Journal of Environmental Psychology 32 (2012) 19–29

Contents lists available at SciVerse ScienceDirect

Journal of Environmental Psychology

journal homepage: www.elsevier.com/locate/jep

# Decisions from experience reduce misconceptions about climate change

Varun Dutt*, Cleotilde Gonzalez

*Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, 4609 Winthrop Street, Pittsburgh, PA 15213, USA*

ARTICLE INFO

*Article history:*
Available online 3 November 2011

*Keywords:*
Climate change
Decisions from description
Decisions from experience
Repeated feedback
STEM
Simulators

ABSTRACT

Research has shown widespread misconceptions in public understanding of the dynamics of climate change: A majority of people incorrectly infer that carbon-dioxide ($CO_2$) concentrations can be controlled by stabilizing emissions at or above current rates (correlation heuristic), and while emissions continuously exceed absorptions (violation of mass balance). Such misconceptions are likely to delay actions that mitigate climate change. This paper tests a way to reduce these misconceptions through experience in a dynamic simulation. In a laboratory experiment, participants were randomly assigned to one of two conditions: *description*, where participants performed a $CO_2$ stabilization (CS) task that provided them with a $CO_2$ concentration trajectory over a 100 year period and asked them to sketch the corresponding $CO_2$ emissions and absorptions over the same period; and *experience*, where participants performed the same task in a dynamic climate change simulator (DCCS), followed by the CS task. In both conditions, half of the participants were science and technology (STEM) majors, and the other half were non-STEM. Results revealed a significant reduction in people's misconceptions in the *experience* condition compared to the *description* condition. Furthermore, STEMs demonstrated better performance than non-STEMs. These results highlight the potential for using experience-based simulation tools like DCCS to improve understanding about the dynamics of climate change.



## 1. Introduction

Despite strong scientific consensus about the causes and risks of climate change, the general public exhibits a complacent attitude towards actions benefiting Earth's climate (Bostrom, Morgan, Fischhoff, & Read, 1994; Leiserowitz, 2007; Read, Bostrom, Morgan, Fischhoff, & Smuts, 1994; Weber, 2006). For example, based upon evidence from surveys of people's beliefs and policy preferences, a large majority would likely advocate *wait-and-see* preferences: They would like to delay significant actions to reduce greenhouse gas emissions until impacts have been more convincingly demonstrated (Kull, 2001; Leiserowitz, 2007; Sterman & Booth Sweeney, 2002, 2007). For example, 60% of participants in a U.S. survey chose either the option "until we are sure that global warming is really a problem, we should not take any steps that would have economic costs," or the option "its effects will be gradual, so we can deal with the problem gradually" (Kull, 2001). This wait-and-see preference is also seen among policymakers: "Slow the growth of greenhouse gas emissions (GHGs), and – as the science justifies – stop, and then reverse that growth" (G. W. Bush, 2/14/02; Jones, 2002). According to Jones (2002), George W. Bush believed that climate mitigation actions could be taken at a slow pace until science confirmed climate change as a real problem.

Furthermore, some scientists also seem to possess a stronger wait-and-see (inaction) view on climate change. For example, Fred Singer, professor emeritus of environmental sciences at the University of Virginia and an ex-member of the U.S. National Advisory Committee on Oceans and Atmosphere, recently commented: "Human activities are not influencing the global climate in a perceptible way. Climate will continue to change, as it always has in the past, warming and cooling on different time scales and for different reasons, regardless of any human action" (Singer, 2009, p. 1). Thus, Singer argues that human activity has no influence on climate change whatsoever, which would result in inaction rather than a slow wait-and-see action. Moreover, climate initiatives like the Kyoto Protocol and Clear Skies, which have pledged to mitigate the global warming problem, have also expressed support for wait-and-see preferences: The Kyoto Protocol's proposed reductions in emissions fall short of the proposed targets and Clear Skies' initiative promotes even further greenhouse gas emissions growth (Sterman & Booth Sweeney, 2002, 2007).

Wait-and-see preferences would work well in simple systems with short delays between the detection of a problem and the implementation of corrective actions. For example, one can afford to wait-and-see when boiling beans until steam builds up and the

* Corresponding author. Tel.: +1 412 628 1379; fax: +1 412 268 6938.
*E-mail addresses:* varundutt@cmu.edu, varundutt@yahoo.com (V. Dutt).

0272-4944/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvp.2011.10.003

cooker whistles because there is a short delay between the whistle and removing the cooker from the flame. Unfortunately for a complex system like Earth's climate, there are much longer delays between the decision to mitigate emissions and the corresponding changes in atmospheric GHG concentrations (IPCC, 2007; Sterman, 2008; Sterman & Booth Sweeney, 2002, 2007). Prior research shows that people often ignore long feedback delays in complex systems (Sterman, 1989), and those who exhibit wait-and-see preferences might be acting under the implicit misconception of very short delays in Earth's climate system (Sterman, 2008; Sterman & Booth Sweeney, 2002, 2007). As there are long feedback delays, however, wait-and-see preferences become problematic. Because even if mitigation actions are taken, atmospheric $CO_2$ accumulation would continue to rise until emissions fell below the absorptions rate. Average atmospheric temperature would then peak, and consequences such as rising sea levels and thermal expansion would continue (Meehl et al., 2005; Wigley, 2005). Therefore, wait-and-see preferences are likely to cause abrupt, persistent, and costly regime changes on Earth in the future (Alley et al., 2003; Scheffer, Carpenter, Foley, Folkes, & Walker, 2001).

Prior research has shown that people's misconceptions about the climate system are related to their own deficient mental models[1]: The general public lacks training in climatology and has little understanding of climate processes (Bostrom et al., 1994; Kasemir et al., 2000; Kempton, 1997; Morgan, Fischhoff, Bostrom, & Atman, 2002; Palmgren, Morgan, de Bruin, & Keith, 2004; Read et al., 1994). In this paper, however, we argue that people's misconceptions are due to a more fundamental limitation of their mental models: A weak understanding of accumulation and mass balance concepts rather than to the particulars of climatology and the climate system. Cronin, Gonzalez, and Sterman (2009) have demonstrated that as the relationship between inflows and outflows become more complex, people tend to rely more on simple but erroneous heuristics. According to Cronin et al. (2009), people rely on the "correlation heuristic," whereby they wrongly infer that the system's accumulations are positively correlated to its inflows.

Through laboratory studies, Sterman (2008) and Sterman and Booth Sweeney (2007) have shown that people's wait-and-see preferences on climate are related to their reliance on the correlation heuristic. For climate, relying on the correlation heuristic means wrongly inferring that an accumulation ($CO_2$ concentration) follows the same path as the inflow ($CO_2$ emissions); hence, stabilizing emissions would rapidly stabilize the concentration, and emissions cuts would quickly reduce the concentration and damages from climate change. Consequently, people who rely on the heuristic would demonstrate wait-and-see preferences because they would significantly underestimate the delay between reductions in $CO_2$ emissions and in the $CO_2$ concentration. They would also underestimate the magnitude of emission reductions needed to stabilize the concentration. Furthermore, Sterman and Booth Sweeney (2007) have also shown that people's wait-and-see preferences are also related to the violation of mass balance, whereby people incorrectly infer that atmospheric $CO_2$ concentration can be stabilized even when emissions exceeds absorptions. Violating mass balance leads to wait-and-see preferences because people think the current state of the climate system, where emissions are double that of absorptions (IPCC, 2007), would not pose a problem to future stabilization.

[1] By "mental model" we mean a person's inferences or judgments about the networks of causes and effects that describe how a system operates which include the system's boundary (i.e., factors are considered endogenous or exogenous) and its time horizon. Therefore, in this paper, the term "mental model" refers to participants' inferences about shapes of $CO_2$ emissions and absorptions overtime.

Prior research has evaluated people's misconceptions and related wait-and-see preferences in terms of correlation heuristic reliance and mass balance violation in a one-shot paper-and-pencil climate stabilization (CS) task (Sterman, 2008; Sterman & Booth Sweeney, 2007). In the CS task, participants are asked to sketch $CO_2$ emissions and absorptions that would stabilize the $CO_2$ concentration to an attainable goal by the year 2100. In this problem, they are given the concentration's starting value in the year 2000, and its historic trends and emissions between the years 1850 and 2000. Sterman and Booth Sweeney (2007) report that about 70% of participants (about 60% of whom had backgrounds in science, technology, engineering, and management (STEM), and a majority of the rest in economics) sketched $CO_2$ emissions that were positively correlated with the $CO_2$ concentration. Moreover, 74% of participants violated mass balance in their responses either by failing to keep emissions greater than absorptions before the concentration stabilized in the year 2100; or failing to make emissions equal to absorptions when the concentration reached 2100.

