that this emission is of major concern. The units for the flows used here are Gigatons of Carbon per year. Again, remember that we are building a simple model, so smaller flows, such as volcanic activity accounting for 0.2GtC/year, aren't considered here.

So how does this help us understand what we can do about greenhouse gas emissions? We can use this simple model to build a simple physical analogue of this system to help us understand the problem space. Let's consider the Terrestrial, Oceanic, and Geological carbon stock as water in a container. The more water that is in the tub, the more carbon is held in the stock. The atmosphere is represented by a smaller and separate container, with a much smaller amount of water in it. The containers are connected by tubes that allow the water to transfer between them. These represent the natural emissions, the natural absorptions and the anthropogenic — or human-related — emissions.

The rate at which the water flows through the tubes is determined by a pump, which we can control. If we only pump the natural emissions, the atmospheric carbon level will rise. Likewise, if we only pump the natural absorptions, the atmospheric carbon level will fall. And if we only pump the anthropogenic emissions, the atmospheric carbon level will rise. Note that it is moving at a very slow rate compared to the natural processes. If we pump the natural emissions and the natural absorptions at similar rates — even if those rates change — the amount of carbon in the atmosphere will remain stable. This is called 'dynamic' equilibrium. This scenario is similar to what was happening on Earth before the industrial revolution.

Let's start pumping the anthropogenic emissions to see what happens. Very slowly at first, but eventually a rapid rise. Remember, although it was a relatively small number, it flows each year and over time that can lead to a big increase in the atmospheric carbon stock level - or the atmospheric carbon concentration. Looking at this graph, it becomes plain to see that something needs to change in this system if we are to reach a stable atmospheric carbon concentration of, for example, 400ppm. One of the suggested changes is for countries to reduce their emissions.

Let's have a look at what the anthropogenic emissions pump was doing in the example before. We are using a simple scale here that one revolution of the pump occurs every time one gigaton of carbon is released each decade. The industrial revolution kickstarted an increase in anthropogenic carbon emissions, and since the Second World War emissions have continued to rise at an increased rate. One policy response is to reduce emissions back to 1990 levels, or rates.

The question then becomes, will reducing emissions back to 1990 rates, help to level out the greenhouse gas concentrations? In groups of three, you will explore the dynamics of this simple system. One person will take on the role of Natural Absorptions, one will take on the role of Natural Emissions, and one will take on the role of Anthropogenic Emissions. You might also have a fourth person to help instruct and observe.

Good luck!

[END]

C.2 Workbook experiment video transcript

Note: the Simulation workshop had two videos - one for the physical simulation and one for the mental simulation experiment. The audio track for both videos is identical.

In this activity, you are going to build a model of the natural carbon cycle to look at the role that burning fossil fuels has on the level of carbon in the atmosphere.

Your model includes: one small tub, one larger tub, pipes for connecting, three hand pumps, and water. There are some rules for the activity: explore the activity by using the model, while working as a group, there are no 'wrong' answers, and remember to have fun!

In the Earth system, carbon moves naturally between major stocks, such as the oceans, soil, plants and the atmosphere. The atmosphere holds many times less carbon than the combined stocks in the oceans, soil and plants, but changes in the atmospheric carbon level can change the Earth's climate. In the natural cycle, the carbon additions and removals from the atmosphere occur at approximately the same. The burning of fossil fuels adds carbon to the atmosphere, but the amounts are relatively small compared to those involved in the natural cycle.

The amount of carbon in the atmosphere has remained approximately the same for the last 10,000 years. However, the level has increased significantly over the last 200 years due to human activity, such as the burning of fossil fuels.

You're now ready to work through the activity. Good luck!

[END]

Appendix D Reflections on trial workshops

In this section, I briefly discuss the evolution of the workshop design. This evolution occurred as a reflection of student behaviour in the workshops described in this thesis, and more importantly the workshops that were used as trial workshops. The complete list of workshop timings and participant cohorts are described in Table D.1, extending Table 4.13.

Table D.1: Summary of all workshop timings and courses

Workshop/Time		Class/group (College)
Trial 1	Sem 1, 2013	VCUG3001 Unravelling Complexity (All colleges)
Checklist	Sem 1, 2013	ENVS1001 Environment and Society (Science) COMP3530 Systems Engineering for Software Engineers (Comp.S) ENGN2225 Systems Engineering Design (Engineering)
Experimentation	Sem 1, 2013	Last two groups: ENGN1211 Discovering Engineering (Engn)
Trial 2	June 2013	High School Outreach (Engineering)
Trial 3	June–November 2013	International System Dynamics Conference (Workshop) ENVS8003 Climate Change Vulnerability and Adaptation (Science) Primary School Outreach (Science Week) NYSF - National Youth Science Forum (national HS program)
Linear Pretest	Sem 1, 2014	ENGN1211 Discovering Engineering (Engineering)
Simulation	Sem 1, 2014	ENGN1211 Discovering Engineering (Engineering)
Dialogue	Sem 1, 2014	ENVS1001 Environment and Society (Science) VCUG3001 Unravelling Complexity (All colleges)
Linear Cover Stories	Sem 1, 2014	ENGN2226 Systems Engineering Analysis (Engineering)

Note: I was the convener of ENGN2225 in 2013, and VCUG3001 and ENGN2226 in 2014.

In this section of the appendix, I describe the evolution of the assessment tasks and activities in relation to five observations:

1. Assessment confusion
   students simply being confused by what the test was asking them to do

2. Group discussion
   experimentation with structured discussion prior to assessment

3. System assembly
   experimentation with participants constructing the system from scratch

4. Towards a learning activity
   experimentation with focus on learning rather than assessment

5. Linear dynamics
   experimentation with pretests and linear exercises with different cover stories

D.1 Test confusion

Trial workshop 1 was the first real activity outside of the informal activities that I had run within my research group. It was also the first group that did not have any training in systems thinking. The assessment task for the activity is shown in Figure D.1. It shows a scenario where the carbon concentration drops and stabilises at approximately 340ppm, and asks participants to graph and write the required trajectory for this to occur. This is the carbon stabilisation (340ppm) exercise from Sterman & Booth Sweeney (2002; 2007).

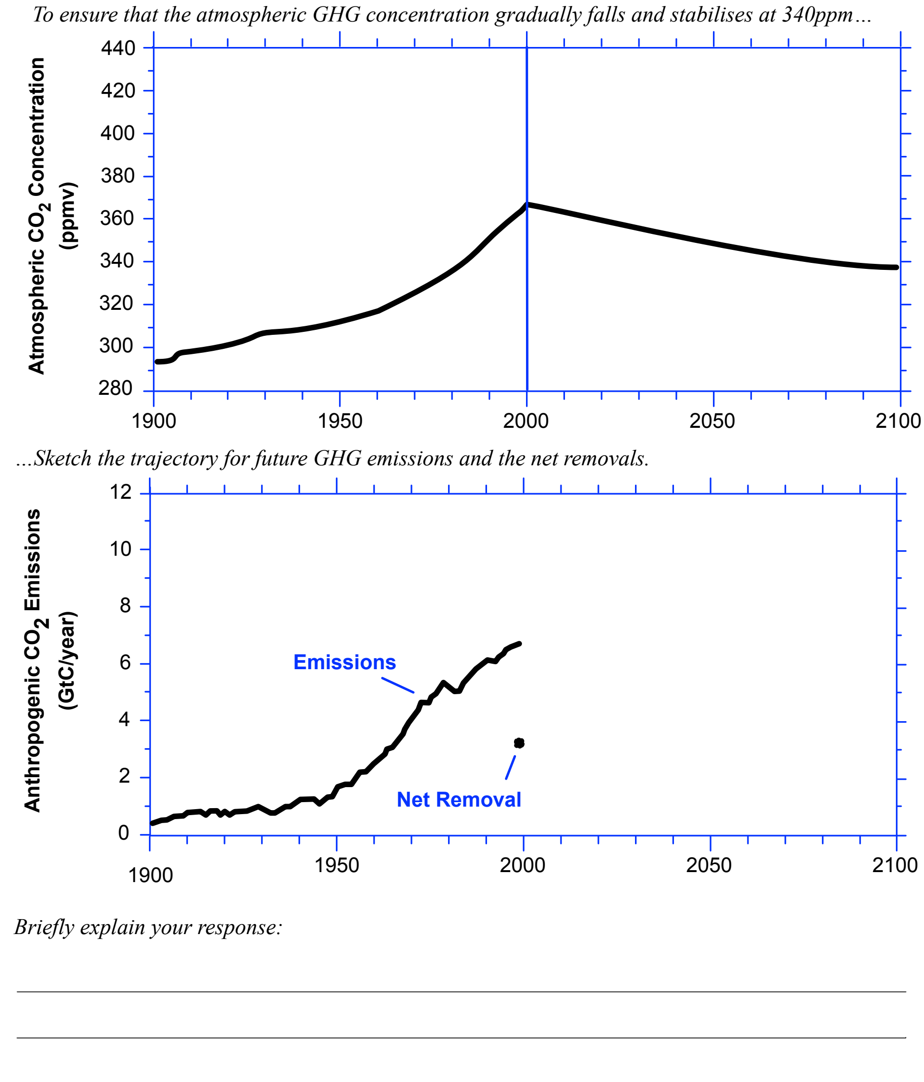


Figure D.1: Example of the test used in Trial Workshop 1. The assessment task was based on the activity in Sterman & Booth Sweeney (2002; 2007). A correct response requires the future net removal trajectory and the future net removal to immediately intersect.

