Correct me if I'm wrong: Groups outperform individuals in the Climate Stabilization Task

Belinda Xie[1, 2*], Mark J. Hurlstone[1], Iain Walker[3]

[1]University of Western Australia, Australia, [2]University of New South Wales, Australia, [3]University of Canberra, Australia

Submitted to Journal:
Frontiers in Psychology

Specialty Section:
Environmental Psychology

Article type:
Original Research Article

Manuscript ID:
394670

Received on:
08 May 2018

Revised on:
19 Sep 2018

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

All authors contributed to the conception and design of the study; MH created the software for running the experiment and produced the figures for the paper; BX conducted the experimental sessions, performed the data analysis, and wrote the paper; MH and IW edited the paper and reviewed the results.

Keywords

climate stabilization task, Mental Models, group decision-making, Carbon dioxide accumulation, stock-flow tasks, Emissions reduction

Abstract

Word count: 167

Avoiding dangerous climate change requires ambitious emissions reduction. Scientists agree on this, but policy-makers and citizens do not. This discrepancy can be partly attributed to faulty mental models, which cause individuals to misunderstand the carbon dioxide (CO_2) system. For example, in the Climate Stabilization Task (hereafter, ‘CST’) (Sterman & Booth-Sweeney, 2007), individuals systematically underestimate the emissions reduction required to stabilize atmospheric CO_2 levels, which may lead them to endorse ineffective ‘wait-and-see’ climate policies. Thus far, interventions to correct faulty mental models in the CST have failed to produce robust improvements in decision-making. Here, in the first study to test a group-based intervention, we found that success rates on the CST markedly increased after participants deliberated with peers in a group discussion. The group discussion served to invalidate the faulty reasoning strategies used by some individual group members, thus increasing the proportion of group members who possessed the correct mental model of the CO_2 system. Our findings suggest that policy-making and public education would benefit from group-based practices.

Funding statement

The work reported here was supported by a grant from the Climate Adaptation Flagship of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia awarded to Mark Hurlstone and Iain Walker.

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: Yes

Please provide the complete ethics statement for your manuscript. Note that the statement will be directly added to the manuscript file for peer-review, and should include the following information:

- Full name of the ethics committee that approved the study
- Consent procedure used for human participants or for animal owners
- Any additional considerations of the study in cases where vulnerable populations were involved, for example minors, persons with disabilities or endangered animal species

As per the Frontiers authors guidelines, you are required to use the following format for statements involving human subjects: This study was carried out in accordance with the recommendations of [name of guidelines], [name of committee]. The protocol was approved by the [name of committee]. All subjects gave written informed consent in accordance with the Declaration of Helsinki.
For statements involving animal subjects, please use:
This study was carried out in accordance with the recommendations of 'name of guidelines, name of committee'. The protocol

was approved by the 'name of committee'.

If the study was exempt from one or more of the above requirements, please provide a statement with the reason for the exemption(s).
Ensure that your statement is phrased in a complete way, with clear and concise sentences.

This study was carried out in accordance with the recommendations of the National Statement on Ethical Conduct in Human Research, Human Ethics Office at the University of Western Australia. The protocol was approved by the Human Ethics Office at the University of Western Australia. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Correct me if I'm wrong: Groups outperform individuals in the Climate Stabilization Task

Belinda Xie[1,2*], Mark J. Hurlstone[1] and Iain Walker[3]

[1]*School of Psychological Science, University of Western Australia, Crawley, Western Australia, Australia*
[2]*School of Psychology, University of New South Wales, Sydney, New South Wales, Australia*
[3]*School of Psychology, University of Canberra, Bruce, Australian Capital Territory, Australia*

*** Correspondence:**
Belinda Xie
belinda.xie@unsw.edu.au

Keywords: climate stabilization task, mental models, group decision-making, carbon dioxide accumulation, stock-flow tasks, emissions reduction.

Abstract

Avoiding dangerous climate change requires ambitious emissions reduction. Scientists agree on this, but policy-makers and citizens do not. This discrepancy can be partly attributed to faulty mental models, which cause individuals to misunderstand the carbon dioxide (CO_2) system. For example, in the Climate Stabilization Task (hereafter, 'CST') (Sterman & Booth-Sweeney, 2007), individuals systematically underestimate the emissions reduction required to stabilize atmospheric CO_2 levels, which may lead them to endorse ineffective 'wait-and-see' climate policies. Thus far, interventions to correct faulty mental models in the CST have failed to produce robust improvements in decision-making. Here, in the first study to test a group-based intervention, we found that success rates on the CST markedly increased after participants deliberated with peers in a group discussion. The group discussion served to invalidate the faulty reasoning strategies used by some individual group members, thus increasing the proportion of group members who possessed the correct mental model of the CO_2 system. Our findings suggest that policy-making and public education would benefit from group-based practices.

1 Introduction

To avoid dangerous climate change, average global temperature must not exceed a critical threshold, defined in the Paris Agreement as 1.5-2ºC above pre-industrial levels (UNFCCC, 2015). However, countries' current climate pledges are guaranteed to overshoot this threshold (Mauritsen & Pincus, 2017), indicating that current national emissions policies are grossly inadequate. In a democracy, implementing effective mitigation policy is a two-step challenge: policy-makers must craft appropriate policies and those policies must then receive political and electoral support (Dreyer, Walker, McCoy, & Teisl, 2015). Both steps require policy-makers, politicians, and citizens to understand the CO_2 system. Unfortunately, most individuals lack this knowledge and consequently underestimate the measures required to mitigate climate change (e.g., Guy et al., 2013; Martin, 2008; Sterman & Booth-Sweeney, 2002).

To reason about emissions policy (in the context of mitigating climate change), an individual must understand how CO_2 emissions contribute to climate change. For example, someone who accepts the scientific consensus would: (1) recognize that global temperature is increasing, (2) attribute that increase to human CO_2 emissions, and (3) predict that emitting more CO_2 will further increase temperature. This knowledge structure is called a 'mental model' (Doyle & Ford, 1998; Sterman, 1994). A mental model represents the causal relationships within a system, and is used to describe, explain, and predict system behavior (Doyle & Ford, 1998; Sterman, 1994). Although crucial for decision-making, mental models are constrained by cognitive limits (Doyle & Ford, 1998; Sterman & Booth-Sweeney, 2002) and can never represent the full complexity of the real world. The human decision-maker is thus likely to make imperfect decisions about complex problems.

The Climate Stabilization Task (hereafter, 'CST') represents the complex problem of choosing the appropriate level of climate change mitigation. The CST is a decision-making task in which participants are told that atmospheric CO_2 concentration is increased by CO_2 emissions (largely from human activities), decreased by CO_2 absorption (largely by oceans and plants), and stabilized when the rate of CO_2 emissions equals the rate of CO_2 absorption. Participants are also told that atmospheric CO_2 concentration has increased since the Industrial Revolution, because the rate of CO_2 emissions has increased to double the rate of CO_2 absorption. Participants are then presented with a hypothetical scenario (Fig. 1A) in which atmospheric CO_2 concentration gradually rises to 400ppm, then stabilizes by the year 2100. Next, participants must sketch trajectories of CO_2 emissions and CO_2 absorption that would correspond with this hypothetical scenario (Fig. 1B).

The 'principle of accumulation' states that, at any given time, the level of some accumulating stock (in this case, atmospheric CO_2 concentration) is the difference between its inflow (rate of CO_2 emissions) and outflow (rate of CO_2 absorption) (Martin, 2008). Thus, to stabilize atmospheric CO_2 concentration, the rate of CO_2 emissions must decrease to equal the rate of CO_2 absorption. However, participants often erroneously assert that stabilizing CO_2 emissions is sufficient to stabilize atmospheric CO_2 concentration (Fig. 1C). Known as 'pattern-matching', this occurs when participants ignore CO_2 absorption, believing that the pattern of atmospheric CO_2 concentration should 'match' the pattern of CO_2 emissions. Repeated studies find that only 6-44% of participants answer the CST correctly, with many falling prey to the abovementioned pattern-matching heuristic (Boschetti, Richert, Walker, Price, & Dutra, 2012; Guy et al., 2013; Moxnes & Saysel, 2009; Newell et al., 2015; Sterman & Booth-Sweeney, 2002, 2007). Low success rates on the CST are observed not only for members of the general public (Boschetti et al., 2012), but also for individuals who are a good proxy for policy-makers—namely stakeholders of a project researching climate change impacts (Boschetti et al., 2012) and Masters students studying system dynamics at the Massachusetts Institute of Technology (Sterman & Booth-Sweeney, 2002).