Sterman (2008) and Sterman and Booth Sweeney (2007) made a qualitative claim that using simulation-based tools would likely help people correct their misconceptions about Earth's climate. Other researchers also suggest that experiencing the adverse consequences of climate change is likely to improve people's understandings of the climate system (Weber, 2006). However, the efficiency of simulation tools in reducing people's reliance on the correlation heuristic and the violation of mass balance has only been demonstrated in some initial attempts (Dutt & Gonzalez, 2009, in press; Moxnes & Saysel, 2009). Moxnes and Saysel (2009) used a simulated computer task where participants were required to stabilize the $CO_2$ concentration by making emissions decisions every 10 simulated years starting in the year 2010. After every 10 years elapsed, participants could see the changes in the concentration as a result of their decisions. Moxnes and Saysel (2009) demonstrated that better emission decisions are possible through providing repeated feedback about decision actions and outcomes to participants. Feedback empowers participants to try new hypotheses and also to understand the cause-and-effect relationships between their decisions and outcomes.

Building on these results, we developed a very simplified but interactive computer-based simulation of the climate system called the Dynamic Climate Change Simulator (DCCS), and used it to collect data on how participants control the atmospheric $CO_2$ accumulation to a goal under different conditions of feedback delays (Dutt & Gonzalez, 2009, in press). The two types of manipulated feedback delays employed in the DCCS were the natural delays in $CO_2$ absorptions, and the frequency with which multiannual emission policies are revised for a simulated climate system. We found that participants improved their control of the $CO_2$ concentration through experiences gained in DCCS, where these experiences might have enabled participants to revise their existing mental models. But again, the efficiency of simulation tools against more traditional descriptive methods has not been fully demonstrated.

## 2. Current research

Given people's widespread misconceptions about the climate system, research is critically needed that shows how their misperceptions can be overcome through experience in simulation tools. The main objective of this paper is to evaluate whether or not experiencing repeated outcome of decisions (i.e., feedback) in DCCS reduces participant's misconceptions about our climate.

DCCS provides repeated feedback on the changes in the $CO_2$ concentration each year as a result of $CO_2$ emission and absorption policies set by participants, allowing participants to observe the results of their decisions as they try to control the concentration to a goal. We compare participants' responses to the CS task after previously making repeated decisions in DCCS to other participants' responses in the CS task where they were not given the DCCS experience. We expect that repeated feedback in DCCS will affect the responses made in the CS task. The hypothesis is:

**H1**. There will be a reduction in misconceptions (relying on the correlation heuristic and violating mass balance) for participants who receive repeated feedback compared to those who do not receive repeated feedback.

Research has observed misconceptions in the CS task among participants both with and without a scientific (STEM) background (Sterman & Booth Sweeney, 2007). Sterman and Booth Sweeney (2007) have suggested that the technical background of STEM participants does not help them reduce their reliance on the correlation heuristic or their violation of mass balance. But they have not tested STEMs' and non-STEMs' misconceptions systematically; they have not tested the background or the relationship of background to performance in the CS task. Considering the widespread misconceptions prevalent among scientists, general public, and policymakers (Nordhaus, 1994), it becomes important to determine the value of a STEM education in reducing misconceptions. In this regard, psychological research has found that novices often focus on the surface (i.e., irrelevant) features of a problem, rather than on more fundamental underlying structural features (Chi, Feltovitch, & Glaser, 1981; Gonzalez & Wong, in press; Schoenfeld, 1982). According to Schoenfeld (1982), this difference in focus on the surface versus the structure is affected by a person's background in mathematics and sciences. A person with a STEM background possesses far greater experience in mathematical and scientific problem-solving compared to someone with a non-STEM background. The mathematical background is expected to help STEMs focus more on the task structure compared to non-STEMs, and enable STEMs to nurture fewer misconceptions about the climate system compared to non-STEMs. The hypothesis is:

**H2**. There will be a reduction in misconceptions (relying on correlation heuristic and violating mass balance) for people from STEM backgrounds compared to people from non-STEM backgrounds.

## 3. Methods

To test our hypotheses, we conducted a laboratory experiment using STEM and non-STEM participants who either performed the CS task only (*description* condition), or completed DCCS then followed by the CS task (*experience* condition).

### 3.1. Participants

One hundred and twenty participants from Carnegie Mellon University and from the surrounding Pittsburgh area were invited to participate through an online advertisement and were randomly assigned to either the *description* or *experience* condition (60 participants in each condition). Out of the 60 participants in each condition, 30 were from STEM backgrounds and 30 were from non-STEM backgrounds. STEM backgrounds included majors in the fields of science, technology, engineering, management, economics, and medicine. Non-STEM backgrounds included majors in the fields of the arts, social sciences, and the humanities. Among the 120 participants, 2 were pursuing Ph.D. degrees, 48 were pursuing Masters or MBA degrees, and 70 were pursuing undergraduate degrees. Fifty-five participants were females. The mean age was 23 years (S.D. = 6), and ages ranged from 18 to 55 years. All participants received a flat compensation of $5 for participating in the experiment.

### 3.2. The CS task

In the CS task used here, participants were first told that the amount of $CO_2$ in the atmosphere is affected by anthropogenic $CO_2$ emissions (emissions resulting from human activity), and natural processes that gradually absorb $CO_2$ from the atmosphere (for example, $CO_2$ is used by plant life and dissolves in oceans). Furthermore, participants were shown the historic trend of emissions and the resulting $CO_2$ concentration over a 150 year period from 1850 to 2000. They were also told that in the year 2000, the absorption of atmospheric $CO_2$ by natural processes was half of $CO_2$ emission. As a result, atmospheric $CO_2$ concentrations increased from preindustrial 1850 levels of about 600 GtC to about 769 GtC in 2000. Fig. 1A shows the graphs with the historic trend. Participants were also graphically shown a scenario in which the $CO_2$ concentration gradually rose to 938 GtC, about 22% higher than its year 2000 level, and then stabilized by the year 2100 (see Fig. 1B). Participants were also provided a separate graph showing $CO_2$ emissions from 1900 to 2000 and $CO_2$ absorption from the atmosphere in 2000. They were then asked to sketch the likely course of future $CO_2$ absorptions and emissions between 2001 and 2100 that corresponded to the concentration scenario in Fig. 1B. Finally, participants were asked to clearly explain the reasons for which they drew their $CO_2$ absorptions and emissions.

The 938 GtC stabilization value in Fig. 1B was taken from the IPCC's Fourth Assessment Report (FAR), which considers 938 GtC (=450 ppmv) to be a realistically attainable climate goal for the future (IPCC, 2007). Obviously, the objective in the CS task was not to test a participant's knowledge of future $CO_2$ emissions and absorptions (which no one really knows), or to make predictions on future trends (as climate scientists would do). Rather, the goal was to test whether or not participants' $CO_2$ emission and absorption responses reflect misconceptions from relying on the correlation heuristic or violating mass balance.

The CS task used in this paper is identical to the one used by Sterman and Booth Sweeney (2007) and Sterman (2008) with one minor difference: In their task, the "Anthropogenic $CO_2$ emissions" and "net removals" (i.e., the inflow and outflow) were provided to participants in units of GtC/year and the $CO_2$ concentration in units of ppmv; whereas, we consistently used GtC for the $CO_2$ concentration and GtC/year for $CO_2$ emissions and absorptions.[2] The different units of measurement used by Sterman and Booth Sweeney (2007) could be somewhat responsible for any misconceptions in their study. Participants may likely infer that the $CO_2$ concentration is unrelated to $CO_2$ emissions and absorptions.

### 3.3. Dynamic Climate Change Simulator (DCCS)

DCCS was developed based on previous research with generic stock-and-flows control tasks (Gonzalez & Dutt, 2011). The simulation, its validation, and full functions are explained in different publications (Dutt & Gonzalez, 2009, in press). For the purpose of the current research, we created the same $CO_2$ concentration scenario in DCCS as that depicted in the CS task. In DCCS, people made $CO_2$ emission and absorption decisions each simulated year and got feedback on how their decisions effected the $CO_2$ concentration (see Fig. 2). Participants made decisions by entering values for $CO_2$ emission and absorption (Fig. 2.1) during each simulated year from 2001 to 2100 (Fig. 2.2) by clicking the "Make Decision" button. The task needed participants to enter $CO_2$ emissions and absorptions so that the resultant $CO_2$ concentration would closely

[2] GtC/year = $10^9$ tons of Carbon/year or billion tons of Carbon/year and ppmv = parts per million by volume.