In Trial workshop 1, it became apparent that the graphical task was causing too much confusion to demonstrate any meaningful learning outcomes. 28 participants across two sessions went through the activity and completed the assessment tasks. Many participants provided responses that could be categorised as correct, but interpretation of their written description was difficult and inconsistent with their graph. Table D.2 shows the types of problems that were encountered in the interpretation of results, ranging from furiously crossed out results, to results that did not use the 'net removals' dot[20].

20. Whether or not participants would intuitively use the net removals dot was an investigation in

Table D.2: Problems observed in the Trial workshop 1 responses

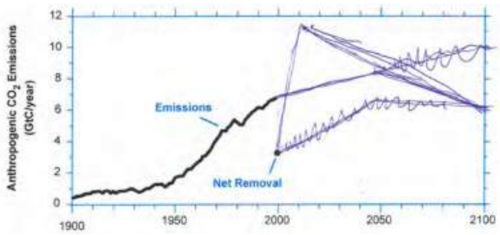
Frequency (n=28)	Description
8 (28.5%)	multiple graphical responses, crossed out or re-written
6 (21.5%)	no removals shown, or removals that did not consider the net removals dot
2 (7.1%)	no written description to justify the graphical response
15 (53.6%)	responses where the net removals did not intersect the emissions

Many of these problems appeared to be caused by some internal inconsistencies in the graphical task, as well as the relationship between the assessment task and the T&P model. The *Anthropogenic CO2 Emissions* graph used in Figure D.1 appeared to confuse participants, as although the y-axis is measured the GtC/year, the y-axis label is *Anthropogenic CO2 Emissions*, but the line label is just *Emissions.* During the activity, participants asked whether the line label referred to the *anthropogenic* emissions or *all* emissions. The data source would suggest that it reflects the anthropogenic emissions, as the natural emissions would push it to approximately 9GtC/year. The "net removals" dot, indicates the algebraic sum of the natural and anthropogenic graphs. The distinction in the activity is trivial, as the correct response is any graph where the removals exceed the emissions enough to lower the concentration to 340ppm. However, this confusion appeared to be tripping up participants in their responses.

Further to the apparent inconsistency in the graph, the net removal dot shows the current rate of removal from all sinks, but in the T&P system the only source of removal was the natural removal pump. This led participants to remark that the natural removals would need to increase independently from the natural additions. The 'correct' response in Box D.1 confusingly demonstrates this logic, where the removals suddenly exceed a rising emissions rate. This also demonstrates the 'crossing-out' of solutions, described in Table D.2.

Sterman and Booth Sweeney (2007 p.218)

Box D.1: A correct response demonstrating an unintended solution. Whilst this solution would see the atmospheric carbon stock fall, the net removals would have to increase at an unprecedented (and unimaginable) rate.



Response #01009

Basically, you would expect emissions to increase in the short run and then decrease after about 2040 as renewable technologies and social behaviour adjusts.

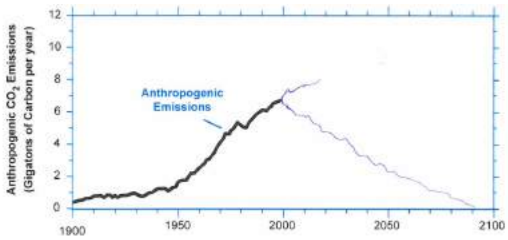
To help make the learning outcome of the activity clearer, the Net Removals dot was removed and the test adjusted in later assessment tasks.

D.2 Group discussions and pretests

In the Checklist experiment, there still appeared to be confusion with the assessment task. The confusion appeared to be around the graphical and written consistency of responses. Having put a considerable number of participants through these workshops (477 participants), I tried different approaches in the last two sessions of the workshops used in the Checklist experiment, listed as *Experimentation* in Table D.1.

First, I wanted to investigate the influence of group discussions on the responses. This was done by discussing the possible solutions to the test before the assessment sheet was completed. This strategy appeared to improve graphical results. With 22 responses, 19 responses demonstrated graphs that would be classified as ‘correct’. Many written statements also appeared to demonstrate a coherent mental model about the problem, as demonstrated in Box D.2.

Box D.2: A correct response with consistent causal logic



Response #30151

The rate of anthropogenic emissions will need to decrease, whilst the rates of natural absorptions and emissions should continue at a constant rate that ensures 'dynamic equilibrium'

Observing that a large proportion of participants that could arrive at a much clearer response with a relatively small prompt was an interesting observation. There were two main concerns that I had with this approach. The first was the repeatability. In the discussion, I had left it open to participants to direct the outcome of the class, and dominant participants may have had an undesired effect on the rest of the class. If this was the case, it did not necessarily help improve individual participants' understanding about stocks and flows. The second concern was that these dominant participants could have directed the group towards an incorrect answer, meaning that the activity may have indirectly confused participants.

I also wanted to investigate whether responses changed throughout the activity. I had initially made the decision to use Sterman & Booth Sweeney's observation that understanding of stocks and flows was poor as the starting point for the study. Together with observations that a pre-test/post-test methodology could have other hidden effects, such as priming students, meant that no 'pre-test' was used in the T&P activity. The effect of using a pre-test was, however, an unknown consideration in the workshop design. As a trial, I gave the same assessment task to participants before (test 1) and after (test 2) the activity, as well as after the discussion of solutions (test 3).

There were 26 students in this trial group. Over the three tests, the frequency of 'correct' and 'decrease' graphical responses increased, and the frequency of 'stabilise' and 'increase' graphical responses reduced. These results are shown in Table D.3, along with the relative changes of responses. Thirteen (50%) of the participants did not change their response between test 1 and test 3. 9 participants (35%) improved their response between test 1 and test 2, and a further 2 participants (8%) improved their response between test 2 and test 3.

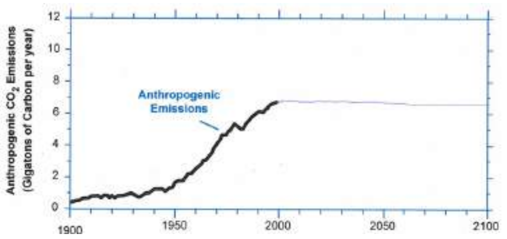
Table D.3: Frequency and changes of responses in the Experimentation workshop

Graphical Response	Test 1	Test 2	Test 3
Correct	6 (23%)	11 (42%)	12 (46%)
Decrease	6 (23%)	6 (23%)	7 (27%)
Stabilise	7 (27%)	4 (15%)	3 (12%)
Increase	7 (27%)	5 (19%)	4 (15%)
Change in response	**Test 1 – Test 2**	**Test 2 – Test 3**	**Test 1 – Test 3**
More correct	14 (54%)	23 (88%)	13 (50%)
Remain the same	9 (35%)	2 (8%)	10 (38%)
Less correct	3 (12%)	1 (4%)	3 (12%)

Note: column percentages may not total 100% due to rounding; n=26

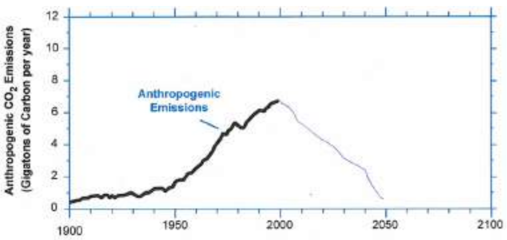
These results led to two noteworthy suggestions for the design of the T&P activity. First, that the activity had an apparent effect of improving some participants' graphical responses (test 1: 23% to test 2: 42%). An example is provided in Box D.3, where the participant clearly improves both his or her understanding of the problem and his or her ability to describe a possible solution.