Figure 1 | Graphical illustration of the CST. Participants are presented with the graph in **(A)** showing the increase of atmospheric CO_2 concentration since the year 1900 up until the year 2000. Following 2000, the graph depicts a hypothetical scenario in which atmospheric CO_2 concentration increases to 400ppm before stabilizing by the year 2100. Next, participants are presented with the graph in **(B)** and asked to sketch the trajectories of CO_2 emissions and CO_2 absorption from years 2000 to 2100 that they believe would be consistent with the hypothetical scenario. The graph in **(C)** shows a typical participant response to the CST, where the blue line represents the participant's estimate of CO_2 absorption, and the purple line represents the participant's estimate of CO_2 emissions**.** As the rate of CO_2 emissions exceeds the rate of CO_2 absorption, atmospheric CO_2 concentration will increase, not stabilize. This is an example of the so-called 'pattern-matching' heuristic, whereby the pattern of CO_2 emissions is assumed to 'match' the pattern of atmospheric CO_2 concentration. The correct CO_2 emissions trajectory, given the participant's estimate of CO_2 absorption, is depicted by the dashed yellow line. The rate of CO_2 emissions decreases to equal the rate of CO_2 absorption, an equilibrium that would stabilize atmospheric CO_2 concentration. This response is consistent with the principle of accumulation, which states that the level of a stock at any given time is the difference between its inflow and its outflow.

Most interventions to correct decision-makers' mental models of the CO_2 system—as indexed by responses on the CST—have been unsuccessful (e.g., Pala & Vennix, 2005; Reichert, Cervato, Niederhauser, & Larsen, 2015). Using analogies (e.g., a bathtub in which the water level represents atmospheric CO_2 concentration) (Guy et al., 2013; Moxnes & Saysel, 2009; Newell et al., 2015) and promoting 'global thinking' over 'local thinking' (Fischer &

Gonzalez, 2015; Weinhardt, Hendijani, Harman, Steel, & Gonzalez, 2015) have produced minor improvements in CST performance. A formal university course in system dynamics was more successful (Pala & Vennix, 2005; Sterman, 2010), but this intervention is too resource-intensive to be applied on a large scale. Although these results seem discouraging, all interventions so far share the limitation of characterizing decision-makers as individuals. However, real-world decision-making is a social, group-based process informed by the beliefs of others (Tranter, 2011). Previous research shows that groups are better able than individuals to attenuate cognitive biases and decision heuristics (Kugler, Kausel, & Kocher, 2012; Schulze & Newell, 2016), as well as to identify, evaluate, and resolve competing hypotheses (Larrick, 2016; Trouche, Johansson, Hall, & Mercier, 2015). These benefits notwithstanding, it is important to note that groups do not outperform individuals in all tasks. However, groups do perform consistently better than individuals on intellective, 'truth-wins' problems in which the sole correct answer can be determined through logic, and then explained to convince others (i.e., the truth 'wins') (Davis, 1973; Laughlin, Hatch, Silver, & Boh, 2006). The CST is one such problem, as understanding the principle of accumulation leads to only one demonstrably correct solution (i.e., the rate of emissions equaling the rate of absorption).

The aim of the current study was to test whether an intervention involving group decision-making can improve performance on the CST. To address this question, we administered a computerized version of the CST to staff and students from the University of Western Australia (N = 141). Participants were given background information about the CO_2 system, and then presented with the hypothetical scenario in which atmospheric CO_2 concentration stabilizes by the year 2100 (Fig. 1A). The decision-making component was administered at two time points, T_1 and T_2. At T_1, participants were presented with four graphs (Fig. 2) and asked to select the graph that would produce the hypothetical scenario described. After selecting a graph, participants typed a brief explanation for their decision. At T_2, participants were randomly allocated to one of three experimental conditions, before entering an anonymized online chatroom for 10 minutes. In the individuals condition (N = 21), participants reflected on their initial answer and explanation by themselves before making a decision about the four graphs again. In the dyads (N = 40) and groups (N = 80) conditions, either two or four participants, respectively, inspected each other's initial answers and explanations, then engaged in a discussion to reach a consensus decision on which of the four graphs is correct.

There were three key predictions. Firstly, it was predicted that the individual reflection would have no effect on decision-making, such that the success rates of individuals would not increase from T_1 to T_2. Secondly, it was predicted that the dyad discussion would benefit decision-making, such that the success rates of dyads would increase from T_1 to T_2. Thirdly, it was predicted that the group discussion would benefit decision-making to a greater extent than the dyad discussion, such that the success rates of groups would increase from T_1 to T_2, and this increase would be greater than that observed for dyads.

A secondary aim of the current study was to examine whether individual performance on the CST can be explained by a person's (1) demographic characteristics, (2) climate change knowledge and attitudes, and/or (3) personality and cognitive style. These constructs influence performance on comparable tasks that tap similar reasoning skills, but their effects on CST performance are unclear. Participants therefore completed a pre-test questionnaire assessing several individual differences measures. This was an exploratory feature of the current study and accordingly we made no specific predictions about the relationships between the following variables and CST performance.

Most studies find no relationship between demographic variables and performance on tasks similar to the CST (e.g., Moxnes & Saysel, 2004; Sterman, 2008). However, other studies find that younger participants (Browne & Compston, 2015), males (Browne & Compston, 2015; Ossimitz, 2002; Reichert et al., 2015), or students studying STEM degrees (Booth-Sweeney & Sterman, 2000; Browne & Compston, 2015) perform better than older participants, females, or students studying non-STEM degrees. Age, sex, and field of education were therefore included in the questionnaire.

The questionnaire included a measure of climate change knowledge, as task-specific knowledge is associated with better performance on some stock-flow tasks (Strohhecker & Größler, 2013), but appears unrelated to performance on the CST (Moxnes & Saysel, 2004). A measure of climate change attitudes was also included to rule out the possibility that participants choose a graph based on their own ideology, rather than stock-flow reasoning. For the same reason, a measure of 'environmental worldview', or one's beliefs about humanity's relationship with nature (Price, Walker, & Boschetti, 2014), was also included.

Two personality variables are related to the ability to overcome bias by prior belief (Homan et al., 2008; West, Toplak, & Stanovich, 2008), and may therefore benefit performance on the CST. 'Active open-mindedness' describes an individual's tendency to spend sufficient time on a problem before giving up, and to consider new evidence and the beliefs of others (Haran, Ritov, & Mellers, 2013). 'Need for cognition' is the psychological need to structure the world in meaningful and integrated ways, and is associated with expending greater mental effort and enjoying analytical activity (Cacioppo & Petty, 1982).

Lastly, three aspects of cognitive style may be relevant to task performance. 'Cognitive reflection' is the ability to resist reporting the first answer that comes to mind (Frederick, 2005), and may therefore protect against pattern-matching. 'Global processing' is a way of perceiving the world that favors the organized whole, whereas 'local processing' favors component parts and details (Weinhardt et al., 2015). Previous studies have produced conflicting results on the relationship between processing style and stock-flow reasoning (Fischer & Gonzalez, 2015; Weinhardt et al., 2015). Systems thinking refers to the tendency to understand phenomena as emerging from complex, dynamic, and nested systems (Thibodeau, Frantz, & Stroink, 2016). It is positively related to the ability to comprehend causal complexity and dynamic relationships (Thibodeau et al., 2016), as well as pro-environmental attitudes (Davis & Stroink, 2016; Lezak & Thibodeau, 2016).

A third and final aim relates to Sterman's (2008) contention that the widespread, global preference for 'wait-and-see' or 'go-slow' approaches to emissions reduction can be linked to misunderstanding the complex CO_2 system. We therefore included a policy preference question in the pre-test questionnaire, which was subsequently repeated at post-test, after completion of the CST. Participants answered the question, "Which of these comes closest to your view on how we should address climate change?", with one of three options: 'wait-and-see' (wait until we are sure that climate change is really a problem before taking significant economic action), 'go-slow' (we should take low-cost action as climate change effects will be gradual), or 'act-now' (climate change is a serious and pressing problem that requires significant action now). If poor understanding of the climate system is indeed responsible for complacent attitudes towards emissions reduction, then we expect participants who answer the CST incorrectly to be more likely to prefer 'wait-and-see' or 'go-slow' policies at post-test. Conversely, those who answer the CST correctly should be more likely to select the 'act-now' option.