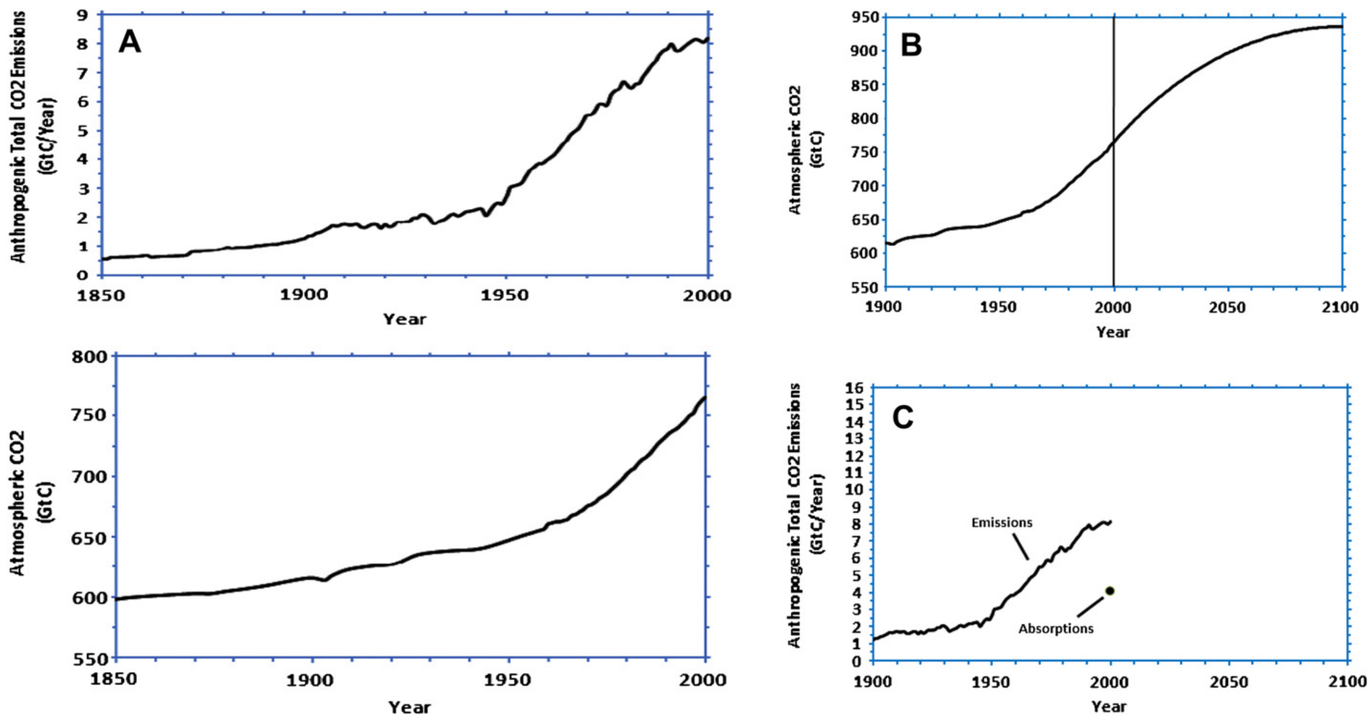


**Fig. 1.** (A) Figs. 1 and 2 that were given to participants as part of the instructions in the CS task. (B) The trajectory of $CO_2$ concentration given to participants over a 200 year period (from 1900 to 2100). (C) Participants had to sketch the trajectory of $CO_2$ absorptions and emissions corresponding to the trajectory of $CO_2$ concentration given in Fig. 1(B).

follow the $CO_2$ concentration scenario (shown in Fig. 1B) for every year until the year 2100. The yearly value of the $CO_2$ concentration goal was derived from the concentration scenario and was displayed as a red line on the atmospheric tank (Fig. 2.3). The values of the upper and lower bounds of the annual goal were displayed on the left hand side of the tank and an acceptable range of ±0.5 GtC was assumed around it. The ±0.5 GtC range was taken from Sterman and Booth Sweeney (2007) as a reasonable difference

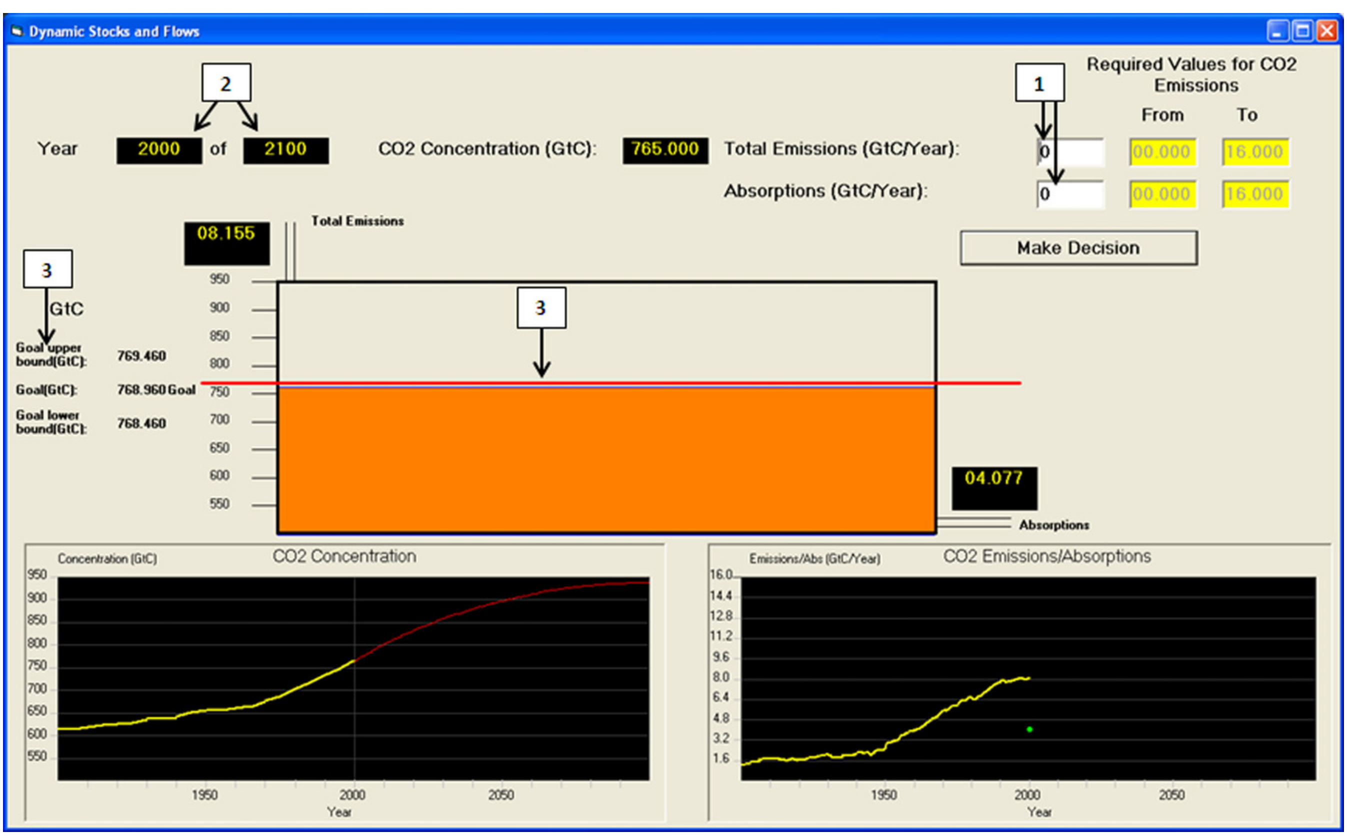


**Fig. 2.** The Dynamic Climate Change Simulation (DCCS).

between $CO_2$ emissions and absorptions at which the concentration stabilized at a yearly goal value. This range also helped us to classify participants according to whether or not their yearly $CO_2$ emission and absorption decisions violated mass balance.

