

Reducing the Linear Perception of Nonlinearity: Use of a Physical Representation

VARUN DUTT* and CLEOTILDE GONZALEZ

Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

Research shows that while judging accumulations of quantities over time (e.g., money in a bank account or CO₂ in Earth's atmosphere), people assume that the shape of the accumulation is similar to the shape of the inflow (i.e., people rely on a correlation heuristic). Relying on the correlation heuristic is particularly worrisome for Earth's climate, as judging the CO₂ accumulation according to its emissions (inflow) would underestimate the actual (nonlinear) increase. This misperception undermines the seriousness of climate problem and results in wait-and-see behavior. We report two experiments where the effectiveness of a physical representation is compared with graphical and text representations in reducing people's underestimation of nonlinear accumulation in different contexts and problems. A physical representation presents an accumulation using a picture that works as a metaphor. In the first experiment, participants drew the shape of an accumulation over time relying on physical or graphical representations in one of two contexts: carbon dioxide and marbles. Although the participants underestimated the accumulation in both contexts, underestimations were reduced in the physical representation compared with the graphical representation. In the second experiment, we extended the evaluation of physical representations against both text and graphical representations in two different climate problems (with linearly increasing or decreasing inflow). Again, underestimations of the accumulation were reduced in the physical representation compared with the other two representations, regardless of the nature of the problem. We discuss implications of using the physical representation for improving people's estimates of nonlinear CO₂ accumulation. Copyright © 2011 John Wiley & Sons, Ltd.

KEY WORDS linear thinking; climate change; physical representation; graphical representation; text representation; correlation heuristic

Consider the story of an ancient king and the inventor of chess. When the inventor presented the chess game to the king, the king was so impressed that he offered him a reward. The inventor asked for rice grains for each square of the chess board, such that he would get one grain for the first square of the board, two for the second, four for the third, and so on (i.e., 2^{n-1} , where n is the square number starting at $n=1$). The king thought that the inventor was modest, but he belatedly realized how difficult it would be to meet the request halfway into the exercise. Assuming that 60 grains weigh 1 gram, meeting the inventor's request would amount to 153 billion tons of rice grains, which would need 31 million cargo ships capable of holding 5000 tons each. The king failed to perceive the nonlinearity in the request and underestimated the accumulation of rice.

Examples of accumulation of nonlinear quantities are pervasive, and the king is not alone regarding the difficulties of understanding nonlinear accumulation. There is currently growing evidence that a large majority of adults fail to perceive the effects of accumulation in nonlinear problems in different contexts (Cronin & Gonzalez, 2007; Cronin, Gonzalez & Serman, 2009; Dörner, 1980; Dörner, Kimber & Kimber, 1997). Research has shown that people have no intuitive feeling for processes that develop nonlinearly, regardless of how common these problems are in the real world (Cronin et al., 2009). For example, when participants were asked to infer or sketch the shape of a nonlinear accumulation due to changes in the inflow (a rate that increases the

accumulation) and the outflow (a rate that decreases the accumulation), more than half responded as if its shape was linear (Cronin et al., 2009). According to Cronin et al. (2009), participants could be classified as relying on an intuitive but erroneous heuristic, called the "correlation heuristic" (CH), where they incorrectly assume that a quantity's accumulation should "look like" or have the same shape as the inflow. According to Cronin et al. (2009), if the inflow is linear, then people relying on CH will infer the accumulation's shape to be linear as well. Moreover, their reliance cannot be attributed to their inability to interpret graphs, contextual knowledge, motivation, and cognitive capacity.