2 Method

Ethical approval to conduct the experiment was granted by the Human Ethics Office at the University of Western Australia (UWA) (RA/4/1/6298).

2.1 Participants

One hundred and forty one members of the campus community at the UWA were recruited to take part in the experiment using the Online Recruitment System for Economic Experiments (ORSEE; Greiner, 2015), an open-source web-based recruitment platform for running decision-making experiments. The ORSEE database at UWA contains a pool of over 1,500 staff and students from a range of academic disciplines. Participants were recruited by issuing electronic invitations to randomly selected individuals in the ORSEE database to attend one of several advertised experimental sessions. Participants' ages ranged from 17 to 74 (*Mdn* = 21.00, *M* = 23.80, *SD* = 7.34) and just over two thirds of participants were female (69.5%). About half studied a degree-specific major under the Faculty of Science (54.5%), but the Business School (15.7%), Engineering, Computing, and Mathematics (14.2%), and Arts (13.4%) faculties were also well represented. Participants were paid $10AUD for attending the experiment.

2.2 Design

The experiment manipulated two independent variables: group size (individuals [$n = 1$] vs. dyads [$n = 2$] vs. groups [$n = 4$]) and time (T_1 vs. T_2). Group size was a between-participants variable, whereas time was a within-participants variable. Participants were allocated to the different group size conditions in a quasi-random fashion (see below). There was a minimum of 20 cases per group size: $21 \times 1 = 21$ participants in the individuals condition; $20 \times 2 = 40$ participants in the dyads condition; and $20 \times 4 = 80$ participants in the groups condition.

2.3 Apparatus, materials, and procedure

The experiment was conducted between May and August 2016 in the Behavioral Economics Laboratory at UWA (http://bel-uwa.github.io), a computerized laboratory designed for carrying out collective decision-making experiments. There were 27 experimental sessions in total, with a minimum of two and a maximum of eight participants per session. Group sizes were randomly pre-determined before each session but were subject to change in the event that some participants failed to attend. For example, if eight participants were invited to a session, and the goal was to run two groups of four participants, but only six participants attended, then four participants were allocated to the group condition, and two participants were allocated at random either to the dyads condition or the individuals condition.

As participants arrived to each experimental session, they were randomly seated at a workstation containing two computer terminals. This random seating allocation in turn determined the group size condition to which the participants were allocated. The workstations were separated from each other by privacy blinds to prevent participants from observing one another's responses, and participants knew that face-to-face communication was prohibited. Participants read an information sheet and provided written informed consent, after which the experimenter provided an overview of the structure of the session. Using the left computer terminal on their workstation, the participants then completed the individual differences questionnaire (see Supplementary Materials, Section 1), which was executed on an internet browser using Qualtrics survey software. The questionnaire took approximately 20 minutes to complete.

Once all participants had completed the questionnaire, they received verbal instructions from the experimenter to minimize their internet browser, which revealed the electronic instructions for the CST (see Supplementary Materials, Section 2). The first page foreshadowed what the task would involve. The second page defined CO_2, CO_2 emissions, atmospheric CO_2 concentration, and CO_2 absorption. The third page described how CO_2 emissions and CO_2 absorption, respectively, increase and decrease atmospheric CO_2 concentration, and why atmospheric CO_2 concentration has increased since the Industrial Revolution. The final page presented the decision-making situation. It described a hypothetical scenario in which atmospheric concentration rises from its current level of 400ppm to stabilize at 420ppm by the year 2100. Participants were then confronted with four graphs depicting the same trajectory of CO_2 absorption, but different trajectories of CO_2 emissions (Fig. 2), and were required to choose the graph that would give rise to the hypothetical scenario. Graph D (Fig. 2D) is the correct response, as it is the only graph that depicts the rate of CO_2 emissions decreasing to equal the rate of CO_2 absorption. We used a multiple-choice format because it is less cognitively-taxing than the version of the CST in which participants sketch trajectories, but it nevertheless produces equivalent results (Sterman & Booth-Sweeney, 2007).

Figure 2 | The four response alternatives in the multiple-choice version of the CST. All graphs show the same CO_2 absorption trajectory with a different CO_2 emissions trajectory. Graph **(A)** depicts the typical pattern-matching response in which CO_2 emissions rise and then stabilize. Graph **(B)** is a less obvious form of pattern-matching in which CO_2 emissions immediately stabilize. Graph **(C)** approximates the correct answer as CO_2 emissions decrease, but not to the level required to achieve stabilization. Graph **(D)** is the correct response, because CO_2 emissions decrease to equal CO_2 absorption, thus stabilizing atmospheric CO_2.

The decision-making component of the CST was executed as a z-Tree (Fischbacher, 2007) program, which was administered on the right computer terminal of each participant's workstation. The CST required a decision at two different points in time: T_1 and T_2. At T_1, all participants completed the task individually, irrespective of the group size condition to which they had been allocated. The experimental procedure at this time point was therefore identical across all three group size conditions. Participants first read the electronic instructions on the left computer terminal, before indicating on the right computer terminal which of the four graphs they believed would stabilize atmospheric CO_2 concentration (Fig. 2). There was no time limit for this component of the task.

Once participants had registered their T_1 graph choice, a text field appeared on screen with a prompt to use the keyboard to type out a brief explanation for why they thought that the graph they had chosen was correct. Participants were allocated 5 minutes to complete this component of the task and a counter in the top right-hand corner of the terminal display indicated the time remaining for participants to supply their written explanations.

At T_2, participants were allocated to their group size conditions (participants were naïve to this second component of the task until they reached this stage). Participants assigned to the dyads or groups conditions were required to discuss the decision problem with their one partner or three group members, respectively, for a fixed period of ten minutes in order to reach a consensus decision regarding the correct solution. They were first given six guidelines for a productive group discussion (adapted from a study by Schweiger, Sandberg and Ragan, 1986), as shown in Fig. 3A. They then entered an online chatroom in which they could communicate with one another. The chatroom interface was divided into two panels: the player decisions panel and the communication panel (Fig. 4). The player decisions panel, to the left of the terminal display, presented the T_1 graph choices and explanations of each group or dyad member under a pseudonym (Leda, Triton, Portia, or Sinope) to preserve participant anonymity. In the communication panel, to the right of the terminal display, dyad and group members could communicate with one another by typing messages into a text entry field. These messages were posted in the communication panel under the group or dyad member's designated pseudonym. A timer in the top right corner of the terminal display showed how much time remained. After ten minutes had elapsed, one group or dyad member was chosen randomly by the computer to register the group's or dyad's consensus decision.

Figure 3 | The instructions given to participants in the groups and dyads conditions (A) and individuals condition (B) at T_2.

Figure 4 | Graphical illustration of the chatroom communication interface.

The procedure at T_2 was different in the individuals condition. Participants in this condition were instructed to reflect on their T_1 decision for ten minutes, alone. They were instructed to approach this reflection with a skeptical mind, to question their original assumptions, and to consider alternative explanations (Fig. 3B). The chatroom interface was once again divided into two panels, this time labelled the decision panel and the reflections panel. In the decision panel, to the left of the terminal display, participants could inspect their T_1 decision and explanation. In the reflections panel, to the right of the terminal display, participants were able to record reflections on their T_1 decision. This panel was essentially the same as the communication panel for participants in the groups and dyads conditions, except that it was used to record reflections, rather than to communicate with group or dyad members. A timer in the top right corner of the terminal display once again indicated how much time remained. After ten minutes had elapsed, participants were required to indicate once again which graph they deemed to be correct.

After submitting the T_2 decision, all participants completed a post-test questionnaire. Participants in the dyads or groups conditions were asked to choose one of the four CST graphs again, in response to the question; “If the group answer is not what you would have chosen, which answer would you have chosen?”. The post-test questionnaire also contained the climate change knowledge and attitudes questions asked in the individual differences questionnaire at the beginning of the experiment.

The CST took approximately 30 minutes to complete, and the entire experimental session lasted approximately 60 minutes.