### 3.4. Violating mass balance and relying on the correlation heuristic

The climate problem presented in the CS task and DCCS has a simple structure with one $CO_2$ concentration, one $CO_2$ emission inflow, and one absorption outflow. $CO_2$ emissions add to the existing $CO_2$ concentration each year, and $CO_2$ absorption subtracts from it. When $CO_2$ emissions are more than absorptions, the concentration increases; and when $CO_2$ emissions are less than absorptions, the concentration decreases. Given that absorptions are half of emissions at the start of the CS and DCCS (year 2000) tasks, the concentration is increasing. In order to stabilize the concentration at 938 GtC in 2100, one needs to keep emissions above absorptions before 2100, and equalize emissions and absorptions in 2100. Therefore, a trajectory that keeps $CO_2$ emissions greater than the absorption between years 2001 and 2099, and equalizing them in 2100 is the only way that will not violate mass balance. If a participant's response did not satisfy any of the two previous constraints, then it violated mass balance. To classify participants as violating mass balance, we visually determined whether or not a $CO_2$ emission trajectory was greater than the $CO_2$ absorption trajectory before year 2100; and if emissions were less than or equal to $\pm 0.5$ GtC/year away from absorption in 2100 (a very small gap). If either of these two conditions was not met, a participant's response was classified as violating mass balance. Two independent raters that were blind to the hypotheses under test evaluated participants' sketched $CO_2$ emissions and absorptions in both the CS task and DCCS. The inter-rater reliability statistic for the two independent raters was Kappa = 0.85 ($p < .001$), 95% CI (0.80, 0.90). This Kappa statistic reveals a satisfactory level of agreement between the two raters (Landis & Koch, 1977).

To determine participants' reliance on the correlation heuristic, we correlated their $CO_2$ emission values over a 100 year period (between years 2001 and 2100) to their $CO_2$ concentration values over the same period in both the CS task and DCCS. We assumed a very conservative threshold of 0.8 for the correlation coefficient value for classifying participants' responses as relying on the correlation heuristic. To find the mean emissions, we separately averaged the $CO_2$ emissions values entered or sketched by participants in different tasks over a 100 year period. Thus, a participant's response was classified as relying on the correlation heuristic if the correlation coefficient between their emissions and concentration trajectory over a 100 year period was more than 0.8.

### 3.5. Experiment's design and dependent measures

Participants were randomly assigned to one of the two between-subjects conditions: *description* or *experience*. In the *description* condition, participants were asked to do only the paper-and-pencil CS task. In the *experience* condition, participants performed the DCCS task first and then did the CS task. The CS task was given to participants in both conditions because it constitutes the main testing phase, which allowed us to evaluate their reliance on the correlation heuristic and violation of mass balance with or without DCCS experience. We wanted to evaluate the influence of two independent measures in this experiment: the experience in DCCS (present in *experience* condition and absent in *description* condition) and the participants' backgrounds (STEM or non-STEM).

To test H1, we compared participants' reliance on the correlation heuristic and their violation of mass balance in the CS task across the *experience* and *description* conditions. Furthermore, we explored the correlation heuristic reliance and mass balance violations among STEM and non-STEM backgrounds between the CS task in the *experience* and *description* conditions, respectively. In order to test H2, we explored correlation heuristic reliance and mass balance violations between STEM and non-STEM backgrounds within the CS task in the *description* condition, and within DCCS and the CS task in the *experience* condition. Later, we also coded and analyzed participants' explanations about the reasoning and procedure they followed while sketching $CO_2$ emissions and absorptions.

### 3.6. Procedure

In the *description* condition, participants were first asked to read the instructions as part of the CS task and the experimenter answered participants' questions at this point, if any. Participants were not given any details on how they could solve the task correctly, only clarification questions about instructions were answered. Then, they were asked to sketch the $CO_2$ emissions and absorptions over the 100 year period. No participant took more than 15 min on the CS task in either the *description* or *experience* conditions. In order to equalize the length of the *description* and *experience* conditions, participants in the *description* condition were given an unrelated task at the beginning of their experiment.

Participants assigned to the *experience* condition were first asked to read the instructions that appeared on a computer screen before they could start in DCCS. The experimenter then answered any questions clarifying the instructions, if any. Participants were told that after DCCS, they would be asked to respond to a short paper-and-pencil task, but they were not shown the CS task at that time. Once participants finished the DCCS, they were handed the CS task immediately after.

## 4. Results

### 4.1. Misconceptions across description and experience

Misconceptions were analyzed by first considering the percentage of participants that relied on the correlation heuristic and violated mass balance. Fig. 3 shows the response of a typical participant in the *description* condition. For this participant, $CO_2$ emission exceeds absorption by a large difference in the year 2100, which violates mass balance. Furthermore, the shape of the $CO_2$

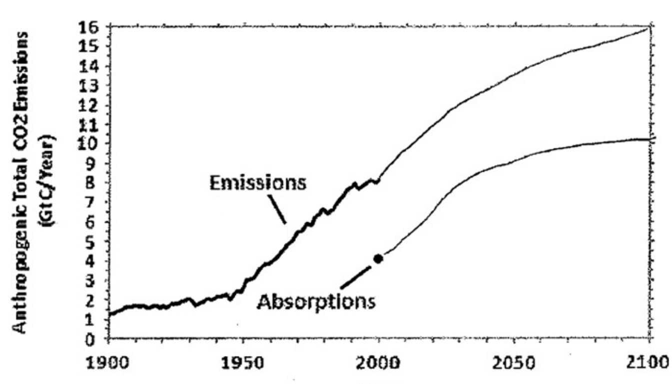


**Fig. 3.** A typical sketch by a non-STEM participant in the *description* condition. The participant sketches $CO_2$ emission much greater than $CO_2$ absorption in the year 2100. The trajectory of $CO_2$ emissions shape is correlated and similar to $CO_2$ goal trajectory that was given to the participant as part of the instructions. The gap between the emission and absorption is more than $\pm 0.5$ GtC/year in the figure.

emissions curve is highly correlated to the shape of the $CO_2$ concentration trajectory given in the problem, showing the participant's reliance on the correlation heuristic. Therefore, by responding with the shapes of $CO_2$ emissions and absorptions curves in Fig. 3, this participant relies on the correlation heuristic and violates mass balance.

Fig. 4 shows the proportion of participants relying on the correlation heuristic in the CS task in the *experience* and *description* conditions. The aggregated result is further divided by STEM and non-STEM participants. Overall, a larger proportion of participants relied on the correlation heuristic in the *description* condition compared to those in the CS task in the *experience* condition ($82\% > 60\%$; $\chi^2\,(1) = 6.82$, $p < .01$, $r^3 = .24$). This finding agrees with hypothesis H1.

Fig. 5 shows the proportion of participants violating mass balance in the CS task in the *experience* and *description* conditions. The aggregated result is further divided by STEM and non-STEM participants. Overall, a larger proportion of participants violated mass balance in the *description* condition compared to those in the CS task in the *experience* condition ($80\% > 57\%$; $\chi^2\,(1) = 7.55$, $p < .01$, $r = .51$). Again, this finding agrees with hypothesis H1.

In order to understand the reasons why participants violate the principle of mass balance, we broke results into constituting subparts as shown in Table 1: Net $CO_2$ emissions in 2100 > 0.5 GtC/year, Net $CO_2$ emissions in 2100 < 0.5 GtC/year, and Net $CO_2$ emissions in 2100 between ±0.5 GtC/year. The Net $CO_2$ emissions in 2100 (denoted by the short form "NET E") equals $CO_2$ emission minus the absorption in the same year. In order not to violate mass balance, the NET E should ideally be equal to 0 GtC/year because participants were asked to stabilize the $CO_2$ concentration in 2100. The assumption of ±0.5 GtC/year in Table 1 serves as a less stringent condition for participants. Thus, a participant's NET E could be as high as 0.5 GtC/year or as low as −0.5 GtC/year and still not be classified as violating mass balance. The Average Absolute NET E refers to the positive value of Average Net $CO_2$ emissions in 2100 (where the averaged is taken over all participants in the respective condition). "VOMB?" refers to the proportion of participants violating mass balance.