Box D.3: A participant's improving graphical and written descriptions over the three tests



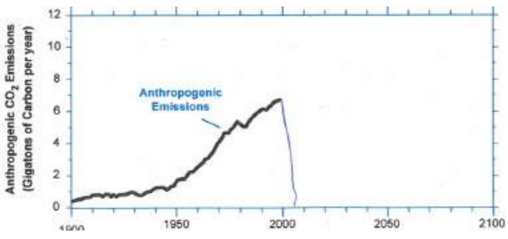
Written Description #40151a - Test 1
(Stabilise graph)

We need to wind back these emissions to approximately year 2000 levels. Worldwide, I expect we've gone way beyond them already.



Written Description #40151b - Test 2
(Correct graph)

Never mind winding it back to year 2000 levels. Provided there is any at all, the atmospheric level will inevitably rise. The only question is, "how fast will it rise."

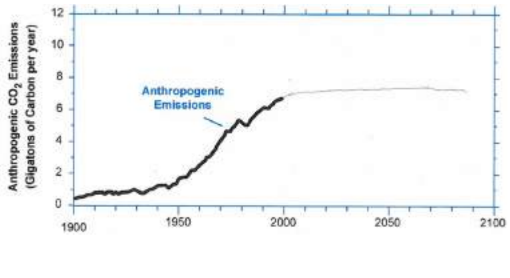


Written Description #40151c - Test 3
(Correct graph)

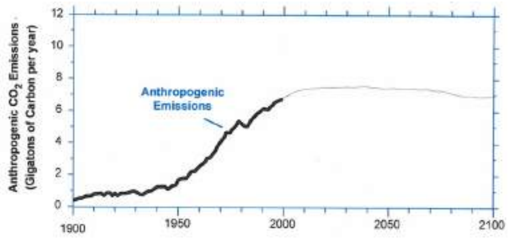
Anthropogenic. emissions are the only net contributor to rising levels (natural components are in balance). The anthropogenic component must be decreased to zero asap.

The second observation is that the activity had the potential to confirm incomplete mental models, seen in the group of participants who did not change their response between test 1 and test 3 (50%). A participant's sequence of responses are shown in Box 8.4, where the participant's views are entrenched through the activity. This behaviour is similar to phenomena of confirmation bias, where the learner looks for cues to confirm his or her mental models rather than challenging their mental models.

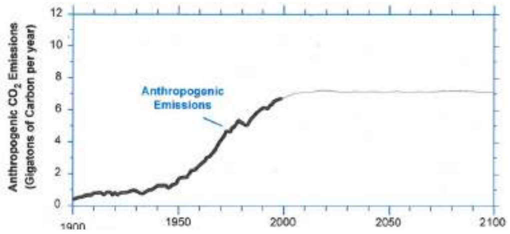
Box D.4: A participant's entrenched understanding



Written Description #40121a - Test 1
(Stabilise graph)
Humanity must reduce it's CO2 emission rate without affecting possible population growth. Modern technology must become cleaner. Also, the cultivation of CO2 absorbing forests could help combat human emission.



Written Description #40121b - Test 2
(Stabilise graph)
Our experiment determined that the natural absorption rate needs to be increased to accommodate for the combined emissions. Reducing anthropogenic emissions would lessen to amount that absorption needs to be increased by.



Written Description #40121c - Test 3
(Stabilise graph)
The natural absorption rate must increase as much as possible to accommodate for anthropogenic emissions. At the same time we must reduce the anthropogenic emissions so that it fits in with the increased absorption rate.

The observations and insights arising out of this trial workshop highlighted the importance of well-placed prompts and informed discussion amongst group members. These results also led to the complete change of approach in further workshops, and the use of *Scenario Cards* in the Dialogue experiment.

D.3 Assembling the system

The main observation from the analysis of the Checklist experiment was that ‘incorrect’ responses were often accompanied by mismatching written and graphical responses (see §5.3.3). In the Checklist experiment, the T&P system was given to the participants completely assembled with labelled components. I had thought that this was a favourable approach for orientating the participants, but the mismatch in the written and graphical responses prompted a reflection on this decision.

I designed Trial Workshop 2 to embrace Papert’s constructionist approach. The activity asked participants to assemble the T&P system from scratch, assisted by a set of workbook prompts. Instructions were brief and the prompts allowed for open-ended construction. The first of the tasks is shown in Figure D.2.

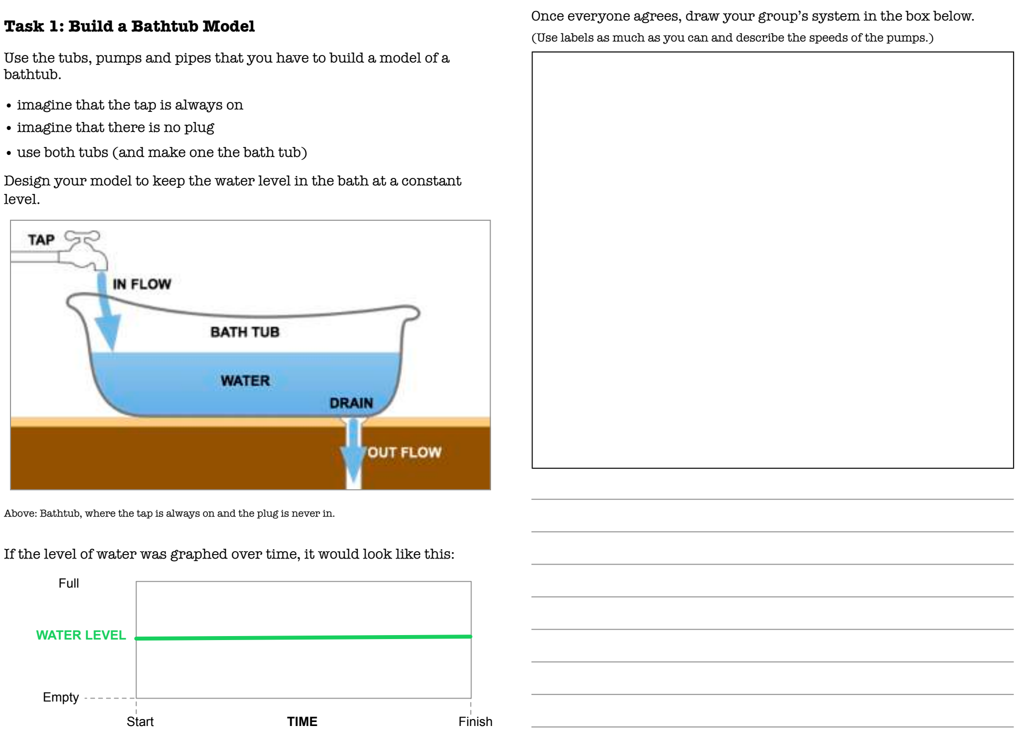


Figure D.2: Instructions for Task 1 from Trial Workshop 2, asking participants to use the T&P components to create a bathtub ‘model’

The activity—in short—was a disaster. The participants were confused about the task and many groups configured the models in novel and unintended ways. Many systems did not represent the system they were trying to describe in any accurate way. Few participants reported that they learnt anything at all in the activity. It was clear that the task of assembling the T&P system required much more scaffolding than could be provided in a 1-hour session. Although the students completing the activity were of high school age, it wasn't at all clear that more senior students would perform any better.

Adding to this difficulty was an observation that I had made previously, but was exacerbated by the general confusion in this activity. It concerned the terminology used to describe the carbon cycle, but was further highlighted with the general confusion in this activity. The term 'anthropogenic' often confused participants, but most of them understood what it meant. However, coupled with the terms 'emissions' and 'absorptions', the conversations within groups were often confused. In later workshops, 'emissions' was replaced with 'additions' and 'absorptions' was replaced with 'removals'. This also had a clearer connection to the container and bathtub metaphors described in Chapter 3.

This workshop helped me to understand the value of presenting a constructed model that participants could begin to manipulate without prior knowledge, and extend their learning from this point. I had thought that assembling the model would be a good starting point, and by assembling the system the participants would better understand the system. However, through this activity it became clear that labelling an already constructed model (that is, explicitly identifying the parts by sticking labels on the parts) would achieve a similar goal without the possibility of total confusion.

This led to a modification in the Workbook experiment that involved placing labels on the system components prior to starting the pumping activities, and could be described as *directed* play rather than *free* play.

D.4 Using a learning activity

Trial Workshop 3 represented a useful shift in the way that I approached the design of the T&P activity in the Workbook workshops, and this section takes a small detour as it describes a series of events that changed my own thinking about the activity.

The concept of having a test to evaluate the participant's understanding had sat uncomfortably with me, as it is widely known that doing well on a test does not mean that the knowledge is retained. Richmond's (2010) generic structure of a the dissipation process comes to mind, as shown in Figure 6.1. The stock of memorised facts—such as those learnt to pass an exam—diminishes over time through a process of forgetting. This decay is related to the size of the stock - the more facts learnt, the faster that facts will be forgotten.