3 Results

3.1 Time 1

The success rates at T_1 (blue bars; Fig. 5) did not differ significantly across the three conditions (χ^2 (2) = 1.72, p = .42, two-sided), and the overall success rate was 44%. This is consistent with the highest previously-reported success rate using the CST (Sterman & Booth-Sweeney, 2002). Table 1 shows the frequency with which participants used various reasoning strategies at T_1 to justify their graph choice. For example, Graph D was frequently accompanied by an explanation correctly describing mass balance principles (88.7% of Graph D responses). Although other strategies were referenced by participants who selected Graph D, every other reasoning strategy was more frequently used to justify an incorrect graph.

Table 1 | The frequency (%) with which different reasoning strategies were adopted, as a function of graph A, B, C, and D choices at T_1

Reasoning strategy and coding criteria	Example participant explanation	Graph A (n = 37)	Graph B (n = 12)	Graph C (n = 30)	Graph D (n = 62)	% of total
Mass Balance (Correct) Description indicating awareness of relationship between emissions and absorption flows and the stock of atmospheric CO_2; terms such as mass balance, accumulation, rate of change.	"It's a mass balance and rates of change situation. For the CO_2 concentration in the atmosphere to stabilise, you need the rates of emission and absorption to equal." "If emissions are greater than the absorption, the amount of CO_2 will increase. If the absorption is greater than the emissions, the amount of CO_2 will decrease."	0.0	0.0	10.0	88.7	41.1
Mass Balance (Incorrect) Description indicating awareness of relationship between emissions and absorption flows and the stock of atmospheric CO_2—but misunderstanding the nature of these relationships.	"In figure D, emissions ended up being the same as absorption, which causes the concentration reducing to 0." "If we keep the same difference between the rate of emission and absorption, the concentration will be the same."	32.4	66.7	56.7	4.8	28.4
Pattern-Matching Description mentioning correlations or similarity of behaviour or patterns among emissions and atmospheric CO_2; indication that emissions should be proportional to changes in atmospheric CO_2.	"As I understand it, there is a direct relationship between CO_2 emissions and the atmospheric concentration…" "If atmospheric concentration increases, that means that CO_2 emissions will also increase." "[Graph A] because it rises and then stabilises"	59.5	33.3	10.0	1.6	21.3
Mathematical Reasoning Using algebraic equations, calculating ratios, or quantifying the absolute values of atmospheric concentration, emissions, and/or absorption.	"To achieve the quantity of 420ppm, should have an increase of 20ppm. The emission should be only 20% higher than the absorption." "The rate of ppm increase from 1990 to 2025 = 120ppm/125yrs = 0.96ppm/yr. The ratio of GtC to ppm is around 4GtC = 1ppm…"	21.6	50.0	16.7	8.1	17.0

Reasonableness of Trajectories						
Indicates belief that the correct trajectory should reflect business-as-usual or personal predictions about future emissions/absorption rates.	"…it would be too idealistic to imply that the change would be immediate and the decline would be as drastic as depicted in options B, C, and D" "With current pressures on countries by the UNFCCC for setting emission reduction targets, countries will take drastic measures to reduce their carbon emissions."	21.6	33.3	20.0	6.5	15.6

Reasoning strategy was inferred from participants' T_1 post-decision explanations. The first column lists the possible reasoning strategies, and the coding criteria for those strategies, whilst the second column gives example participant explanations that conform to each strategy. The third through sixth columns show the percentage of participants who chose graph A, B, C, or D, respectively, and subsequently referenced each reasoning strategy. As any one explanation could conform to more than one reasoning strategy, the values in columns three through six may sum to greater than 100%. The last column shows the percentage of total responses that referenced each reasoning strategy. Again, this sums to more than 100% because multiple responses were possible. For inter-rater reliability information, see Supplementary Materials, Section 3.

Figure 5 | Percentage of correct (Graph D) responses to the CST as a function of time (T_1 vs. T_2) and condition (individuals vs. dyads vs. groups). The bars for T_1 (and T_2 for the individuals condition) represent the percentage of correct individual responses, whereas the bars for the dyads and groups conditions at T_2 represent the proportion of correct dyad and group consensus decisions, respectively. Error bars represent standard errors.

The full coding scheme consisted of five strategies from Sterman and Booth-Sweeney (2007) and their associated coding criteria, plus four additional categories created post-hoc to capture reasoning strategies that did not conform to any previously-defined category. The strategies taken from Sterman and Booth-Sweeney (2007) were: pattern-matching, mass balance, technology, sink saturation, and CO_2 fertilisation (for details see Sterman and Booth-Sweeney, 2007; Table 7). Two additional categories defined in their coding scheme (energy balance and inertia/delays) were not used by any of our participants, and therefore were not included here. Two of the new categories were simply the reverse of the categories identified by Sterman and Booth-Sweeney (2007): mass balance—incorrect (incorrect understandings of mass balance) and technology—reverse (technology will increase emissions, rather than enable emissions reduction). The final two categories, mathematical reasoning and reasonableness of trajectories, were created by the authors on the basis of an analysis of participants' responses. In this paper, we only report on the five most popular strategies (technology, sink saturation, CO_2 fertilisation, and technology—reverse were used by less than 3% of total participants and were therefore excluded from the current analysis).

The literature tends to attribute incorrect answers on the CST to the 'pattern-matching' heuristic (e.g., Cronin et al., 2009; Sterman, 2008; Sterman & Booth-Sweeney, 2007). Pattern-matching was indeed the most popular reasoning strategy for participants who selected the typical pattern-matching graph of Graph A (Fig. 2A). However, pattern-matching was not the most popular incorrect reasoning strategy overall. As shown in the last column of Table 1, across all responses, 'incorrect mass balance' principles were applied more frequently than pattern-matching. 'Mathematical reasoning' and 'reasonableness of trajectories' were also common, especially for participants who chose Graph B. The popularity of these strategies suggests that errors on the CST are not exclusively caused by rash, heuristic decisions (i.e., pattern-matching)—even participants who used deliberate and effortful approaches (e.g., unsuccessfully trying to relate CO_2 emissions with CO_2 absorption, or calculating ratios) failed to reach the correct answer.

3.2 Time 2

There was more heterogeneity in success rates across conditions at T_2 (orange bars; Fig. 5), and the overall success rate of 59% was marginally higher than at T_1. The success rate for dyads (65.0%) did not differ significantly from that for individuals (38.1%) (χ^2 (1) = 2.97, p = .121, two-sided) or groups (75.0%) (χ^2 (1) = .48, p = .731, two-sided). However, the success rate for groups was significantly higher than for individuals (χ^2 = 5.67, p = .028, two-sided). A more diagnostic set of comparisons involves contrasting the difference in success rates between T_1 and T_2 for each condition, separately. For individuals, there was no change in success rates over time (p = 1.00, McNemar, two-sided). For dyads, the success rate of dyad consensus decisions at T_2 (65.0%) was numerically, but not significantly, higher than that of individual dyad member decisions at T_1 (52.5%) (p = .125, McNemar, two-sided). Recall that after dyads submitted their consensus decision, individual dyad members were prompted for the answer they would have chosen, regardless of what the consensus decision was. Again, there was no difference in success rates between answers given by individual dyad members at T_1 and then at this post-test stage (both 52.5%) (p = 1.00, McNemar, two-sided). For groups, the success rate of group consensus decisions at T_2 (75.0%) was significantly higher than that of individual group member decisions at T_1 (41.3%) (p < .001, McNemar, two-sided). Furthermore, the success rate of individual group members' post-test answers (61.3%) was significantly higher than the success rate of individual group members at T_1 (41.3%) (p = .002, McNemar, two-sided). Thus, the group discussion reliably improved CST success rates, while the individual reflection and dyad discussion did not.

Analyzing the content of T_2 reflections and discussions (in the same way as analyzing the content of T_1 explanations) sheds light on how groups derived their decision-making advantage over individuals. Despite explicit instructions to "approach this with a skeptical mind", "question your own assumptions", and "consider alternative arguments", 80% of participants in the individuals condition did not type anything during the ten-minute reflection period. Of 21 individuals, only one typed an alternative argument, and only three changed their answers at T_2. Individuals failed to self-reflect, thus preventing them from recognizing their answer was incorrect, or considering why a different answer may be correct. This is consistent with previous research characterizing individuals as "cognitively lazy" decision-makers who rarely challenge an answer that 'feels' right (Larrick, 2016; Trouche et al., 2015).