As seen in Table 1, a large proportion of participants in the *description* condition (80%) wrongly inferred that the $CO_2$ concentration still could be stabilized in 2100 if $CO_2$ emissions could be either greater than or less than $CO_2$ absorptions (column labeled: "Net E > 0.5 or Net E < −0.5"); whereas, in the *experience* condition, a smaller proportion demonstrated the same incorrect inference in the CS task (57%; $\chi^2\,(1) = 11.63$, $p < .001$, $r = .31$) and in DCCS (26%; $\chi^2\,(1) = 43.25$, $p < .001$, $r = .49$). Similarly, a large proportion of participants wrongly inferred that the $CO_2$ concentration could still be stabilized in 2100 if $CO_2$ emissions exceed $CO_2$ absorptions (column labeled: "Net E > 0.5") in the *description* condition (78%); whereas, in the *experience* condition, a smaller proportion of participants showed this wrong inference in the CS task (50%; $\chi^2\,(1) = 11.63$, $p < .001$, $r = .31$) and in DCCS (18%; $\chi^2\,(1) = 43.25$, $p < .001$, $r = .49$).

Finally, participants in the CS task and DCCS (*experience* condition) had a significantly smaller Average Absolute Net E in the year 2100 compared to participants in the *description* condition ($U = 1133.00$, $Z = -3.536$, $p < .001$, $r = -.32$ and $U = 542.50$, $Z = -6.772$, $p < .001$, $r = -.62$, respectively). Thus, the experience gained in DCCS helped participants to reduce mass balance violations. Furthermore, the reduction helped participants to perform better in the following CS task in the *experience* condition compared to that in the *description* condition.

[3] This refers to the effect size unless otherwise mentioned.

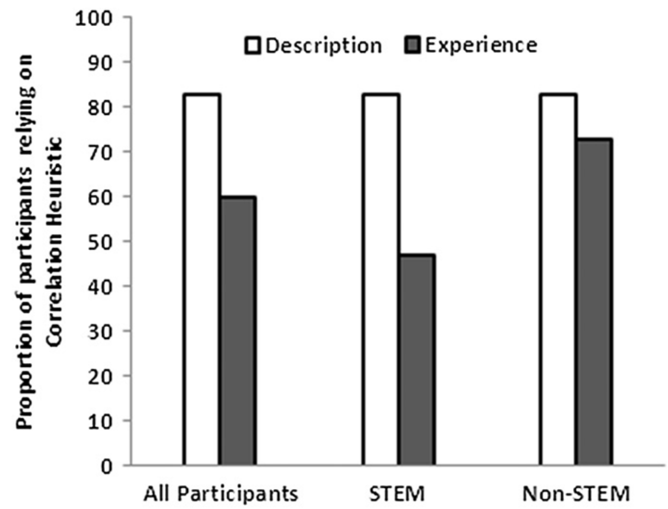


**Fig. 4.** Proportion of participants relying on correlation heuristic in the CS task in the *description* and *experience* conditions. The figure also shows the breakup of correlation heuristic for STEM and non-STEM backgrounds.

Altogether, participants in the *experience* condition showed fewer misconceptions (i.e., less reliance on the correlation heuristic and less violation of mass balance) in the CS task compared to those in the CS task in the *description* condition.

### 4.2. Interaction between background and condition

With the additional factor of participants from both STEM and non-STEM backgrounds, DCCS's exact influence in the *experience* condition becomes difficult to untangle if there are significant interactions between participants' backgrounds and the condition given. As the main effect of experience in DCCS significantly reduced reliance on the correlation heuristic and violation of mass balance in the CS task in the *experience* condition, we investigated the effects of DCCS's experience among both STEM and non-STEM backgrounds. There was a significant interaction between the condition (*experience* or *description*) in the CS task and participants' backgrounds (STEM or non-STEM) for reliance on the correlation heuristic ($F$(1,

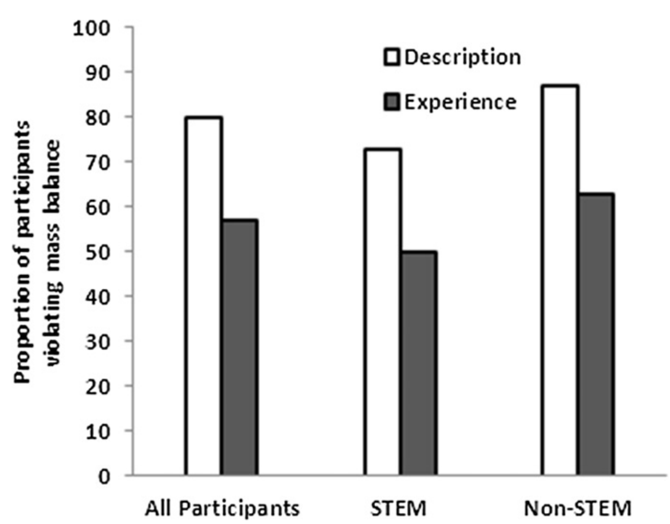


**Fig. 5.** Proportion of participants violating mass balance in the CS task in the *description* and *experience* conditions. The figure also shows the breakup of participants violating mass balance within STEM and non-STEM backgrounds.

**Table 1**
Conservation of mass balance and conformance.

| Task | Average Absolute NET E | Net $CO_2$ emissions in 2100 (called "NET E") = $CO_2$ emission − absorption in 2100 (GtC/year) | | | | | | | | VOMB? | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Net E > 0.5 ($A$) | | Net E ≥ −0.5 and Net E ≤ 0.5 | | Net E < −0.5 ($B$) | | Net E > 0.5 or Net E < −0.5 ($A$) + ($B$) | | | |
| | | $N$ | % | $N$ | % | $N$ | % | $N$ | % | $N$ | % |
| CS (description) | 5.40 (3.76) | 47 | 78 | 12 | 20 | 01 | 02 | 48 | 80 | 48 | 80 |
| CS (experience) | 2.93 (3.37) | 30 | 50 | 26 | 43 | 04 | 07 | 34 | 57 | 34 | 57 |
| DCCS (experience) | 0.66 (1.58) | 11 | 18 | 44 | 73 | 05 | 08 | 16 | 26 | 18 | 30 |
| CS (description, STEM) | 4.63 (3.63) | 21 | 70 | 08 | 27 | 01 | 03 | 22 | 73 | 22 | 73 |
| CS (description, non-STEM) | 6.21 (3.77) | 26 | 87 | 04 | 13 | 00 | 00 | 26 | 87 | 26 | 87 |
| CS (experience, STEM) | 2.18 (2.77) | 12 | 40 | 11 | 37 | 07 | 23 | 19 | 63 | 15 | 50 |
| CS (experience, non-STEM) | 3.68 (3.78) | 18 | 60 | 11 | 37 | 01 | 03 | 19 | 63 | 19 | 63 |
| DCCS (experience, STEM) | 0.56 (1.16) | 06 | 20 | 21 | 70 | 03 | 10 | 09 | 30 | 09 | 30 |
| DCCS (experience, non-STEM) | 0.77 (1.93) | 05 | 17 | 23 | 77 | 02 | 07 | 07 | 23 | 09 | 30 |

*Note:* Net $CO_2$ emission in 2100 (called "Net E") = $CO_2$ emission in 2100 − $CO_2$ absorption in 2100 and should be ideally equal to 0 GtC/year. Average Absolute Net E = | Average $CO_2$ emission in 2100 − Average $CO_2$ absorption in 2100 | when the different values are averaged over all participants in a specific task. Net E > 0.5; Net E >= −0.5 and Net E <= 0.5; and, Net E < −0.5 refer to when the Net $CO_2$ emission in year 2100 were greater than 0.5, in-between −0.5 and 0.5, or less than −0.5 GtC/year. The value of 0.5 GtC/year is the same as used by Sterman and Booth Sweeney (2007). The assumption of ±0.5 GtC/year makes the requirement on Net $CO_2$ emission in 2100 to equal 0, to be less stringent for participants. Thus, a participant's Net $CO_2$ emission in 2100 could be as high as 0.5 GtC/year or as low as −0.5 GtC/year and still she would not commit violation of mass balance. "VOMB?" refers to the proportion of participants that violate mass balance, i.e., whether or not $CO_2$ emissions > absorptions before year 2100, and, Net $CO_2$ emission in 2100 <0.5 GtC/year. A value in round brackets, "()", indicates the standard deviation around the average value.

116) = 3.6, $p < .05$). As shown in Fig. 4, the difference in reliance between the CS task in the *description* and *experience* conditions is greater for STEM participants (36%) than for non-STEM participants (10%). This difference was significant for STEMs (CS task, *description* (83%) > CS task, *experience* (47%) with $\chi^2$ (1) = 8.86, $p < .001$, $r = .38$), but not for non-STEMs (CS task, *description* (83%) = CS task, *experience* (73%) with $\chi^2$ (1) = 0.37, *ns*, $r = .08$).