I wanted to highlight the importance of the T&P metaphor as a tool for concrete learning, not as another learning exercise that will be forgotten. I wanted participants to leave the exercise seeing the world around them in terms of tubs, pumps and interactions, the same way that systems thinkers see the world in terms of stocks, flows and feedback. I removed the assessment task from the activity, and set about building a better learning activity.

In the Trial Workshop 2, I tried to reduce the frequency of mismatched written and graphical responses by prompting participants to construct the system. In the Trial Workshop 3, I tried to achieve the same goal by giving participants experience of distinguishing stocks and flows. Many of the mismatched responses suggested that the participants were "pattern matching" – using the "correlation heuristic". If participants could see graphical examples of stock-and-flow behaviour over time before drawing their own graphs, then perhaps the frequency of mismatched responses could be reduced.

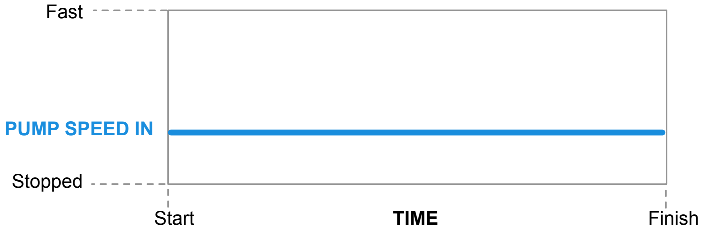
In the checklist experiment, only an A4 prompting sheet was given (see material in §A.1). In Trial Workshop 3, tasks from the prompting sheet were used with graphical representations laid out in a similar way to the assessment task. The two tasks shown in Box D.3 are conceptually the same tasks as those given in the first two tasks of the A4 prompting sheet used in the Checklist activity. However, the graphical representation prompt participants to consider the what the behaviour over time graph would look like.

a)

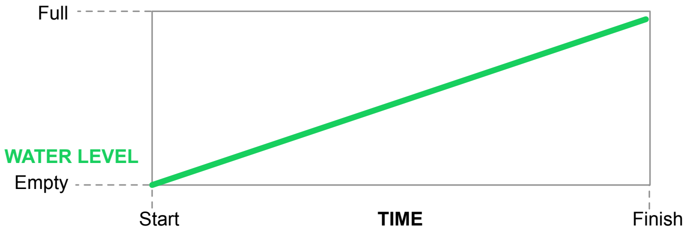
Task 1: IN Pumps

1. Hook up one pump (the IN pump) so that you can move water from the tub to the cylinder.
2. Pump the water at a slow and steady rate from the tub to the cylinder until the water level reaches the top of the cylinder.

If the pump goes like this...



Then the level of the water in the cylinder should go like this...



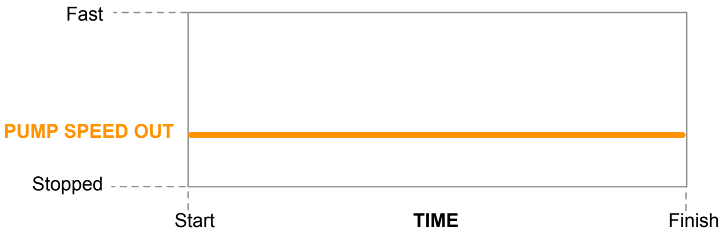
Check this by playing with the pumps.

b)

Task 2: OUT pumps

1. Hook up an OUT pump into your system, with the cylinder still full.
2. Pump the water OUT at a slow and steady rate back into the tub.

If the pump goes like this...



Draw what you think will happen to the water level in the cylinder...

Check your answers by playing with the pumps, and discuss.

Q. What would happen if the OUT pump went twice as fast?
Q. What would happen if the task went for twice as long?

Figure D.3: Example of the first two tasks in the Trial Workshop 3 workbook. a) Task 1 checks the behaviour of one additions pump; b) Task 2 gets participants to predict the behaviour of a removals pump

In all, there were nine tasks that progressively went from a one-pump system shown in Figure D.3 to a three-pump system described in the T&P system of the carbon cycle. Participants demonstrated an understanding of how the T&P metaphor (the conceptual source domain) worked, before applying it to carbon cycle (the conceptual target domain). Participants were thinking through the problems in terms of the T&P system. As Draper (2010) describes, the systems concepts were apparently transparent.

The discussions that followed this trial workshop were also the richest that I had observed. The first I recount in §6.4. The second took place after I presented the T&P activity at the 2013 *International System Dynamics Conference* in Boston. The small group of participants had a deep and engaging discussion about the universality of the activity as a learning tool for systems concepts, particularly in primary schools. Computer simulation has rightly dominated the field of system dynamics. Although the computer allows for powerful simulation, it can be seen as a 'black box', where commands are processed without students necessarily understanding the underlying mechanism. The T&P activity was seen as a useful experiential introduction to systems by ST educators.

It was at this conference that I started a conversation, with Anne LeVigne and Tracy Benson from the Waters Foundation and Lees Stunz from the Creative Learning Exchange, on developing the T&P activity into an broader educational activity. This conversation continues, and they have prototyped and modified their own versions for use in primary schools.

These discussions prompted me to remember an observation that I had made before the formal phase of my dissertation experiments began, and I was still developing the T&P activity. Barry Newell ran an informal exercise (which I would later develop into the T&P activity) as a tutorial demonstration in *ENGN3410 Engineering Sustainable Systems* in semester 2, 2012.

It involved the pumps, a bucket and a beer glass. We were using the activity to complement the weekly reading on the systems basics (chapter 1 of Meadows, 2008) to help students identify simple system behaviour. In a relatively free-form demonstration of ‘bathtub’ dynamics, we had asked the students to operate the pumps as fast as possible. Of course, the water level didn’t change significantly even though the pump operators were furiously pumping.

Many weeks later, we were discussing the system archetypes in a tutorial. We were talking about the ‘escalation’ archetype, and a student reminded the class about the activity we had run with the pumps, remarking that the competing increased effort in the balancing loops were like the pumps getting faster and faster, and the only way to stop this was to ‘let go’ (stop pumping). Similar to the primary school student, this student had used the pumps as a problem solving tool, a tool for concrete learning, and as a lens through which to see the world.

These observations led to further changes in the activities. The workbook appeared to provide a useful prompt for participants. The number of tasks were reduced from nine to five, removing some duplication. However, I also wanted to explore some additional aspects of the experiment: first, the effect of asking the group to agree to a solution before undertaking the individual test. This became the GD treatment in the Dialogue experiment. Second, the effect of prompting participants towards the solution by presenting them with a number of scenarios showing a graph with a matching written description. This became the SC treatment in the Dialogue experiment. Finally, I wanted to investigate the effect of playing with the pumps or just viewing a pictorial representation of the pump system. This became the Mental simulation treatment in the Simulation workshop.

To make any valid comparison to the earlier workshops, I decided to keep the assessment task from the Checklist workshops in the experiment. My experiences did, however, prompt a number of subtle changes. First, in line with the group diagram treatment, I wanted to investigate whether the individual thought that the group had affected their answer, so a 5-point Likert scale from ‘not at all’ to ‘completely’ was used to ask to what extent did you agree with your group. Second, I wanted to test the participant’s causal logic separate from the graphical response and written description as another way of interpreting their understanding of the exercise. These are shown in §4.3.3. Third, I the y-axis numerical scale from the graph was replaced with a relative scale, as many students in the Checklist experiment had used the numbers to calculate the solution rather than the conceptual source domain. This change is shown in §4.3.2.

The insights that I gained through these trial workshops do not contribute directly to the results presented in Chapter 5, but do contextualise how the workshops developed through challenging my own mental models about the problem central to this thesis.

D.5 Linear dynamics

A linear dynamics task was used on two occasions towards the end of all workshops: once as a pretest, and once to investigate the role of cover stories on the assessment task. A pretest was intentionally not part of the learning activity. However, after observing the mismatch between written and graphical responses in the earlier workshops, and the great improvement in results observed in trial workshops, I decided to conduct a pretest in the Simulation experiment.

I also didn't want to prime participants around the outcome of the T&P activity in the pre-test, so a simpler linear dynamics or 'bathtub' test was used (see Box 2.1 as an example). The intention of including this test was to investigate whether participants had the intuition to complete the visual integration task when it involved linear dynamics, before testing the more complicated non-linear visual integration task presented in the carbon cycle task. The graphical component of the test is shown in Figure D.4.