By contrast, all groups entertained at least two reasoning strategies in their group discussions (except one group in which all group members selected Graph D at T_1). Group discussions contained a mean of 2.30 different reasoning strategies, compared to 1.15 for dyad discussions, and 0.24 for individual reflections (χ^2 (2) = 33.48, p = .000, two-sided). Furthermore, we have tentative evidence that group members helped correct other members' faulty reasoning strategies. For example, in one group, one participant's misunderstanding of mass balance principles ("If emission rate gets close to absorption, concentration will decrease below 400[ppm]") was corrected by two other participants who explained the principle of accumulation ("…but when emissions is greater than absorption, then concentration will increase"). Groups were more likely than individuals (χ^2 (1) = 23.89, p = .000, two-sided) and dyads (χ^2 (1) = 8.29, p = .010, two-sided) to refer to the correct reasoning strategy of mass balance (even if no group member had referenced mass balance in their T_1 explanation). This supports previous findings showing that exposure to diverse perspectives motivates group members to critically evaluate all arguments (Trouche et al., 2015), thus increasing the likelihood that the correct decision will be discussed and judged to be correct (Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006).

The data seem to speak against—but do not rule out—the alternative explanation that groups only benefited from the increased probability of having one member who knew the correct answer. Eight individuals, 15 dyads, and 15 groups contained at least one member who chose Graph D at T_1. If the effect was due merely to the presence of a correct member, we would expect equal performance between dyads and groups, and also for dyads to outperform individuals at T_2 (which they did not, χ^2 (1) = 2.97, p = .121, two-sided). Furthermore, two groups gave the correct consensus decision at T_2, despite having no members who gave the correct decision at T_1 (an example of 'process gain', in which interpersonal interaction between multiple individuals yields an outcome better than that of any single individual, or even the sum of all individuals; Hackman & Morris, 1975).

We were also able to rule out effects of individual differences. Binary logistic regression analyses were conducted to determine whether the various individual difference variables measured at pre-test subsequently predicted performance on the CST at T_1. To satisfy the assumption of a dichotomous dependent variable, CST performance was coded as either incorrect (selecting Graphs A, B, or C) or correct (selecting Graph D). Age, actively open-minded thinking, and supporting the policy of 'government regulation of CO_2 as a pollutant' were the only significant independent predictors (p < .05). These variables were subsequently combined into a set of predictors and subjected to a further binary logistic regression analysis. The full model was statistically significant compared to the constant-only model, χ^2 (3) = 21.47, p < .001, Nagelkerke's R^2 = .19. Prediction accuracy was 61.9% (50.8% for correct responses, 70.5% for incorrect responses). Thus, the full model with these three predictors was barely above chance at predicting correct answers. This poor predictive performance is noteworthy in revealing that performance on the CST is largely immune to the influence of demographic, attitudinal, personality, and cognitive variables.

Recall that the individual difference questions about climate change knowledge and attitudes were presented again at the post-test phase. Although CST performance was not significantly predictive of answers to any of these items, in light of the third aim of our study it is worth discussing the answers to the policy preference question, “Which of these comes closest to your view on how we should address climate change?”. At pre-test, 69.1% of participants answered ‘act-now’, 30.9% answered ‘go-slow’, and no participant selected the ‘wait-and-see’ option. Excluding the wait-and-see option, answering the CST correctly did not significantly predict post-test responses ($\chi^2(3) = 7.19$, $p = .066$, two-sided). However, there was an increase in the percentage of individuals who selected ‘act-now’ from pre-test to post-test across all conditions (overall, 80.9% ‘act-now’ at post-test).

4 Discussion

Repeated studies employing the CST reveal that individuals systematically underestimate the emissions reduction required to stabilize atmospheric CO_2 levels. So far, interventions to increase success rates on the CST have been individual-focused and largely ineffective. We sought to examine whether group reasoning could increase success rates on this task. It was predicted that success rates from T_1 to T_2 would (1) not increase for individuals, (2) increase for dyads, and (3) increase for groups to a greater extent than for dyads. The first and third predictions were statistically supported, but there was only qualitative support for the second prediction. The ten-minute group discussions among four participants significantly improved success rates, whereas ten minutes of individual reflection or dyad discussion did not. By analyzing individual justifications at T_1, and reflections and discussions at T_2, we found that groups benefited from exposure to multiple perspectives and the opportunity to communicate, which facilitated the falsification of incorrect reasoning strategies. We also found that incorrect reasoning strategies were numerous, and not limited to the oft-reported pattern-matching strategy. Lastly, we ruled out two alternative explanations—groups did not improve merely due to a size advantage in the number of members who knew the correct response, nor were individual differences in demographics, climate change knowledge, personality, or cognitive style responsible for any given individual’s CST success.

There are some potential limitations of the current study that merit consideration. Firstly, with a minimum of twenty cases in each condition as units for the statistical analysis, our experiment may have had insufficient statistical power to detect a significant improvement in success rates for dyads from T_1 to T_2. Using a larger sample size may reveal that the numerical, yet non-significant, increase in success rates observed with our dyad members reflects some real benefit of the dyad discussion. Secondly, the group advantage observed in the current study was obtained using a multiple-choice response format, which differs from the conventional CST procedure in which participants must sketch trajectories of emissions and absorption. It therefore remains open whether the results reported here would generalize to an experimental scenario employing this more complicated response format. It is possible, for example, that the uncharacteristically low rates of susceptibility to the pattern matching heuristic observed in the current study are an artifact of our unorthodox response format. Thus, we may expect higher initial rates of pattern-matching when returning to the original CST procedure, but it remains to be seen whether this would eliminate or attenuate the group advantage witnessed here. Thirdly, our intervention at T_2 in the dyad and group conditions afforded more than merely the opportunity for individuals to communicate with one another—participants were also afforded the chance to read the T_1 explanations of their dyad partner or group members. Although it is our conviction that the opportunity to engage in communication was instrumental to the group decision making advantage, we cannot preclude the possibility that mere exposure to the alternative perspectives of others may confer an advantage in itself, compared to individual reasoning alone. A group condition in which participants are exposed to their group members' T_1 decision explanations—without engaging in any subsequent discussion—would reveal whether mere exposure to multiple perspectives can produce a group benefit. Finally, we could not provide strong evidence for Sterman's (2008) argument that policy-makers' and citizens' deficient mental models of the climate system are responsible for complacent attitudes towards emissions reduction. Answers to the CST did not predict subsequent policy preferences. However, all participants in our sample believed that we must take action against climate change ('act-now' or 'go-slow'), even before completing the CST. We were therefore unable to test the hypothesis that participants who answer the CST incorrectly also deny the need for emissions reduction ('wait-and-see'). However, the overall increase in 'act-now' responses, relative to 'go-slow' responses, from pre-test to post-test implies some diffuse benefit of merely completing the CST. Future studies employing a sample with more heterogeneous pre-existing policy beliefs will provide a stronger test for the hypothesized link between accurate mental models of the CO_2 system and support for urgent emissions reduction.

Given the abovementioned concerns about the generality of our results, one direction for future work is to determine whether and how our findings generalise to other stock-flow tasks, especially the original CST procedure, in which participants sketch trajectories of emissions and absorption by themselves. Furthermore, although we have shown a benefit on CST performance of reasoning in groups of four members, an additional avenue for future work will be to examine whether this advantage extends to larger groups. On the one hand, we might expect that increasing the number of group members will improve CST performance, because of an increase in information-processing capacity and diversity of perspectives (Charness & Sutter, 2012; Cohen & Thompson, 2011). On the other hand, we might expect that as the group size reaches some critical point, CST performance will begin to decline as the abovementioned benefits of group decision making will be outweighed by the costs of coordinating opinions and resolving disputes within the group (Lejarraga, Lejarraga, & Gonzalez, 2014; Orlitzky & Hirokawa, 2001). Identifying the optimal group size for solving the CST will permit more robust recommendations about how group-based practices should be incorporated into the decision-making process.