The interaction between condition (*experience* or *description*) in the CS task and backgrounds (STEM or non-STEM) for violating mass balance was not significant ($F$(1, 116) = 0.0, *ns*). As shown in Fig. 5, DCCS experience led to a reduction in mass balance violations for both STEMs and non-STEMs. Therefore, as Fig. 5 and Table 1 indicate, the experience gained in DCCS (*experience* condition) reduced violation of mass balance for both STEMs (23%) and non-STEMs (24%). These differences were significant for both backgrounds: For STEMs, 73% in *description* condition >50% in *experience* condition ($\chi^2$ (1) = 3.50, $p < .05$, $r = .24$); and for non-STEMs, 87% in *description* condition > 63% in *experience* condition ($\chi^2$ (1) = 4.36, $p < .05$, $r = .27$).

### 4.3. Misconceptions among STEMs and non-STEMs (Investigating hypothesis H2)

We compared the proportion of misconceptions between STEM and non-STEM participants within the CS task in the *description* and *experience* conditions (hypothesis H2). In the *description* condition's CS task, there was excessive reliance on the correlation heuristic by both STEMs (83%) and non-STEMs (83%), and the difference between these groups was not significant ($\chi^2$ (1) = 0.11, *ns*, $r = .04$) (see Fig. 4). Similarly, as seen in Fig. 5, the difference between STEMs (73%) and non-STEMs (87%) for violating mass balance in the CS task in the *description* condition was not significant ($\chi^2$ (1) = 1.67, *ns*, $r = .17$). These results replicate Sterman's (2008) and Sterman and Booth Sweeney's (2007) findings. In contrast, in the *experience* condition's CS task, the difference in correlation heuristic reliance between STEMs (47%) and non-STEMs (73%) was significant ($\chi^2$ (1) = 4.44, $p < .05$, $r = .27$) (see Fig. 4), but the difference between STEMs (50%) and non-STEMs (63%) for violating mass balance was not ($\chi^2$ (1) = 1.09, *ns*, $r = .14$) (see Fig. 5).

Some of the reasons for the lack of differences between STEMs and non-STEMs for violating mass balance can be seen in Table 1. For example, STEMs and non-STEMs were equally likely to infer that the $CO_2$ concentration could be stabilized even when emissions do not equal the absorptions in 2100 in the CS task in both conditions (CS task, *description*: 27% versus 13% with $\chi^2$ (1) = 2.46, *ns*, $r = .20$; CS task, *experience*: 37% versus 37% with $\chi^2$ (1) = 3.27, *ns*, $r = .23$). Furthermore, STEMs and non-STEMs were equally likely to infer that the $CO_2$ concentration could still be stabilized in 2100 in both conditions when $CO_2$ emissions exceeded absorptions (column labeled: "Net E > 0.5") (CS task, *description*: 70% versus 87% with $\chi^2$ (1) = 2.46, *ns*, $r = .20$; CS task, *experience*: 40% versus 60% with $\chi^2$ (1) = 3.27, *ns*, $r = .23$). Finally, there were no significant differences between STEMs and non-STEMs in the Average Absolute Net E in the *experience* and *description* conditions (CS task, *description*: $U = 357.50$, $Z = -1.372$, *ns*, $r = -.17$; CS task, *experience*: $U = 343.00$, $Z = -1.623$, *ns*, $r = -.21$).

### 4.4. Participants' explanations

As mentioned above, all participants were asked to explain the reason for which they drew the $CO_2$ absorptions and emissions that they did. These explanations help us to investigate participants' existing mental models. Sterman and Booth Sweeney (2007) had previously analyzed participants' written explanations, and we coded our participants' written explanations according to the same procedure they detailed. Explanations were classified under several categories, which we also used: Mass Balance, Correlation Heuristic, Inertia/Delays, $CO_2$ Fertilization, Sink Saturation, and Technology. The definitions and explanations of these categories are provided under each category in Table 2. Mass balance indicated awareness of a relationship between $CO_2$ emission/absorption flows and the concentration of atmospheric $CO_2$ (i.e., where participants wrote terms such as "mass balance"). Correlation Heuristic indicated the incorrect assumption that emissions changes should be proportional to changes in atmospheric $CO_2$ (perhaps with lags or time delays). Inertia/Delays indicated mention of delays in response of system to changes in emissions, atmospheric $CO_2$, and the use of terms such as "delay," "lag," "inertia," etc. $CO_2$ Fertilization indicated the possibility

**Table 2**
Codes for written participant explanations in CS tasks (*experience*, *description*) based on rater Kappa statistics.

| Categories[a] | CS (description)[b] | | CS (experience)[b] | |
|---|---|---|---|---|
| | *N*[c] | %[d] | *N*[c] | %[d] |
| *Mass Balance* | | | | |
| Description indicating awareness of relationship between emission and absorption flows and the concentration of atmospheric $CO_2$; terms such as mass balance. | 20 | 33 | 29 | 48 |
| STEM | 15 | 50 | 16 | 52 |
| Non-STEM | 05 | 17 | 13 | 45 |
| *Correlation Heuristic* | | | | |
| Description mentioning correlations or similarity of behavior or patterns among emissions, atmospheric $CO_2$, indication that emissions change should be proportional to changes in atmospheric $CO_2$ (perhaps with lags or time delays). | 39 | 65 | 22 | 37 |
| STEM | 15 | 50 | 10 | 32 |
| Non-STEM | 24 | 80 | 12 | 41 |
| *Inertia/Delays* | | | | |
| Mention of delays in response of system to changes in emissions, atmospheric $CO_2$; terms such as 'delay,' 'lag,' 'inertia,' etc. | 10 | 17 | 03 | 05 |
| STEM | 05 | 17 | 02 | 06 |
| Non-STEM | 05 | 17 | 01 | 03 |
| *$CO_2$ Fertilization* | | | | |
| Mention of the possibility that $CO_2$ absorptions may rise due to enhanced plant growth, other effects of higher atmospheric $CO_2$ or higher temperatures. | 17 | 28 | 04 | 07 |
| STEM | 09 | 30 | 01 | 03 |
| Non-STEM | 08 | 27 | 03 | 10 |
| *Sink Saturation* | | | | |
| Mention of the possibility that $CO_2$ absorptions may fall due to Carbon sink saturation, e.g., deforestation, ocean saturation, carbon discharge stimulated by higher. | 10 | 17 | 05 | 08 |
| STEM | 04 | 13 | 00 | 00 |
| Non-STEM | 06 | 20 | 05 | 17 |
| *Technology* | | | | |
| Indicates inference that technology will enable emissions reductions (e.g., alternative energy sources) or enhance $CO_2$ absorptions (e.g., anthropogenic carbon capture and sequestration). | 06 | 12 | 04 | 08 |
| STEM | 05 | 17 | 02 | 06 |
| Non-STEM | 01 | 07 | 02 | 10 |

[a] Categories and their definitions were taken from Sterman and Booth Sweeney (2007). A single participant's explanation could be classified into multiple categories. Also, absence of a category in a participant's written explanation does not reveal whether or not the participant was aware of it. Even if the participant's conclusion was wrong but belonged to a particular category, it was classified as part of that category.
[b] The CS task (*description*) and CS task (*experience*) were given in the *description* and *experience* conditions of the experiment, respectively.
[c] *N* refers to the number of participants whose explanation included the category under consideration.
[d] % proportion of participant's explanation out of a total of 60 participants in the CS task (*description*) and 60 participants in the CS task (*experience*).

that $CO_2$ absorptions may increase due to enhanced plant growth and other effects of greater atmospheric $CO_2$ or temperatures. Sink Saturation indicated the possibility that $CO_2$ absorptions may fall due to saturation of carbon sinks (e.g., oceans). Finally, Technology indicated the assumption that technology will more easily enable $CO_2$ emissions reductions (e.g., alternative energy sources) or enhance $CO_2$ absorptions (e.g., anthropogenic carbon capture and sequestration).