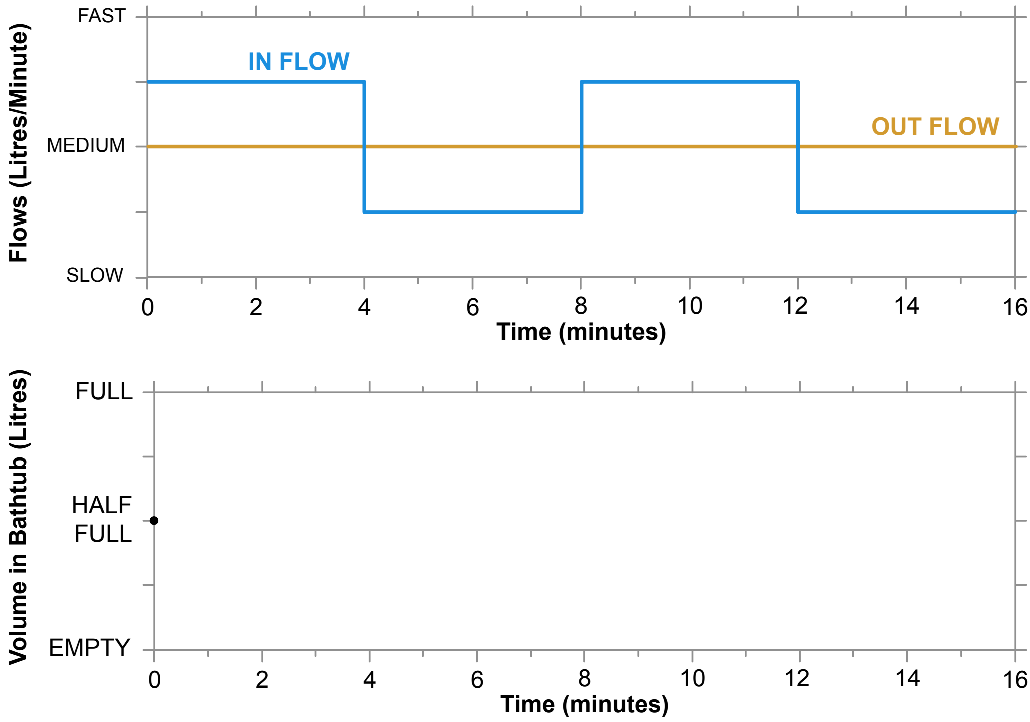


Figure D.4: The linear dynamics task modelled on Booth Sweeney & Sterman (2000), and as described in Box 2.1.

I had some concerns about using the bathtub as a basis for testing because of inconsistencies in the mapping between the metaphor's conceptual source domain to the physical experience of having a bath (see §2.2 for this argument). After interpreting the results of Simulation workshop, where the physical simulation treatment condition had only marginally and not significantly outperformed the mental simulation treatment condition, I had wanted to investigate further the different contexts of the linear dynamics activity, similar to the idea of "cover stories" presented in Cronin et al (2007; 2009) which had shown no significant difference.

As an extrapolation on the bathtub dynamics test, I used a cohort of students (ENGN2226 in 2014) who had likely already completed the T&P activity in ENGN1211 in 2013 to test two further contexts: an abstract version, where the context of the problem was an abstract exercise of stocks and flows, and; a pumps version, where the context of the problem used the T&P pictorial analogue. The same graph was used in the three versions, with the terms 'additions' and 'removals' replacing 'in flow' and 'out flow' in the abstract test. The instructional text varied slightly to fit the context, and is shown in Box D.5.

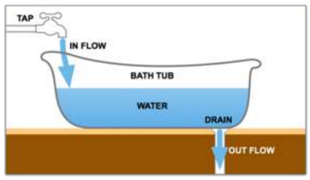
Box D.5: Question text for the three linear treatment conditions. The text was written as similarly as possible to investigate whether the different contexts built on conceptual metaphors—the bathtub and the pumps contexts—performed differently to the abstract context.

Abstract treatment

Consider two flows over time. One flow adds to an accumulation, and the other flow subtracts from that accumulation. The level of the accumulation at any time could be measured at any time by considering the total additions minus the total subtractions from time zero.
The top graph shows the relative rates of the addition and subtraction flows over a period of time. From that information, draw the shape of the behaviour for the accumulation over time on the second graph below. Assume the initial accumulation (at time zero) is half.

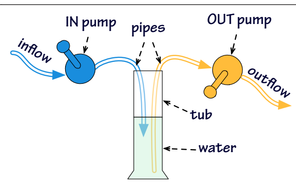
Bathtub treatment

Consider the bathtub shown on the right. Water flows in at a certain rate, and exits through the drain at another rate.
The top graph shows the hypothetical behaviour of the in flow and out flow rates for the bathtub over a period of time. From that information, draw the shape of the behaviour for the volume of water in the tub over time on the second graph below. Assume the initial volume of water in the tub (at time zero) is half.



Pumps treatment

Consider the pumps shown on the right. Water is pumped in at a certain rate, and exits through a pump at another rate.
The top graph shows the hypothetical speeds of the in flow and out flow pump rates for the tub over a period of time. From that information, draw the shape of the behaviour for the volume of water in the tub over time on the second graph below. Assume the initial volume of water in the tub (at time zero) is half.



The investigation behind this experiment was to see whether the two instances of the container metaphor—the bathtub and the pumps—helped participants to mentally simulate the solution based on their *a priori* understanding of these systems.

Coding of linear dynamics task

For the three linear dynamics treatment conditions—abstract, bathtub, pumps—a similar question was given with identical graph with the language contextualised for each test. In flows and out flows were given over time, and participants were asked to graph the accumulation. Not enough detail was given to allow mathematical calculation, and the question asked for participants to describe the nature of the behaviour over time. This section describes the process for preparing data shown in the results.

The graphical responses for the linear dynamics test were categorised by shape and relative amplitude and period using a rubric. A correct response needed to be linear and periodic, with local maxima at t=4,12, with a local minima as half-full at t=0,8,16. Incorrect responses had one or more of the criteria not met. The typical categories are shown in Table D.4. Responses that did not fit into these response categories were recorded as *other.*

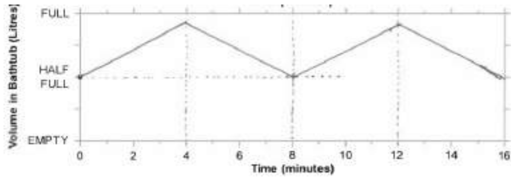
Table D.4: Rubric for categorising graphical responses in the linear dynamics task

Criterion	Category	Correct	Partially incorrect		Pattern match	Periodic	
			points	shape		linear	non-linear
Shape	linear	×	×			×	
	non-linear			×			×
	square-wave				×		
Local maxima	t=4,12	×	×	×			
	other t					×	×
Local minima	half at t=0,8,16	×		×		×	×
	other t or amplitude		×				
	Category code	**correct**	**CSIP**	**CPIS**	**PM**	**PL**	**PNL**

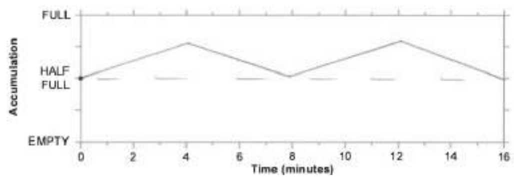
As there was no numerical data given in the test, correct responses could have a local maxima at any value higher than half-full and would still be categorised as *correct*. Partially incorrect responses had either an incorrect local maxima or minima or an incorrect shape (typically non-linear). Pattern-matching responses were typically a square wave, and were coded as such at any amplitude. Periodic linear and non-linear had local minima and maxima at different t. Examples of typical graphical responses are shown Box D.6.