In closing, we note that the final success rate of group consensus decisions at T_2 (75%) is considerably higher than the success rates previously reported with the CST. The success of our group-based intervention suggests that group-based decision-making may help facilitate the two-step implementation of effective emissions policy. First, crafting appropriate mitigation policy requires comprehensive and accurate decision-making, and our results suggest small groups are best-suited for this task. Second, rallying political and electoral support for such policy requires a well-informed population that comprehends the scale of the emissions problem. The general public presently endorses high levels of belief in anthropogenic climate change, but low levels of concern and urgency about climate change mitigation (Akter & Bennett, 2011; Reser, Bradley, Glendon, Ellul, & Callaghan, 2012). In order to bridge this gap between what the public believes about the climate change problem (that it is real and caused by human activities), and the solutions they are willing to support (immediate and significant emissions reduction), their mental models must be changed. Group-based programs, whether informal conversations about climate change or formal public education initiatives, could establish the correct mental model and help mobilize support for effective mitigation policy.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

All authors contributed to the conception and design of the study; MH created the software for running the experiment and produced the figures for the paper; BX conducted the experimental sessions, performed the data analysis, and wrote the paper; MH and IW edited the paper and reviewed the results.

7 Funding

The work reported here was supported by a grant from the Climate Adaptation Flagship of the Commonwealth Scientific and Industrial Research Organisation (CSIRO) of Australia awarded to Mark Hurlstone and Iain Walker.

8 References

Akter, S., & Bennett, J. (2011). Household perceptions of climate change and preferences for mitigation action: the case of the Carbon Pollution Reduction Scheme in Australia. *Climatic Change, 109*(3-4), 417-436. doi:10.1007/s10584-011-0034-8

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

Boschetti, F., Richert, C., Walker, I., Price, J., & Dutra, L. (2012). Assessing attitudes and cognitive styles of stakeholders in environmental projects involving computer modelling. *Ecological Modelling, 247*, 98-111. doi:10.1016/j.ecolmodel.2012.07.027

Browne, C., & Compston, P. (2015). *Rethinking intuition of accumulation*. Paper presented at the Reinventing Life on a Shrinking Earth: Proceedings of the 33rd International Conference of the System Dynamics Society, Cambridge, USA. Retrieved from http://www.systemdynamics.org/conferences/2015/proceed/papers/P1091.pdf

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology, 42*(1), 116-131. Retrieved from PsycArticles.

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *The Journal of Economic Perspectives, 26*(3), 157-176. doi:10.1257/jep.26.3.157

Cohen, T. R., & Thompson, L. (2011). When are Teams an Asset in Negotiations and When are They a Liability? In E. Mannix, A. Neale, & J. Overbeck (Eds.), *Research on managing groups and teams: Negotiation and groups* (Vol. 14, pp. 3-34). Bingley, UK: Emerald Group Publishing.

Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes, 108*(1), 116-130. doi:10.1016/j.obhdp.2008.03.003

Davis, A. C., & Stroink, M. L. (2016). The Relationship between Systems Thinking and the New Ecological Paradigm. *Systems Research and Behavioral Science*, *33*, 575-586. doi:10.1002/sres.2371

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. Retrieved from PsycArticles.

Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for system dynamics research. *System Dynamics Review, 14*(1), 3-29. doi:10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K

Dreyer, S. J., Walker, I., McCoy, S. K., & Teisl, M. F. (2015). Australians' views on carbon pricing before and after the 2013 federal election. *Nature Climate Change, 5*(12), 1064-1067. doi:10.1038/nclimate2756

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental economics, 10*(2), 171-178. doi:10.1007/s10683-006-9159-4

Fischer, H., Degen, C., & Funke, J. (2015). Improving stock-flow reasoning with verbal formats. *Simulation & Gaming, 46*(3-4), 255-269. doi:10.1177/1046878114565058

Fischer, H., & Gonzalez, C. (2015). Making sense of dynamic systems: how our understanding of stocks and flows depends on a global perspective. *Cognitive science, 40*(2), 1-17. doi:10.1111/cogs.12239

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*(1), 114-125. doi:10.1007/s40881-015-0004-4

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

Hackman, J. R., & Morris, C. G. (1975). Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 8, pp. 45-99). New York, NY: Academic Press, Inc.

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making, 8*(3), 188-201. Retrieved from ProQuest.

Homan, A. C., Hollenbeck, J. R., Humphrey, S. E., Van Knippenberg, D., Ilgen, D. R., & Van Kleef, G. A. (2008). Facing differences with an open mind: Openness to experience, salience of intragroup differences, and performance of diverse work groups. *Academy of management Journal, 51*(6), 1204-1222. doi:10.5465/AMJ.2008.35732995

Korzilius, H., Raaijmakers, S., Rouwette, E., & Vennix, J. (2014). Thinking Aloud While Solving a Stock-Flow Task: Surfacing the Correlation Heuristic and Other Reasoning Patterns. *Systems Research and Behavioral Science, 31*(2), 268-279. doi:10.1002/sres.2196

Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science, 3*(4), 471-482. Retrieved from http://www.econstor.eu/bitstream/10419/55339/1/685229548.pdf

Larrick, R. P. (2016). The Social Context of Decisions. *Annual Review of Organizational Psychology and Organizational Behavior, 3*(1), 441-467. doi:10.1146/annurev-orgpsych-041015-062445

Laughlin, P. R., Hatch, E. C., Silver, J. S., & Boh, L. (2006). Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. *Journal of Personality and Social psychology, 90*(4), 644-651. doi:10.1037/0022-3514.90.4.644

Lejarraga, T., Lejarraga, J., & Gonzalez, C. (2014). Decisions from experience: How groups and individuals adapt to change. *Memory & Cognition, 42*(8), 1384-1397. doi:10.3758/s13421-014-0445-7

Lezak, S. B., & Thibodeau, P. H. (2016). Systems thinking and environmental concern. *Journal of Environmental Psychology, 46*. doi:10.1016/j.jenvp.2016.04.005

Martin, J. F. (2008). *Improving understanding of climate change dynamics using interactive simulations.* (Master of Science in Technology and Policy), Massachusetts Institute of Technology, Cambridge, MA. Retrieved from http://dspace.mit.edu/bitstream/handle/1721.1/43742/263169453-MIT.pdf?sequence=2

Mauritsen, T., & Pincus, R. (2017). Committed warming inferred from observations. *Nature Climate Change, 7*, 652-655. doi:10.1038/nclimate3357

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C., Frieler, K., Knutti, R., . . . Allen, M. R. (2009). Greenhouse-gas emission targets for limiting global warming to 2 C. *Nature, 458*(7242), 1158-1162. doi:10.1038/nature08017

Moxnes, E., & Saysel, A. K. (2004). Misperceptions of global climate change: Information policies. *Working Papers in System Dynamics, 1*(4), 1-51. Retrieved from http://bora.uib.no/bitstream/handle/1956/1981/WPSD1.04Climate.pdf?sequence=1

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

Newell, B. R., Kary, A., Moore, C., & Gonzalez, C. (2015). Managing the Budget: Stock-Flow Reasoning and the CO2 Accumulation Problem. *Topics in cognitive science, 8*(1), 138-159. doi:10.1111/tops.12176

Orlitzky, M., & Hirokawa, R. Y. (2001). To Err is Human, to Correct for it Divine: A Meta-Analysis of Research Testing the Functional Theory of Group Decision-Making Effectiveness. *Small Group Research, 32*(3), 313-341. doi:10.1177/104649640103200303

Ossimitz, G. (2002). *Stock-flow-thinking and reading stock-flow-related graphs: An empirical investigation in dynamic thinking abilities*. Paper presented at the Proceedings of the 2002 International System Dynamics Conference.