Two independent raters, who were blind to the hypotheses under test, coded participant explanations. The inter-rater reliability statistics for the two independent raters was Kappa, Mass Balance = 0.95 ($p < .001$), 95% CI (0.89, 1.00); Kappa, Correlation Heuristic = 0.97 ($p < .001$), 95% CI (0.92, 1.00); Kappa, Inertia = 1.00 ($p < .001$), 95% CI (1.00, 1.00); Kappa, $CO_2$ Fertilization = 0.91 ($p < .001$), 95% CI (0.81, 1.00); Kappa, Sink Saturation = .92 ($p < .001$), 95% CI (0.81, 1.00); and Kappa, Technology = .91 ($p < .001$), 95% CI (0.78, 1.00). These Kappa statistics reveal a satisfactory level of agreement between the two raters on their individual categorizations (Landis & Koch, 1977), hence the same levels of categorization was used for subsequent analysis of participant explanations (any disagreements between raters were resolved by meeting and active discussion). Table 2 displays the frequency and proportions of explanations for the CS task in the *description* and *experience* conditions, then further broken down by participants' backgrounds. A participant's explanation could belong to more than one category at the same time.

The differences between explanations in the *description* and *experience* conditions for the Mass Balance, Inertia/Delays, Sink Saturation, and Technology categories were insignificant ($\chi^2$ (1) = 2.79, *ns*, $r$ = .15; $\chi^2$ (1) = 4.23, *ns*, $r$ = .19; $\chi^2$ (1) = 1.91, *ns*, $r$ = .13; $\chi^2$ (1) = 0.37, *ns*, $r$ = .06, respectively). The proportion of participants suggesting Correlation Heuristic in their explanations, however, was significantly lower in the *experience* than in the *description* condition (37% versus 65%; $\chi^2$ (1) = 9.64, $p < .01$, $r$ = .28). The proportion of explanations suggesting an increase in $CO_2$ absorptions due to $CO_2$ Fertilization was also significantly larger in the *description* than in the *experience* condition (28% versus 7%; $\chi^2$ (1) = 9.76, $p < .01$, $r$ = .29). The latter two differences consistently show why participants suffered from greater misconceptions in the *description* condition than in the *experience* condition.

The difference between the *description* and *experience* conditions in the proportion of Correlation Heuristic explanations was significant among non-STEMs, but not STEMs (non-STEM 80% versus 40%: $\chi^2$ (1) = 9.25, $p < .001$, $r = .40$; STEM 50% versus 32%: $\chi^2$ (1) = 1.98, *ns*, $r = .18$). On the other hand, the significant difference between the *description* and *experience* conditions for $CO_2$ Fertilization explanations was a result found among STEMs and non-STEMs (STEM 30% versus 3%: $\chi^2$ (1) = 7.97, $p < .01$, $r = .36$; non-STEM 27% versus 10%: $\chi^2$ (1) = 2.59, *ns*, $r = .21$). Therefore, there was an increase in the Correlation Heuristic explanations from the *description* to the *experience* conditions for non-STEMs; whereas, there was a reduction in $CO_2$ Fertilization explanations from the *description* to the *experience* condition for STEMs.

Finally, if STEMs demonstrated fewer misconceptions with their Correlation Heuristic and Mass Balance responses than non-STEMs, there should be significant differences in these two categories within the *description* condition between the types of backgrounds. Moreover, if DCCS is an effective manipulation, then the proportions of Correlation Heuristic and Mass Balance explanations should be respectively less and more for STEMs and for non-STEMs in the *experience* condition. We tested for these expectations. In the *description* condition, the proportion of Mass Balance explanations made by STEMs were significantly larger than those made by non-STEMs, and the proportion of Correlation Heuristic explanations were significantly less for STEMs than for non-STEMs (50% versus 17% with $\chi^2$ (1) = 7.50, $p < .01$, $r = .35$; 50% versus 80% with $\chi^2$ (1) = 5.93, $p < .05$, $r = .31$ respectively). In the *experience* condition, however, differences in the use of both categories were insignificant between STEMs and non-STEMs (52% and 45% with $\chi^2$ (1) = 0.28, *ns*, $r = .07$; 32% and 41% with $\chi^2$ (1) = 0.54, *ns*, $r = .10$). These results highlight the fact that STEMs showed fewer misconceptions in their explanations than non-STEMs in the *description* condition, but not in the *experience* condition.

## 5. Discussion and conclusions

One main and consistent result of our study is that acquiring experiential feedback in the Dynamic Climate Change Simulator (DCCS) helps to reduce participants' misconceptions about the way the climate system in the subsequent Climate Stabilization (CS) task works. Reducing misconceptions about Earth's climate is likely to reduce wait-and-see preferences (Bostrom et al., 1994; Sterman, 2008; Sterman & Booth Sweeney, 2002, 2007).

Experiential feedback in DCCS enables participants to test several hypotheses they might have about how $CO_2$ emission and absorption processes affect the $CO_2$ concentration. It is likely that the ability to test several hypotheses repeatedly about the cause-and-effect relationship in DCCS enables them to understand that the concentration increases when $CO_2$ emissions are greater than absorptions, decreases when emissions are less than absorptions, and stabilizes at a particular value when emissions equal absorptions. Consequently, participants are able to apply this understanding in the subsequent CS task. We do find some evidence for this reasoning from our results. For example, the proportion of participants' explanations indicating the correlation heuristic was far less in the *experience* condition's CS task than that in the *description* condition. Therefore, it seems that the experience gained in DCCS enabled participants to decrease their reliance on the correlation heuristic and also enabled them to improve their performance in the CS task that followed.

The above explanations about feedback in DCCS are also supported by similar findings in the literature for other dynamic tasks (Cronin et al., 2009; Dutt & Gonzalez, in press; Gonzalez, Lerch, & Lebiere, 2003; Moxnes & Saysel, 2009). For example, Cronin et al. (2009) suggest that participants can get increasingly accurate answers in simple dynamic tasks with even just "correct/incorrect" feedback given for repeated attempts. In the first attempt, only 15% of their participants answered the accumulation question correct, but 80% of the participants were able to solve the problem correctly by the seventh attempt. Similarly, it appears that participants in our study are able to successfully transfer their experiences through repeated feedback in DCCS to the CS task. Based upon DCCS' success, the use of experiential tools like it is recommended to aid in the process of formulating climate policies. Considering the reduction in correlation heuristic reliance and mass balance violations with experience after DCCS, another important implication is the use of DCCS as a tool to supplement basic education about climate change at all levels of schooling.

However, we also found differential effects of using DCCS for participants with science (STEM) and non-science (non-STEM) backgrounds: The decrease in correlation heuristic reliance between the CS task in the *description* and *experience* conditions was present for STEMs, but absent for non-STEMs. Given that neither STEMs or non-STEMs relied on the correlation heuristic during DCCS, what we can conclude is that only STEMs were able to retain and transfer the experiential knowledge previously acquired to the CS task. In contrast, non-STEMs reduced their reliance on the correlation heuristic after performing in DCCS, but this reduction was not seen in non-STEMs in the *description* condition. One probable reason is that STEMs are more commonly exposed to dynamic problems that require attaining an equilibrium or control of an accumulation as part of their prior training and curriculum. For example, science courses in thermodynamics and physics cover concepts of mass balance and energy balance as part of their curriculum. When STEMs are asked to perform in DCCS, they are likely reminded of these equilibrium concepts that they previously learnt. This background might enable STEMs to do better in the CS task compared to those that did not gain experience in DCCS. On the other hand, courses in the humanities and social sciences do not explicitly cover concepts of mass balance and energy balance. Thus, non-STEMs may not necessarily have the prior knowledge base to rely on. As learning new concepts might take time, non-STEMs are unlikely to be able to carry forward these concepts and do better in the CS task.

One could also argue that the reduction in correlation heuristic reliance among STEMs may also be due to their prior education in mathematics and sciences, which help them see the underlying stock-and-flow structure of the problem (Chi et al., 1981).[4] STEMs may be able to recognize the stock-and-flow structure and how the associated flows ($CO_2$ emissions and absorptions) affect the $CO_2$ concentration. This explanation is supported by prior research, which has shown that people with math and science expertise can learn faster and generate more meaningful categories, superior performance, better use of problem metaphors, and a deeper understanding of the problem structure compared to people without the same expertise (Chi et al., 1981; Schoenfeld, 1982). For example, Schoenfeld (1982) has shown that college freshmen students and college faculty members generate very different classifications for problems in geometry and algebra. College freshmen students classify these problems based upon surface similarities (e.g., based upon circles, functions, or whole numbers); whereas, college faculty members with prior experience show a deeper understanding of the problem structure in their classifications.