Box D.6: Examples of graphical responses in the linear dynamics task

Correct Responses

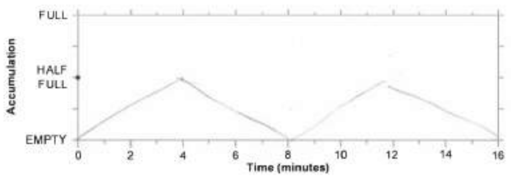


Response #700111

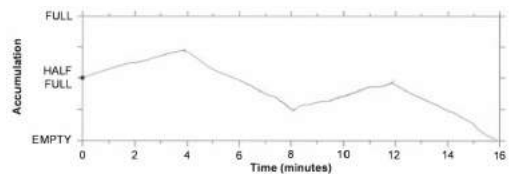


Response #700520

Correct shape, incorrect points (CSIP) responses

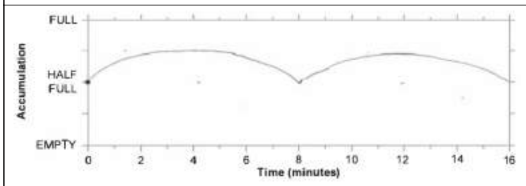


Response #700112

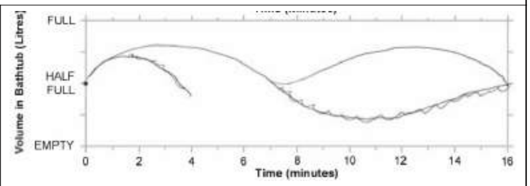


Response #700207

Correct points, incorrect points (CPIS) responses

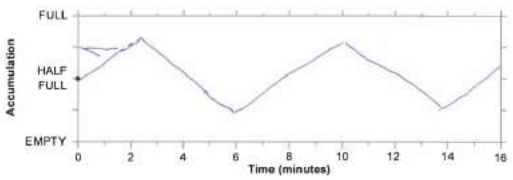


Response #700824

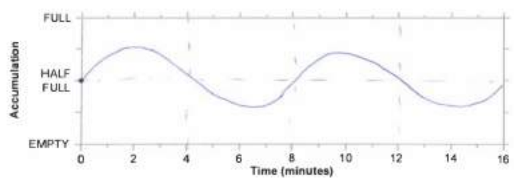


Response #700910

Periodic Linear (PL) and Periodic Non-Linear (PNL) responses

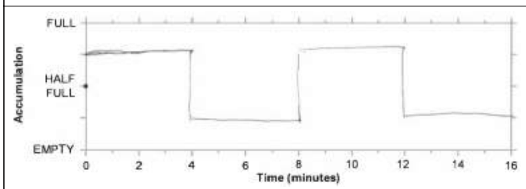


PL Response #700314

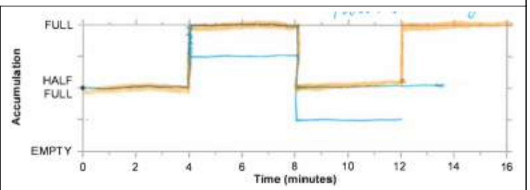


PNL Response #700621

Pattern Match (PM) responses

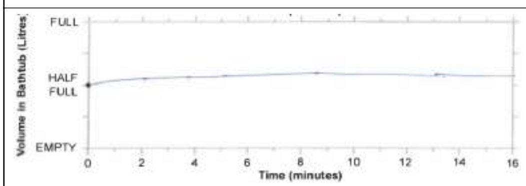


Response #700519

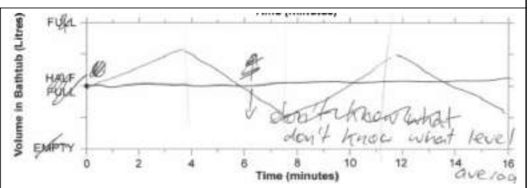


Response #700103 (inverted PM)

Other responses



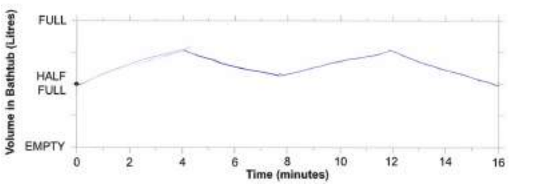
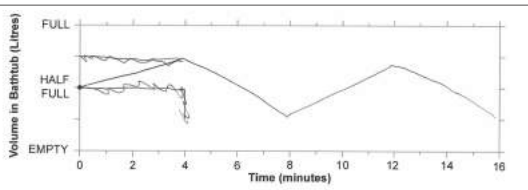
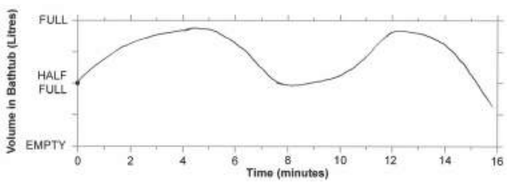
Response #700728



Response #700211

The written descriptions were not as easily categorised as the carbon cycle written descriptions, largely due to the periodic nature of the solution. For example, Box D.7 shows three responses which have similar written descriptions, but demonstrate different graphical representations. *Respons*e #600113 changes direction at the correct points (t=4, 8 and 12), but the bathtub is one-quarter full at t=16. Response #600407 shows a non-linear shape with local minima and maxima not aligned with the correct times. Although the written descriptions supported the graph, in the majority of responses the description could have matched multiple graph types. For this reason, the written descriptions were not coded.

Box D.7: Three different linear dynamics graphs with similar descriptions

	Response #600705 Correct graph The water is fluctuating between half-full and more than half-full as the flow changes.
	Response #600113 Correct shape, Incorrect Points graph It is oscillating from above half full to a bit below half full.
	Response #600407 Periodic, non-linear graph The volume is increasing, then decreasing, then increasing and then decreasing again.

Pretest experiment

The linear pretest was used in the Simulation experiment to investigate whether pre-existing knowledge or understanding was skewing results. The same pretest (bathtub) was used for all participants. The task was relatively easier than the carbon cycle task, as the test involved linear flow rates.

175 participants completed the pretest. The profile data for the pretest experiment is referenced in Table D.6. Participants were randomly assigned to the carbon cycle treatment groups, not based on the pretest results. Approximately half of participants drew a correct pretest graph (53.1%). Slightly more

participants in the physical simulation treatment condition (54.9%, n=82) drew a correct pretest than the mental simulation treatment condition (51.6%, n=93). The result is not significant ($p = 0.76$).

However, the pretest was an indicator of performance for the carbon cycle graphical response in both treatment conditions, shown in Table D.5. Participants with a correct linear pretest graph were four times more likely to draw a correct anthropogenic additions graph ($p < 0.01$, OR: 4.04 95%CI [1.99,8.51]), indicating that prior knowledge is a key predictor for whether participants can draw the correct carbon cycle graph.

Table D.5: Linear pretest and carbon cycle graph n=175

Linear pretest graph			Carbon cycle graph		
Correct	93	(53%)	Correct	48	(52%)
			Incorrect	45	(48%)
Not correct	82	(47%)	Correct	17	(21%)
			Incorrect	65	(79%)
Σ	175			*p < 0.01*	

Linear cover stories workshop

The Linear cover stories workshop was used to investigate how context influences understanding. Three treatment conditions—Abstract, Bathtub, and Pumps—were used with similar written and graphical tests. Profile data of these participants were collected using the same process as for the carbon cycle activity.

Participants in this workshop were similar in profile across all treatment conditions within the cover story workshop, with no profile category showing a significant bias. Participants were largely STEM students, under 21, male, and later in their degree program, as shown in Table D.6.

Table D.6: Summary of profile data across all workshops (category percentages)

Profile category		Cover stories workshop						Bathtub Pretest	
		Abstract		Bathtub		Pumps			
Age	≤21	55	(74.3%)	57	(73.1%)	54	(74.0%)	162	(92.6%)
	≥22	18	(24.3%)	21	(26.9%)	18	(24.7%)	12	(6.9%)
	NA	1	(1.4%)	0	(0.0%)	1	(1.4%)	1	(0.6%)
Sex	female	17	(23.0%)	18	(23.1%)	18	(24.7%)	36	(20.6%)
	male	56	(75.7%)	60	(76.9%)	55	(75.3%)	137	(78.3%)
	NA	1	(1.4%)	0	(0.0%)	0	(0.0%)	2	(1.1%)
Field of study	STEM	71	(95.9%)	76	(97.4%)	73	(100.0%)	171	(97.7%)
	other	1	(1.4%)	1	(1.3%)	0	(0.0%)	3	(1.7%)
	NA	2	(2.7%)	1	(1.3%)	0	(0.0%)	1	(0.6%)
Degree progress	early	28	(37.8%)	26	(33.3%)	31	(42.5%)	168	(96.0%)
	later	45	(60.8%)	52	(66.7%)	42	(57.5%)	6	(3.4%)
	NA	1	(1.4%)	0	(0.0%)	0	(0.0%)	1	(0.6%)
Language	English	51	(68.9%)	45	(57.7%)	46	(63.0%)	143	(81.7%)
	other	22	(29.7%)	33	(42.3%)	27	(37.0%)	31	(17.7%)
	NA	1	(1.4%)	0	(0.0%)	0	(0.0%)	1	(0.6%)
	Σ	74		78		73		175	

Note: percentages may not total 100% due to rounding.

The graphical responses were categorised as shown in Table D.4. Across the treatment conditions, the Bathtub (62.8%) and Pumps (60.3%) treatment conditions had proportionally more correct responses than the Abstract treatment condition (43.2%), shown in Table D.7. This result is not is not significant ($p = 0.27$).