Studying our understanding of nonlinear accumulations is important not only in simple mathematical problems but also for global problems with serious socio-economic impact such as those concerning Earth's climate. For example, well-educated participants with backgrounds in science and mathematics rely on correlational reasoning when judging changes in an accumulation of carbon dioxide (CO₂) (Dutt & Gonzalez, 2010; Serman, 2008; Serman & Booth Sweeney, 2007). In these studies, participants were given a problem where the CO₂ accumulation was shown to change nonlinearly as a result of both emission (inflow) and absorption (outflow) over time. They were asked to sketch the emission and absorption trajectories that would produce the given CO₂ accumulation trajectory. Participants relying on CH, however, misperceived the dynamics of the future CO₂ accumulation. They assumed that if one is to stabilize the accumulation at a level greater than *status quo*, then emissions should rapidly increase and stabilize at a higher level as well (Dutt & Gonzalez, 2010; Serman, 2008; Serman & Booth Sweeney, 2007). Thus, participants based their inferences solely on the shape of emission trends and did not

*Correspondence to: Varun Dutt, Dynamic Decision Making Laboratory, 4609 Winthrop Street, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA. E-mail: varundutt@cmu.edu

consider the shape of both emission and absorption trends together with the accumulation's initial value.

Moreover, as a consequence of relying on CH when emissions increase linearly over time, participants' correlational (linear) reasoning would grossly underestimate the actual nonlinear increase in the accumulation. Underestimation is a serious problem that undermines the urgency of the climate change problem and likely encourages "wait-and-see" behavior (Dutt & Gonzalez, in press, 2010; Sterman, 2008). Wait-and-see behavior becomes particularly worrisome for Earth's climate because of the inherent long delays between mitigation actions and corresponding changes in atmospheric CO₂ (Intergovernmental Panel on Climate Change [IPCC], 2007; Sterman, 2008; Sterman & Booth Sweeney, 2007). Given such delays, atmospheric CO₂ accumulation would continue to rise until emission equals the absorption rate. Average atmospheric temperature would then peak, and rising sea level from ice melt and thermal expansion would continue (Meehl et al., 2005; Wigley, 2005). Therefore, current wait-and-see policies 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).

In this paper, we hypothesize that people's reliance on CH and their consequent underestimation of nonlinear accumulation is influenced by the format in which information is communicated. Research indicates that when accumulation problems are presented using text or function graphs, responses often rely on CH (Cronin & Gonzalez, 2007; Cronin et al., 2009). This evidence extends to CO₂ accumulations in climate problems (Sterman & Booth Sweeney, 2002, 2007). However, research is critically needed on what presentation formats could reduce reliance on CH (Cronin et al., 2009). Here, we motivate and investigate the use of a physical representation, which presents a problem using a picture as a metaphor.

Experiment 1 evaluates a physical representation against a conventionally used graphical representation in nonlinear problems in two different contexts: a generic non-climate context (i.e., accumulation of marbles in a container) and a specialized climate context (i.e., accumulation of CO₂ in Earth's atmosphere). Experiment 2 builds upon the results of Experiment 1 and investigates the effectiveness of the physical representation against both a text and a graphic representation in two nonlinear problems that differ in their dynamics: an increasing function where the inflow increases linearly while the outflow is constant and a decreasing function where the inflow decreases linearly while the outflow is constant. We believe that the use of physical representation will improve people's understanding of nonlinear CO₂ accumulation in the atmosphere and nonlinear accumulations in other problems and contexts.

EXPERIMENT 1: PHYSICAL AND GRAPHICAL REPRESENTATIONS IN DIFFERENT CONTEXTS

Research shows that reliance on CH in nonlinear problems represented with graphs is robust and increases as the

problem's complexity increases (Cronin & Gonzalez, 2007; Cronin et al., 2009; Dörner, 1980; Dörner et al., 1997). Because climate is a complex system and makes extensive use of graphical representations for communicating climate change, it is likely that people would also rely on CH. For example, the IPCC (2001a) report has a number of graphical figures that illustrate CO₂ emission scenarios and their corresponding CO₂ accumulation in the atmosphere projected over time. Figure 1 shows an example from the IPCC (2001a) Synthesis Report's Summary for Policymakers. The figure shows different hypothesized CO₂ emission trajectories over a 300-year period. Each of these CO₂ emission trajectories leads to a nonlinear projection for CO₂ accumulation over time (please refer to IPCC, 2001a report for other examples).