The fact that simulation tools have differential success based upon prior scientific background has important implications. First,

[4] By stock-and-flow structure we mean the ability to recognize the accumulation and its corresponding inflows and outflows, and to be able to determine how flow processes affect the accumulation in a problem.

if experience needs the support of a mathematical background to reduce misconceptions as shown in our results, then it is possible that many policymakers, who lack the needed scientific training and mathematical backgrounds, might make climate policies while relying on the correlation heuristic. Second, the design of simulation tools (e.g., what features to put in these tools and the length of training in these tools etc.) needs to be carefully considered, given that they would be directed towards both students with and without prior STEM education. Although we can only speculate, perhaps it might be beneficial to extend the length of training for non-STEMs compared to STEMs; as extended practice with simulation tools might yield more benefits for non-STEMs.

Furthermore, we find that there is a reduction in the proportion of participants violating mass balance; however, there was an absence of a corresponding increase in explanations indicating Mass Balance. This inconsistency between what people "do" versus what people "say" might be due to the previously recorded dissociation between explicit and implicit leaning (Berry & Broadbent, 1984). Therefore, the absence of an increase in the proportion of Mass Balance explanations (explicit learning) need not lead to a decrease in the proportion of participants committing a violation of mass balance (implicit learning).

Finally, the problems used in DCCS and in the following CS task in this study were identical. Therefore, there is a possibility that any improvement in participants' performance in the CS task following DCCS is because of the similarity they perceive between these two tasks. Although this observation might not necessarily constitute a problem for the effectiveness of the DCCS manipulation, it does raise an important question for future research: Whether people are learning the stock-and-flow structure of the problem while they are performing DCCS, or whether they are learning the numerical values of $CO_2$ emissions, absorptions, and $CO_2$ concentration and these shapes over time only (i.e., the surface of the problem). As participants' explanations in the CS task do show an overall reduction in their misconceptions about correlation heuristic in our results, DCCS is believed to affect the way participants observed the stock-and-flow structure in the problem. As part of future research, however, we would like design problems for the *experience* condition's CS task that are different from or similar to the one given in DCCS. By doing so, we will be able to test the boundaries of experiential learning in DCCS against similar or novel problems in the CS task. Second, we also plan to discuss misconceptions with participants after their performance in DCCS and before they complete the CS task. This second manipulation is likely to produce even stronger evidence of learning, and perhaps even more drastic reduction in people's misconceptions in the subsequent CS task.

Although we all experience climate in our day-to-day lives, it is possible that we misperceive the association between our decisions and their effects because the climate effects are vastly delayed in time. There is an important need to develop efficient methods of climate risk communication where methods are designed in accordance with people's existing mental models of climate change (Morgan et al., 2002). According to Morgan et al. (2002), simply asking experts what to do for the climate and then passing their view onto lay people generally results in lay people missing the point and becoming confused, disinterested, and even annoyed. In a world where people with non-STEM backgrounds are plentiful and their support is clearly needed, the experts should understand and pay close attention to the underlying mental models, pre-existent knowledge, and needs of lay people. Again here, the use of simulation tools like DCCS is likely to help improve lay people's understanding of the cause-and-effect relationships that govern Earth's climate.

## Acknowledgments

This research was supported by a Defense Threat Reduction Agency's grant to Cleotilde Gonzalez and Christian Lebiere (HDTRA1-09-1-0053). The authors are thankful to Hau-yu Wong, Dynamic Decision Making Laboratory, Carnegie Mellon University for providing comments on the manuscript.

## References

Alley, R., Marotzke, J., Nordhaus, W., Overpeck, J., Peteet, D., Pielke, R., et al. (2003). Abrupt climate change. *Science, 299*, 2005–2010. doi:10.1126/science.1081056.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalized knowledge. *The Quarterly Journal of Experimental Psychology, 36*(2), 209–231. doi:10.1080/14640748408402156.

Bostrom, A., Morgan, M. G., Fischhoff, B., & Read, D. (1994). What do people know about global climate change? Part 1: Mental models. *Risk Analysis, 14*(6), 959–970. doi:10.1111/j.1539-6924.1994.tb00065.x.

Chi, M. T. H., Feltovitch, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.

Cronin, M., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators and citizens. *Organizational Behavior and Human Decision Processes, 108*, 116–130. doi:10.1016/j.obhdp.2008.03.003.

Dutt, V., & Gonzalez, C. (2009). Human "mis"-perception of climate change. In *Proceedings of the human factors and ergonomics society 53rd annual meeting*. San Antonio, TX: Human Factors and Ergonomics Society.

Dutt, V., & Gonzalez, C. (in press). Human control of climate change. *Climatic Change*. doi:10.1007/s10584-011-0202-x

Gonzalez, C., & Dutt, V. (2011). A generic dynamic control task for behavioral research and education. *Computers in Human Behavior, 27*, 1904–1914. doi:10.1016/j.chb.2011.04.015.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*(4), 591–635. doi:10.1016/S0364-0213(03)00031-4.

Gonzalez, C., & Wong, H.-y. (in press). Understanding stocks and flows through analogy. *System Dynamics Review*. doi:10.1002/sdr.470

IPCC. (2007). *Climate change 2007: The physical science basis*. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Jones, A. (2002). *We can't afford to "wait-and-see" on climate change*. Retrieved September 7, 2009 from the sustainability institute web site: http://www.sustainer.org/pubs/columns/07.19.02Jones.html.

Kasemir, B., Dahinden, U., Swartling, A., Schüle, R., Tabara, D., & Jaeger, C. (2000). Citizens' perspectives on climate change and energy use. *Global Environmental Change, 10*, 169–184. doi:10.1016/S0959-3780(00)00022-4.

Kempton, W. (1997). How the public views climate change. *Environment, 39*(9), 12–21. doi:10.1080/00139159709604765.

Kull, S. (2001). *Americans on the global warming treaty*. Retrieved April 21, 2011 from the program on international policy attitudes: www.pipa.org/OnlineReports/GlobalWarming/buenosaires02.00.html.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Leiserowitz, A. (2007). International public opinion, perception, and understanding of global climate change. In *Human Development Report 2007/2008*. United Nations Development Program.

Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A., Buja, L. E., et al. (2005). How much more global warming and sea level rise? *Science, 307*, 1769–1772. doi:10.1126/science.1106663.

Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. (2002). *Risk communication: A mental models approach*. New York: Cambridge University Press.

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

Nordhaus, W. (1994). Expert opinion on climatic change. *American Scientist, 82*, 45–51.

Palmgren, C., Morgan, M. G., de Bruin, W., & Keith, D. (2004). Initial public perceptions of deep geological and oceanic disposal of carbon dioxide. *Environment Science Technology, 38*(24), 6441–6450. doi:10.1021/es040400c.

Read, D., Bostrom, A., Morgan, M. G., Fischhoff, B., & Smuts, D. (1994). What do people know about global climate change? Part 2: Survey studies of educated laypeople. *Risk Analysis, 14*(6), 971–982. doi:10.1111/j.1539-6924.1994.tb00066.x.

Scheffer, M., Carpenter, S., Foley, J., Folkes, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature, 413*, 591–596. doi:10.1038/35098000.

Schoenfeld, A. H. (1982). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education, 13*(1), 31–49.

Singer, S. F. (2009). *Climate change: Natural and unstoppable*. Retrieved September 7th 2009 from the FC Know Web site: http://www.mydigitalfc.com/news-analysis/climate-changenatural-and-unstoppable-837.

Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision making. *Organization Behavior and Human Decision Processes, 43*(3), 301–335. doi:10.1016/0749-5978(89)90041-1.

Sterman, J. D. (2008). Risk communication on climate change: Mental models and mass balance. *Science, 322*, 532–533. doi:10.1126/science.1162574.
Sterman, J. D., & Booth Sweeney, L. (2002). Cloudy skies: Assessing public understanding of global warming. *System Dynamics Review, 18*(2), 207–240. doi:10.1002/sdr.242.
Sterman, J. D., & Booth Sweeney, L. (2007). Understanding public complacency about climate change: Adults' mental models of climate change violate conservation of matter. *Climatic Change, 80*(3–4), 213–238. doi:10.1007/s10584-006-9107-5.
Weber, E. U. (2006). Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic Change, 77*(1–2), 103–120. doi:10.1007/s10584-006-9060-3.
Wigley, T. (2005). The climate change commitment. *Science, 307*, 1766–1769. doi:10.1126/science.1103934.