Pala, Ö., & Vennix, J. A. (2005). Effect of system dynamics education on systems thinking inventory task performance. *System Dynamics Review, 21*(2), 147-172. doi:10.1002/sdr.310

Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quéré, C., . . . Wilson, C. (2013). The challenge to keep global warming below 2 °C. *Nature Climate Change, 3*, 4-6. doi:10.1038/nclimate1783

Price, J. C., Walker, I. A., & Boschetti, F. (2014). Measuring cultural values and beliefs about environment to identify their role in climate change responses. *Journal of Environmental Psychology, 37*, 8-20. doi:10.1016/j.jenvp.2013.10.001

Reichert, C., Cervato, C., Niederhauser, D., & Larsen, M. D. (2015). Understanding Atmospheric Carbon Budgets: Teaching Students Conservation of Mass. *Journal of geoscience education, 63*(3), 222-232. doi:10.5408/14-055.1

Reser, J. P., Bradley, G. L., Glendon, A. I., Ellul, M., & Callaghan, R. (2012). *Public risk perceptions, understandings and responses to climate change in Australia and Great Britain*. Retrieved from Gold Coast, Australia: https://terranova.org.au/repository/nccarf/public-risk-perceptions-understandings-and-responses-to-climate-change-in-australia-and-great-britain-final-report/reser_2012_public_risk_perceptions_final.pdf

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social psychology, 91*(6), 1080-1093. doi:10.1037/0022-3514.91.6.1080

Schulze, C., & Newell, B. R. (2016). More heads choose better than one: Group decision making can eliminate probability matching. *Psychonomic Bulletin & Review, 23*(3), 907-914. doi:10.3758/s13423-015-0949-6

Schweiger, D. M., Sandberg, W. R., & Ragan, J. W. (1986). Group approaches for improving strategic decision making: A comparative analysis of dialectical inquiry, devil's advocacy, and consensus. *Academy of Management Journal, 29*(1), 51-71. Retrieved from http://www.jstor.org.ezproxy.library.uwa.edu.au/stable/255859

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

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

Sterman, J. D. (2008). Risk Communication on Climate: Mental Models and Mass Balance. *Science, 322*(5901), 532-533. doi:10.1126/science.1162574

Sterman, J. D. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review, 26*(4), 316-334. doi:10.1002/sdr

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

Strohhecker, J., & Größler, A. (2013). Do personal traits influence inventory management performance?—The case of intelligence, personality, interest and knowledge. *International Journal of Production Economics, 142*(1), 37-50. doi:10.1016/j.ijpe.2012.08.005

Thibodeau, P. H., Frantz, C. M., & Stroink, M. L. (2016). Situating a Measure of Systems Thinking in a Landscape of Psychological Constructs. *Systems Research and Behavioral Science, 33*, 753-769. doi:10.1002/sres.2388

Tranter, B. (2011). Political divisions over climate change and environmental issues in Australia. *Environmental Politics, 20*(1), 78-96. doi:10.1080/09644016.2011.538167

Trouche, E., Johansson, P., Hall, L., & Mercier, H. (2015). The Selective Laziness of Reasoning. *Cognitive science, 40*, 2122-2136. doi:10.1111/cogs.12303

UNFCCC. (2015). *Paris Agreement*. Retrieved from Paris, France: http://unfccc.int/files/meetings/paris_nov_2015/application/pdf/paris_agreement_english_.pdf

Weinhardt, J. M., Hendijani, R., Harman, J. L., Steel, P., & Gonzalez, C. (2015). How analytic reasoning style and global thinking relate to understanding stocks and flows. *Journal of Operations Management, 39–40*, 23-30. doi:10.1016/j.jom.2015.07.003