Table D.7: Linear dynamics results by graphical response (column percentages)

Graphical response	Abstract		Bathtub		Pumps	
Correct	32	(43.2%)	49	(62.8%)	44	(60.3%)
CPIS	5	(6.8%)	5	(6.4%)	5	(6.8%)
CSIP	17	(23.0%)	6	(7.7%)	7	(9.6%)
PM	11	(14.9%)	5	(6.4%)	4	(5.5%)
Flat	0	(0%)	6	(7.7%)	1	(1.4%)
PL/PNL/Other	9	(12.2%)	7	(9%)	12	(16.4%)
Σ	74		78		73	

However, this trend suggests that using a conceptual source domain for problem solving, such as the bathtub or pumps, is more effective than problem solving of abstract concepts. Comparing the Abstract treatment to the others, participants were twice as likely to draw a correct graph ($p = 0.01$ OR: 2.10; 95%CI [1.15,3.86]).

Investigating types of error can help investigate how the problem contexts are used to understand the problem. Participants in the abstract treatment condition were more likely to draw a Correct Shape Incorrect Points (CSIP) graph and Pattern Matching (PM) than any other treatment condition. The CSIP graph shows a misunderstanding around the level of accumulation over time, and the PM graph shows a confusion between stocks and flows.

The bathtub treatment condition had fairly evenly distributed errors (between 6.4% and 9%) across the other categories. A key difference between the bathtub and pump treatment conditions errors is that the flat category is more prevalent than the abstract and pumps treatment conditions. The flat graphical response suggests a conceptual misunderstanding, likely associated with the bathtub conceptual source domain. This could be due to the large surface area of a bathtub, and that changes over time are slower to see than in the pumps context. This is a topic for further investigation. The pumps treatment condition had more periodic non-linear (PNL) responses (7.9%) than any other treatment condition, also likely with the conceptual source domain, as the pumps could be seen dynamically changing over time.

9 Bibliography

ABC RN Breakfast, 2015, radio broadcast, ABC Radio National, Sydney, 19 November. Interview with Richard Denniss and Fran Kelly.

Afternoon Journal, c1980, television series, TV PUC São Paulo and KTV Solucoes, Brazil. Conversation with Seymour Papert and Paulo Friere.

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.

Argyris, C., & Schön, D., 1974, *Theory in practice: Increasing professional effectiveness*, Jossey Bass, San Francisco.

Aronson, E. 1978, *The jigsaw classroom.* Sage, Oxford, England.

Australian Bureau of Statistics, 2014, Review of the sex standard, viewed 12 January 2015, http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Review+of+the+Sex+Standard

Australian National University (ANU), 2013, *Undergraduate Handbook*, The Australian National University, Canberra. p.130

Baker, R.M., 1996, *"Supertutes", "Yes Minister" and Action Research: methods to assist geography teaching.* Paper delivered at the International Geography Congress, The Hague

Bammer, G., 2013, *Disciplining interdisciplinarity: integration and implementation sciences for researching complex real-world problems,* ANU ePress, Canberra.

Bammer, G. & Smithson, M., 2009, *Uncertainty and risk: multidisciplinary perspectives,* Earthscan, London.

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, IPCC.

Barton, B.F., and Marthalee S.B., 1983, 'The Nature and Treatment of Professional Engineering Problems—The Technical Writing Teacher's Responsibility.' *Journal of technical writing and communication* 13.4 : 355-368.

Benson, 2010, 'Changing School Culture: Creating Student-Centered Classrooms', in J Richmond, L Stuntz, K Richmond & J Egner (eds), *Tracing connections: voices of systems thinkers,* isee systems and Creative Learning Exchange, Acton, MA.

Biggs, J.B. & Collis, K.F., 1982, *Evaluating the quality of learning: the SOLO taxonomy,* Academic Press, New York.

Bligh, D.A., 2000, *What's the use of lectures?* Jossey-Bass Publishers, San Francisco.

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R., 1956, *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*, David McKay Company, New York.

Boden, T.A., G. Marland, and R.J. Andres. 2013. *Global, Regional, and National Fossil-Fuel CO2 Emissions.* Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN.

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

Boulton, G., & Lucas, C., 2008, *What are universities for?* League of European Research Universities, viewed 4 December 2013, http://www.leru.org/?cGFnZT00

Box, G.E.P. & Draper, N.R., 1987, *Empirical Model-Building and Response Surfaces*, Wiley, NY.

Braun, W., 2002, The System Archetypes, *The Systems Modeling Workbook* p. 27: Published online at: http://wwwu.uni-klu.ac.at/gossimit/pap/sd/wb_sysarch.pdf.

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

Browne, C., & Rajan, T., 2015, 'Using Distributed Constructionism in Engineering Tutorials: Requirements Engineering' *Proceedings of the 26th Annual Conference for the Australasian Association for Engineering Education.* Torquay.

Buskes, G., Shen, B. & Shallcross, D., 2010, 'Promoting student engagement in lectures through a trial use of response clickers', *Proceedings of the 21st Annual Conference for the Australasian Association for Engineering Education.* p.290.

Chapin, H., 1978, 'Flowers are Red', on *Living Room Suite,* LP, Elektra, New York

Cox, S., 2014, 'Using analogues to explore deformation of the Earth's crust', presentation at *CMBE & CPMS Teaching and Learning Colloquium,* ANU, 4 November 2014

Cronin, M.A, & 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.

Crutzen, P.J., 2002, Geology of mankind, *Nature*, 415(6867):23.

De Bono, E., 2000, *Six thinking hats,* Penguin, London.

Deming, W.E., 2000, *Out of the crisis,* MIT Press, Cambridge, MA.

Dewey, J., 1938, *Experience and education,* Touchstone, New York.

Draper, F. 2010, 'Teaching by wondering around: learning about the world naturally', in J Richmond, L Stuntz, K Richmond & J Egner (eds), *Tracing connections: voices of systems thinkers,* isee systems and Creative Learning Exchange, Acton, MA.

Dyball, R. & Newell, B., 2014, *Understanding human ecology: a systems approach to sustainability,* Routledge, Oxon, UK.

Fiddaman, T., 1997, Feedback Complexity in Integrated Climate-Economy Models. PhD thesis, MIT Sloan School of Management, Cambridge, MA (unpublished).

Ford, A.F., 1999, *Modeling the environment: an introduction to system dynamics models of environmental systems,* Island Press, Washington, D.C.

Forrester, J.W., 1969, *Urban dynamics,* Pegasus Communications, Waltham, MA.

Forrester, J.W., 1970, *Engineering for the benefit of mankind: a symposium held at the third autumn meeting of the National Academy of Engineering,* National Academy of Engineering, Washington D.C.

Forrester, J.W., 1989, 'The Beginning of System Dynamics' Banquet Talk at the international meeting of the System Dynamics Society, Stuttgart, Germany, July 13, viewed 18 August 2013, http://web.mit.edu/sysdyn/sd-intro/D-4165-1.pdf

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

Friere, P., 2005, *Pedagogy of the Oppressed*, 30th Anniversary Edition, Continuum, New York.

Hattie, J., 2013, *Visible learning: A synthesis of over 800 meta-analyses relating to achievement,* Routledge, Oxon, UK.

Hovmand, P., 2014, *Community Based System Dynamics,* Springer, New York.

IPCC, 2005, *IPCC Special Report on Carbon Dioxide Capture and Storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change* [Metz, B., O. Davidson, H. C. de Coninck, M. Loos, and L. A. Meyer (eds.)]. Cambridge University Press, Cambridge, UK

Kapmeier, F., 2004, 'Findings from four years of bathtub dynamics at higher management education institutions in Stuttgart'*, Proceedings of the 2013 International System Dynamics Conference,* Cambridge, MA, System Dynamics Society, Albany, NY.

Kapmeier, F., with Tilebein, M., & Happach, R.M., 2015, 'Bathtub Dynamics revisited: Exploring the engineering domain' *Proceedings of the 2015 International System Dynamics Conference*, Cambridge, MA, System Dynamics Society, Albany, NY.

Keeling, C.D., Piper, S.C., Bacastow, R.B., Wahlen, M., Whorf, T.P., Heimann, M. & Meijer, H.A., 2005, 'Atmospheric CO2 and 13CO2 exchange with the terrestrial biosphere and oceans from 1978 to 2000: observations and carbon cycle implications', pages 83-113, in *A History of Atmospheric CO2 and its effects on Plants, Animals, and Ecosystems*, editors, Ehleringer, J.R., T. E. Cerling, M. D. Dearing, Springer Verlag, New York.

Kelly, P. 2007. 'Green light on the hill is hard to miss', *The Australian*, 4 April 2007.

King, R., 2008, 'Engineers for the future', *Australian Council of Engineering Deans*, Epping

Kolb, D.A. ,1984, *Experiential learning: experience as the source of learning and development*, Prentice-Hall, NJ.

Kuznig, R., 2009, 'The Carbon Bathtub', *National Geographic*, viewed 4 June 2013 at http://ngm.nationalgeographic.com/big-idea/05/carbon-bath

Lakoff, G., 2004, *Don't think of an elephant!: know your values and frame the debate - the essential guide for progressives,* Chelsea Green Publishing, White River Junction, Vermont.