Similarly, the United States' Environment Protection Agency (EPA) explains historic climate change using mathematical graphs detailing nonlinear increases in CO₂ accumulation over time (EPA, 2010), and graphical representations that communicate climate change have also been common in news reports (Schiermeier, 2010).

Prior research has evaluated people's reliance on CH for climate problems that are presented graphically (Sterman, 2008; Sterman & Booth Sweeney, 2007). Participants are asked to sketch the CO₂ emission and absorption that would stabilize the CO₂ concentration to an attainable goal by the year 2100. 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 emission trajectories that were positively correlated with the concentration trajectories. Therefore, these participants robustly relied on CH.

We believe that using physical representations to depict CO₂ accumulation, CO₂ emission, absorption, or average atmospheric temperature over time is likely to improve people's understanding of climate change. A physical representation would present a problem using a series of pictures that work as a "metaphor" to explain changes in the quantity of interest. For example, consider a problem where the length

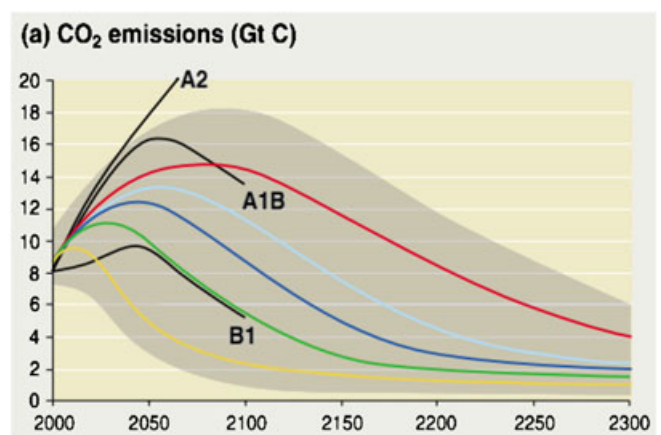


Figure 1. A figure taken from the IPCC Synthesis Report's Summary for Policymakers document. Different emission trajectories (A2, A1B, B1, etc.) are sketched (in gigatons of carbon on the y-axis) with respect to time (in years on the x-axis). Source: IPCC (2001b; Figure SPM 6a, p. 20)

(L) of a square's side is doubled, and one needs to calculate the new area. A physical representation will be one where a single square tile of side L is replaced by four square tiles, each of side L (i.e., to form a square of side $2L$). There is abundant literature in mathematical education that depicts how and why drawings, pictures, and diagrams facilitate people's ability to solve mathematical problems (e.g., Aprea & Ebner, 1999; De Corte, Greer & Verschaffel, 1996; Hall, Bailey & Tillman, 1997; Larkin & Simon, 1987; Schoenfeld, 1992), and we expect that pictorial metaphors are likely to reduce underestimation and reliance on the CH in nonlinear problems. For example, pictorial representations using an array of icons as a metaphor to represent fractions have been effective in reducing people's tendency to neglect the denominator of a fraction while evaluating a probability or a risk (Garcia-Retamero et al., 2010) and aids people with low numeracy skills in forming better judgments about risks in general (Galesic et al., 2009), and aids people with low numeracy skills in forming better judgments about risks in general (Galesic et al., 2009). Also, research has revealed that a picture of a nonlinear problem drawn by students themselves helps them to construct a proper mental representation of the essential elements and relations involved (Pólya, 1945; Schoenfeld, 1992). Making a drawing or diagram, however, does not guarantee that one will find the solution if the representation is incorrect (Van Essen & Hamaker, 1990). According to De Bock, Van Dooren, Janssens, and Verschaffel (2007), an effective method provides students with a "correct ready-made drawing" or a physical representation of a nonlinear problem as a metaphor that communicates its dynamics. Using this idea, Van Dooren, De Bock, Janssens, and Verschaffel (2007) have shown that among school children, the ability to solve nonlinear area problems improves with a physical representation compared with a graphical representation. The performance of the group given a physically represented word problem was superior to the group given a graphical problem. Thus, we expect that

H1 : Estimations of accumulation will be more accurate in a physical representation compared with a graphical representation.