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930. doi:10.1037/a0012842

Figure 1.JPEG

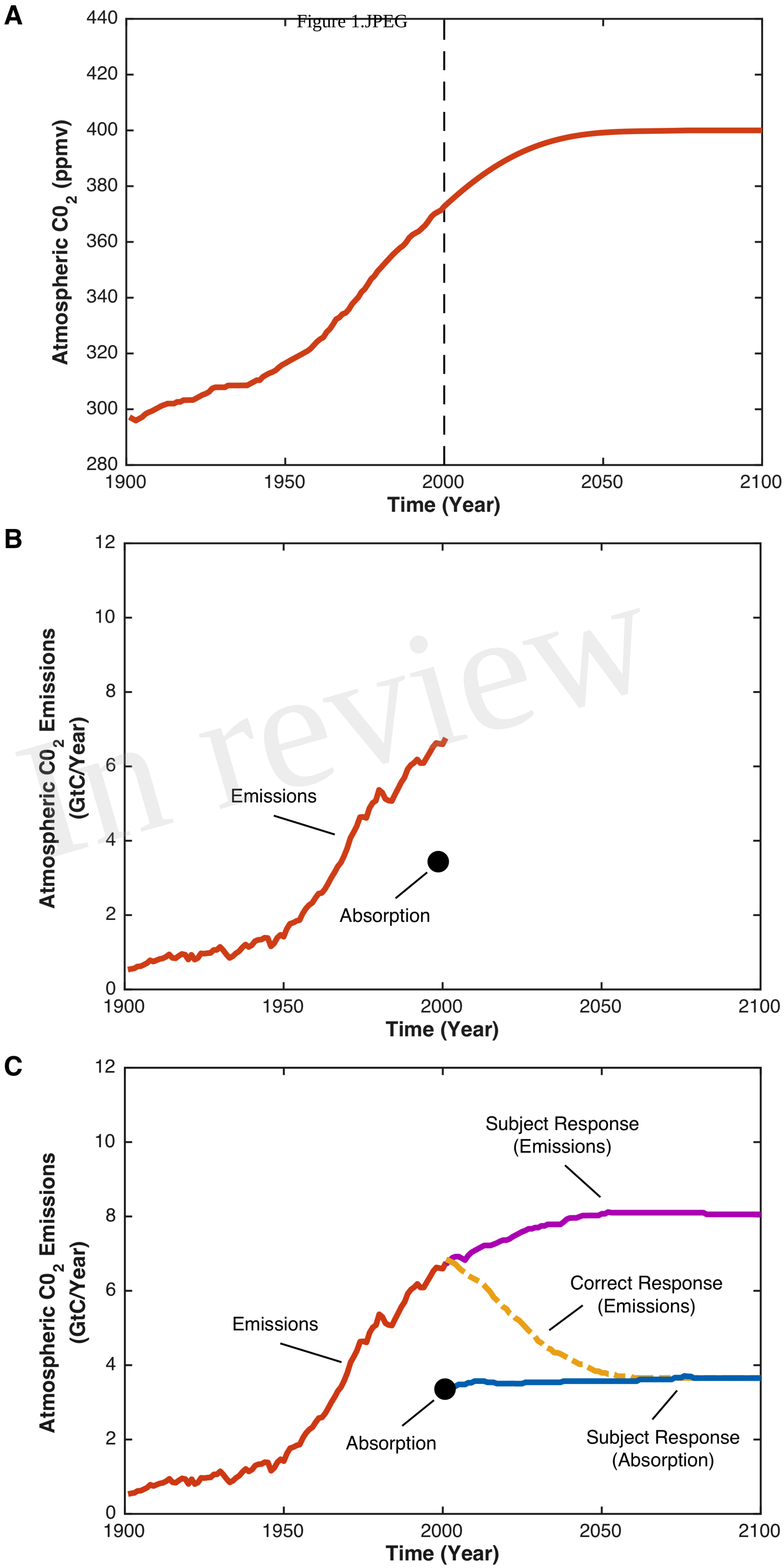

Figure 2.JPEG

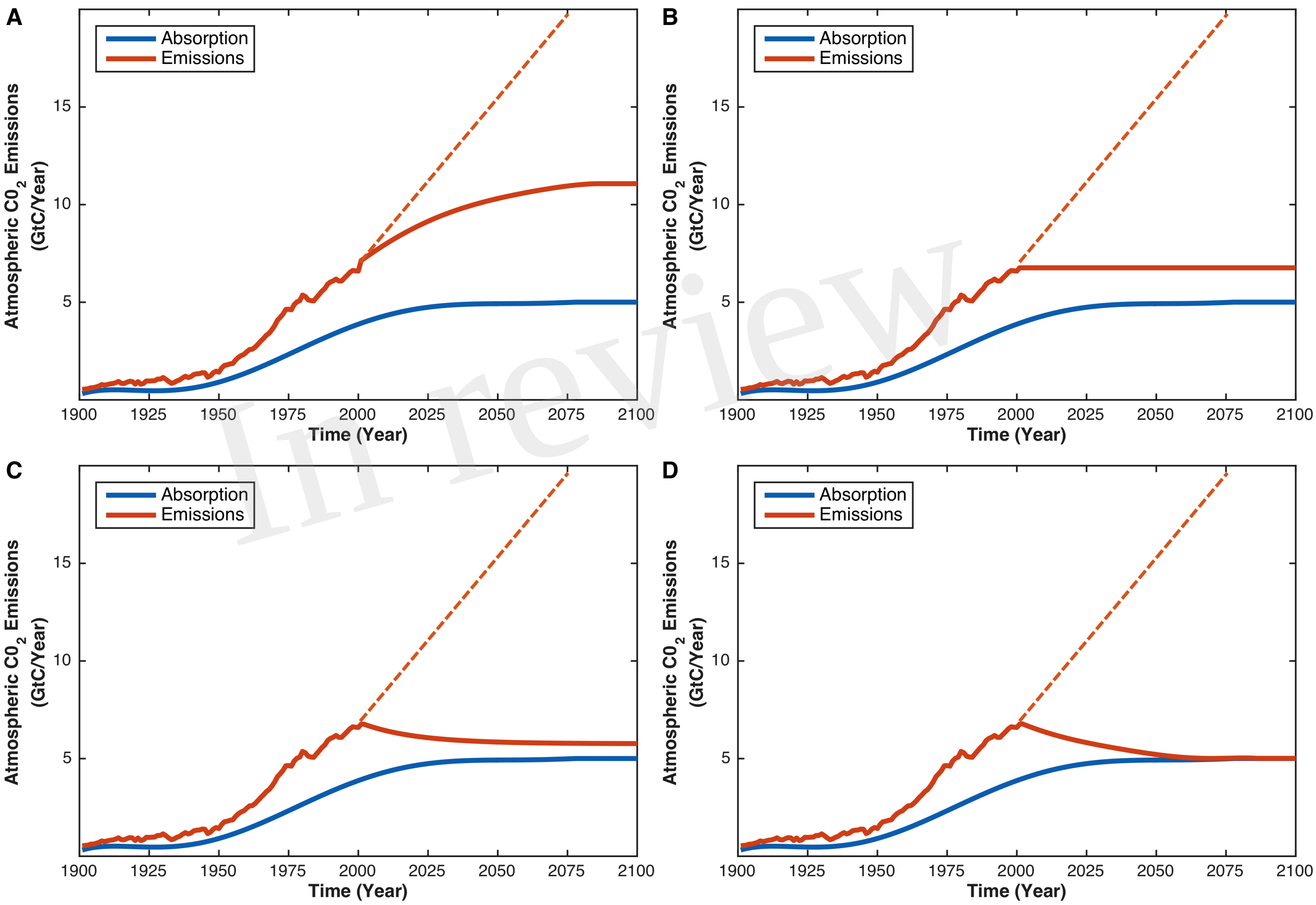

Figure 3.JPEG

A

Group Based Decision Making Task

We are now going to get you to perform a group based decision making task. In the next stage, you will be presented with a display containing two panels. The left panel will identify the graph that you and the other three participants chose earlier, along with the explanation you each gave for doing so. These choices and explanations will be subsumed under a pseudonym that has been assigned to you (visible in the bottom left hand corner of your screen) in order to preserve your annonymity. We would first like you to inspect the decisions and explanations of your group members. After doing so, your group's task is to reach a consensus decision regarding which of the four graphs would stabilise atmospheric concentrations at 420ppm by 2100. To do this, you and your group members will be able to use a chat panel on the right of the display to communicate with one another.

When your group has reached a consensus decision regarding which graph is correct, you will each have to click the continue button in the bottom right corner of the display (the button will only appear after the 10 minutes has elapsed). The group member named **Leda** will then be required to enter the final decision of the group.

Compared to solving problems alone, working in a group and sharing different opinions can lead to a more complete investigation of a problem. In order to work most effectively as a group, here are some guidelines to keep in mind:

1. **Avoid arguing blindly** for your position. Present your thoughts clearly and persuasively, but also pay serious attention to what other people are saying.

2. There is **no competition** involved in this task, so don't think of it as something that can be won or lost by individual group members.

3. **Don't use strategies** like majority voting or picking an answer at random. You want to reach a **consensus** in which everyone understands the problem and agrees on the answer.

4. **Disagreements** are natural and helpful. The more assumptions and opinions you share, the better the decision-making process will be.

5. Approach this with an **open, but sceptical,** mind. No-one, including you, has to be "right", and the group will benefit when members question each other's assumptions from the beginning.

6. Don't be afraid to **speak up and ask questions.** Remember that your goal is to get the right answer (not to please the other members of your group!).

You will only have **10 minutes in the group discussion,** and you must reach a group consensus decision **before the timer runs out** .

B

Reflection Decision Making Task

You will now be given **10 minutes** to reflect on your decision. Use this time to type any additional ideas or thought processes you may have in the communication panel on the right of the display. At the end of 10 minutes, we will ask for your decision again and this final answer can be different to your initial answer. Of course, you can choose your initial answer again.

Reflecting on your own decision-making can lead to a more complete investigation of a problem. In order to think more effectively, approach this with a sceptical mind. This means questioning your own assumptions and considering alternative arguments.

You don't have to make sure that everything you type is "right", but the more assumptions and opinions that you consider, the better the decision-making process will be.

Figure 4.JPEG

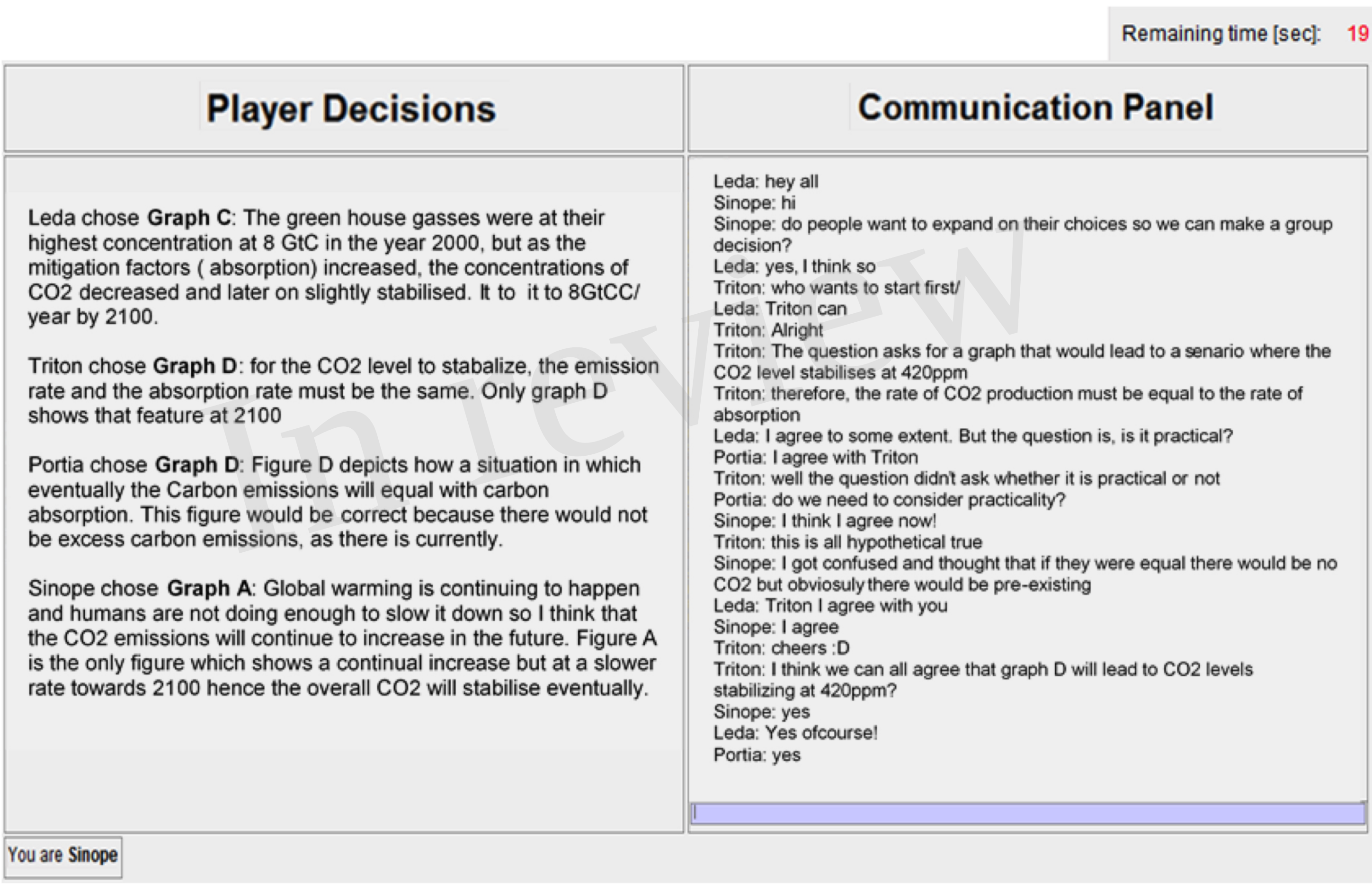

Figure 5.JPEG