Lakoff, G. & Johnson, M., 1999, *Philosophy in the flesh: The embodied mind and its challenge to western thought,* Basic books, New York.

Lakoff, G. & Johnson, M., 2003, *Metaphors we live by,* University of Chicago Press, Chicago.

Lantz, M.E., 2010, 'The use of Clickers in the classroom: Teaching innovation or merely an amusing novelty?' *Computers in Human Behavior*, 26(4), pp. 556-61.

LaVigne, A., Benson, T., and Marlin, S., 2015, 'Developing Understanding of Dynamic Systems Within Early Childhood Settings', *Proceedings of the 2015 International System Dynamics Conference,* System Dynamics Society, Albany, NY

Lyman, F., 1987, 'Think-pair-share: An expanding teaching technique', *MAA-CIE Cooperative News*, 1(1), pp. 1-2.

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

MacArthur, J.R. & Jones, L.L., 2008, ‘A review of literature reports of clickers applicable to college chemistry classrooms’, *Chem. Educ. Res. Pract.*, 9(3), pp. 187-95.

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

Maxwell, T., Andersen, D.F., Richardson, G.P., & Stewart, T.R., 1994, ‘Mental Models and Dynamic Decision Making in a Simulation of Welfare Reform’, *Proceedings of the 1994 International System Dynamics Conference*, Stirling, Scotland, System Dynamics Society, Albany, NY.

May, K.T., 2013, ‘The 20 most popular TED Talks, as of December 2013’, *TEDBlog*, web log post, 16 December 2013, viewed 8 January 2015, http://blog.ted.com/2013/12/16/the-most-popular-20-ted-talks-2013/

Meadows, D.H., 2008, *Thinking in systems,* Wright, D. ed. Chelsea Green Publishing, White River Junction, Vermont.

Meadows, D.H., Meadows, D.L., Randers, J., Behrens III, W.W., 1972, *The Limits to Growth: a report for the Club of Rome's project on the predicament of mankind*, Universe Books, New York.

Meadows, D.H., Robinson, J.M., 2007 (1985), *The Electronic Oracle: Computer Models & Social Decisions*, reprint, John Wiley & Sons, New York.

Morecroft , J.D.W., 1982, A Critical Review of Diagramming Tools for Conceptualizing Feedback System Models. *Dynamica* 8(1)

Morecroft, J.D.W., 1988 ‘System dynamics and microworlds for policymakers’, *European Journal of Operational Research* 35 p.301-320

Morice, C.P., Kennedy, J.J. Rayner, N.A, & Jones, P.D., 2012, ‘Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset’, *J. Geophys. Res.*, 117, D08101.

Newell, B., 1988, *Turtle confusion: Logo puzzles and riddles,* Curriculum Development Centre, Woden, ACT.

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

PA Consulting, 2009, *Dynamic Planning for COIN in Afghanistan*, viewed 3 January 2015 at: http://msnbcmedia.msn.com/i/MSNBC/Components/Photo/_new/Afghanistan_Dynamic_Planning.pdf

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

Papert, S., 2005, 'You can't think about thinking without thinking about thinking about something'. *Contemporary Issues in Technology and Teacher Education*, 5(3/4), 366-367.

Piaget, J., 1973, *To understand is to invent: the future of education,* Grossman Publishers, New York.

Reddy, M.J. 1979, 'The conduit metaphor - a case of frame conflict in our language about language', in A Ortony (ed), *Metaphor and thought,* Cambridge University Press, Cambridge, pp. 284-324.

Richardson, G.P., 2011, 'Reflections on the foundations of system dynamics', *System Dynamics Review*, 27(3), pp. 219-43.

Richardson, G.P. & Pugh, A.L.I., 1981, *Introduction to system dynamics modeling,* Pegasus Communications, Waltham, MA.

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

Robinson, K. 2006, *Do schools kill creativity?*, online video, February 2006, TED Conferences LCC, New York, viewed 16 December 2013, http://www.ted.com/talks/ken_robinson_says_schools_kill_creativity

Rockström, J., Steffen, W., Noone, K., Persson, Chapin, F.S., Lambin, E.F., Lenton, T.M., Scheffer, M., Folke, C. & Schellnhuber, H.J., 2009, 'A safe operating space for humanity', *Nature*, 461(7263), pp. 472-5.

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

Rogers. S., 2010, 'The McChrystal Afghanistan PowerPoint slide: can you do any better?', *The Guardian Online*, viewed 3 January 2015 at: http://www.theguardian.com/news/datablog/2010/apr/29/mcchrystal-afghanistan-powerpoint-slide

Saeed, K., 2013, *System Dynamics: A disruptive science. A conversation with Jay W. Forrester, founder of the field*, viewed online 4 January 2015 at: http://cpe.wpi.edu/images/CPE-New/System_Dynamics_Online_-_forrester2013_interview_edited_short.pdf

Scripps, 2015, 'Atmospheric CO2 Data', *Scripps Institution of Oceanography*, UC San Diego, viewed 11 January 2015, http://scrippsco2.ucsd.edu/data/atmospheric_co2.html

Sedlmeier, P., Brockhaus, F. & Schwarz, M. 2014, 'Visual integration with stock-flow models: How far can intuition carry us?' in P Bender, R Hochmuth, PR Fischer, D Frischemeier & T Wassong (eds), *Mit Werkzeugen Mathematik und Stochastik lernen - Using Tools for Learning Mathematics and Statistics,* Springer Fachmedien Wiesbaden, Wiesbaden.

Senge, P.M., 2006, *The fifth discipline: the art and practice of the learning organization,* Doubleday/Currency, New York.

Shannon, C.E. & Weaver, W., 1949, *The Mathematical Theory of Communication,* University of Illinois Press, Champaign, IL.

Smith, J. & Browne, C.A., 2013, 'Pieces of the puzzle - impacts of using the jigsaw classroom within peer-facilitated tutorials on engineering design projects.' *Proceedings of the 24th Annual Conference for the Australasian Association for Engineering Education.* Gold Coast.

Stager, G., 2005, Papertian Constructionism and the Design of Productive Contexts for Learning, EuroLogo 2005 conference.

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W. & de Wit, C.A., 2015, Planetary boundaries: Guiding human development on a changing planet, *Science*, p. 1259855.

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

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.

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.

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P.M., 2013, 'IPCC, 2013: Summary for Policymakers. In Climate Change 2013: The Physical Science Basis', *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge Univ. Press, Cambridge, UK.

TED, n.d., *The most popular talks of all time,* TED Conferences LCC, New York, viewed 4 December 2015, https://www.ted.com/playlists/171/the_most_popular_talks_of_all

Turner, G. M. 2008, 'A comparison of The Limits to Growth with 30 years of reality', *Global Environmental Change*, 18, pp. 397-411.

Trevelyan, J., 2010, 'Mind the Gaps: Engineering Education and Practice', *Proceedings of the 21st Annual Conference for the Australasian Association for Engineering Education*. Sydney.

Trevelyan, J 2014, *The Making of an Expert Engineer: How to Have a Wonderful Career Creating a Better World and Spending Lots of Money Belonging to Other People*. in Not available. CRC Press/Balkema, London, pp. 581.

UNFCCC, 2009, 'Copenhagen Accord'. *U.N. Framework Convention on Climate Change*. United Nations. 18 December 2009.

UNFCCC, 2015, 'Adoption of the Paris Agreement'. *U.N. Framework Convention on Climate Change*. United Nations. 12 December 2015.

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

Vygotsky, L. S., 1978, *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.∆

Vygotsky, L. S., 1986, *Thought and language* MIT Press, Cambridge, MA.

Warren, K., 2014, 'Agile SD: fast, effective, reliable', *Proceedings of the 2014 International System Dynamics Conference*, Delft, The Netherlands, System Dynamics Society, Albany, NY.

Zhao, Y., 2014, *Who's afraid of the big bad dragon? Why China has the best (and worst) education system in the world,* Jossey-Bass, San Francisco, CA.

Abbreviations

General abbreviations used in text

ABC	Australian Broadcasting Commission
ANU	Australian National University
CCM	collaborative conceptual modelling
CCS	carbon capture and storage
CLE	Creative Learning Exchange
COIN	counter-insurgency
DC	direct current
DMAIC	define, measure, analyse, improve and control
GtC	gigatons of carbon
IPCC	Intergovernmental Panel on Climate Change
MOOCs	massive open online courses
NYSF	National Youth Science Forum
OECD	Organisation for Economic Co-operation and Development
ppm	parts per million
TMI	think, make, improve
UNFCCC	United Nations Framework Convention on Climate Change

Note: see the Glossary on page xv for abbreviations associated with the design of the T&P activity.