If the physical representation can effectively communicate a problem's dynamics and help participants to answer correctly, then the dependence of this effect on a problem's context or cover story is important to determine. That is because prior research has shown that the general public lacks training in climatology and has little understanding of climate processes (Bostrom, Morgan, Fischhoff & Read, 1994; Kasemir et al., 2000; Kempton, 1997; Morgan, Fischhoff, Bostrom & Atman, 2002; Palmgren, Morgan, de Bruin & Keith, 2004; Read, Bostrom, Morgan, Fischhoff & Smuts, 1994). Consequently, the climate context is expected to be unfamiliar compared with other contexts encountered in day-to-day judgments: accumulation of water in a bathtub with a tap adding water and a drain removing water from the tub, of money in a bank account with income and expenditure, or of marbles in a container with marbles being put in and removed.

Current research is not conclusive regarding the effect of context on problem solving. On the one hand, research has shown that people's judgments are often influenced by their familiarity with the context of the problem (Brunstein, Gonzalez, & Kanter, 2010; Gigerenzer & Hug, 1992). For example, Brunstein, Gonzalez, and Kanter (2010) put nonlinear accumulation problems into the medical context and found that context familiarity actually hurts participants' performance. Medical students did worse in the medical context than in a generic context. Other research, however, reveals no influence of context (e.g., Almor & Sloman, 2000; Cronin et al., 2009). For example, Cronin et al. (2009) found no effect of familiarity with the context on participants' reliance on CH. In their design, however, the three contexts used were represented using only graphical representations.

We extend this analysis to a physical representation and its comparison with the graphical representation in both a climate and non-climate context. Given the expected effectiveness of the physical representation and the lack of consensus regarding the effect of context, we expect that

H2 : Estimations of accumulation will be more accurate in a physical representation compared with a graphical representation, regardless of the context used.

Participants

One hundred and thirty-two adults from Pittsburgh, PA were recruited through a website advertisement and participated in this experiment. Forty-four percent were graduate students enrolled in an MS or PhD program or had completed one of these degrees in the past. The rest were undergraduates either enrolled in an undergraduate program or had completed a bachelor's degree. Forty-four participants were women. Ages ranged from 18 to 62 years ($M = 23$ years, $SD = 6$ years). Sixty-two percent of participants self-reported having degrees in STEM, and the rest have non-STEM backgrounds. All participants received a flat \$3 compensation for answering an accumulation problem.

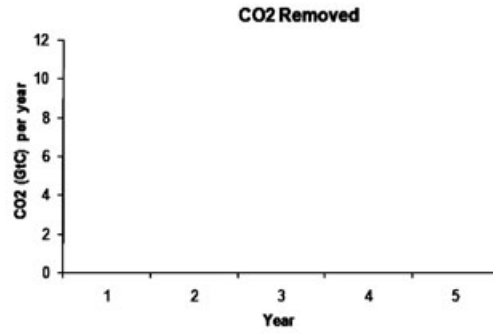
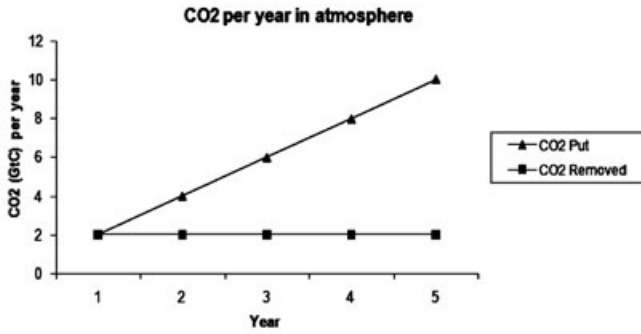
Experimental design

On the basis of the two representations and the two contexts within each, participants were randomly assigned to one of four between-subjects treatments: climate context and physical representation (climate-physical, $N = 25$), climate context and graphical representation (climate-graphical, $N = 25$), marble context and physical representation (marble-physical, $N = 39$), and marble context and graphical representation (marble-graphical, $N = 42$). In each treatment, participants were provided with one accumulation problem that differed in the context and the representation. All four problems were mathematically identical.

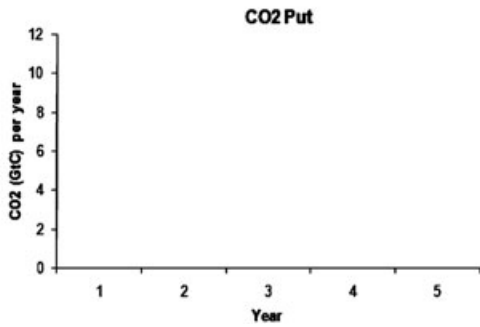
Figure 2 shows an example of the graphical representation of the inflow and outflow provided to participants in the climate context. The graphs provided in the marble context were identical. In the climate context, the CO₂ emission and absorption were the inflow and the outflow. In the marble context, the marbles inserted in and removed out of the container were the inflow and the outflow. In both

The figure given below shows carbon-dioxide (CO₂) put and removed from the atmosphere in each of 5 years (GtC = Giga or 10⁹ tons of carbon).

In the space provided below draw the CO₂ removed from the atmosphere per year:



In the space provided below draw the CO₂ put in the atmosphere per year:



In the space provided below draw the total amount of CO₂ in the atmosphere in each of the 5 years. The initial amount of CO₂ in the atmosphere has been shown as 10 GtC in the first year.

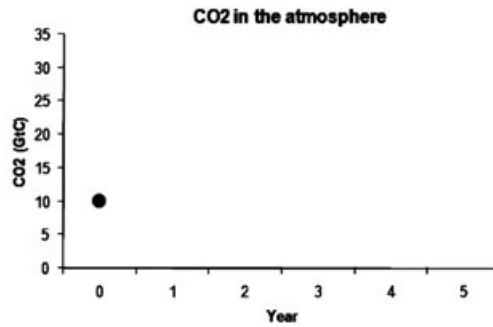


Figure 2. The task presented to participants in the graphical representation for the climate context in Experiments 1 and 2. Participants were provided with graphs showing changes to the inflow and outflow over time and were asked to sketch the accumulation value over the five periods because of the changes in the inflow and the outflow

contexts, the inflow increased linearly over time, whereas the outflow remained constant over all five periods. Participants were first asked to sketch the inflow and the outflow in a blank graph provided. Asking participants to sketch the given inflow and outflow was done to test their understanding of these flows in the problem. Then, they sketched the accumulation in another blank graph provided. As shown in Figure 2, participants were given an initial accumulation value (at time=0), which was depicted as a black dot on the graph. The linearly increasing inflow and the constant outflow result in a nonlinearly (parabolic) increasing accumulation. This problem was also used by Cronin et al. (2009) to demonstrate participants' reliance on CH in the graphical representation.

In the physical representation, participants were shown "opaque" containers (representing Earth's atmosphere) in three states for each period: initial state, inflow, and outflow. The units of inflow and outflow are shown as small circles and with numbers, and the direction of the flow is indicated by arrows (see Figure 3 for the corresponding physical representation of the graphical form for the climate context of Figure 2). Like in the graphical representation, participants sketched the accumulation. Again, the same linearly increasing inflow and constant outflow as in the graphical representation were used. Participants were asked to read the instructions and to sketch the inflow, the outflow, and the accumulation over time.

These problems depicted the changes in the inflow and the outflow over five periods (or years) and asked participants to sketch the curve of the inflow, the outflow, and the resulting accumulation over that time.

Evaluating participants' responses

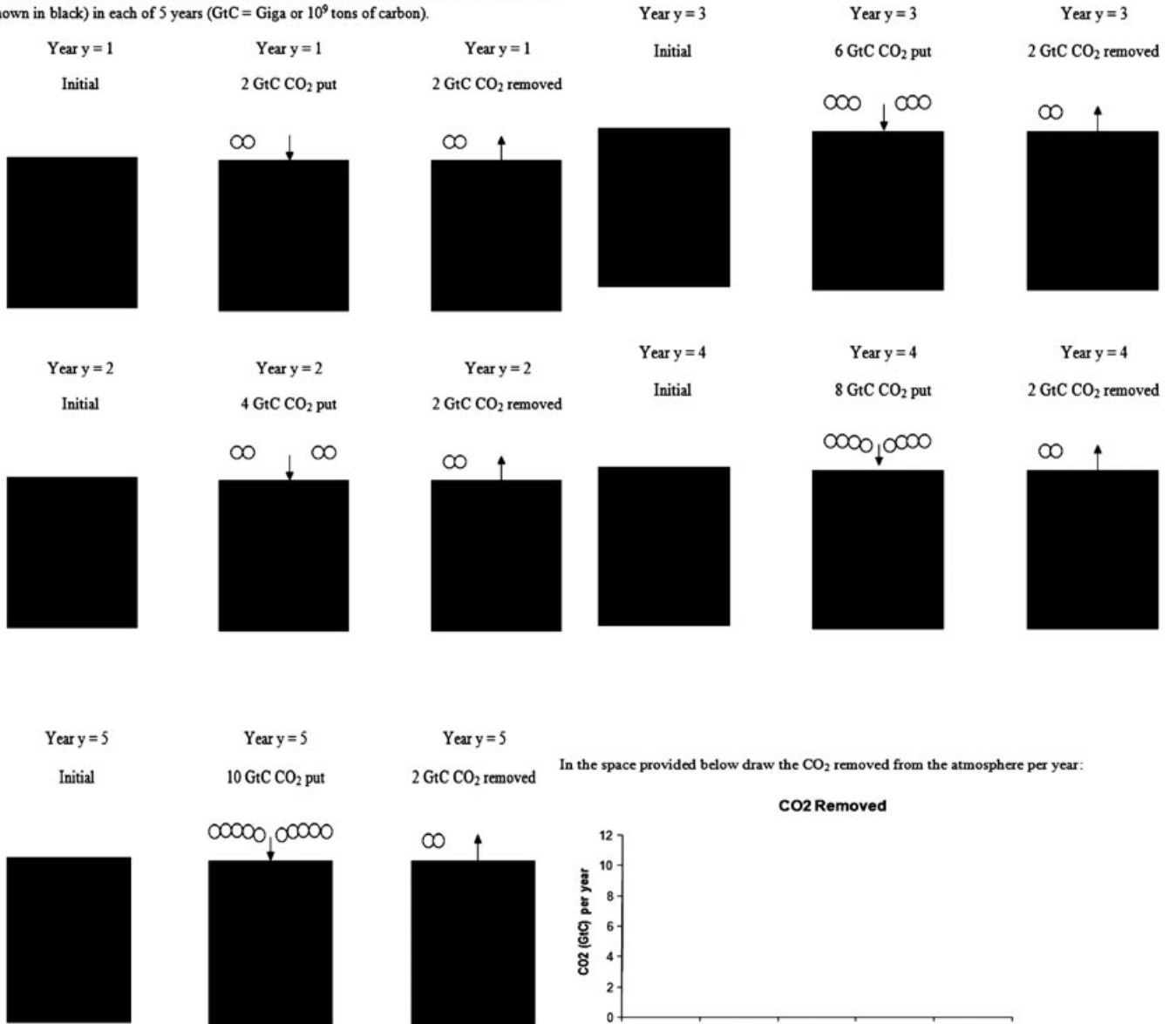
The correct accumulation values were the same in all treatments and could be derived by repeatedly using the following equation:

$$S_T = S_{T-1} + I_T - O_T \tag{1}$$

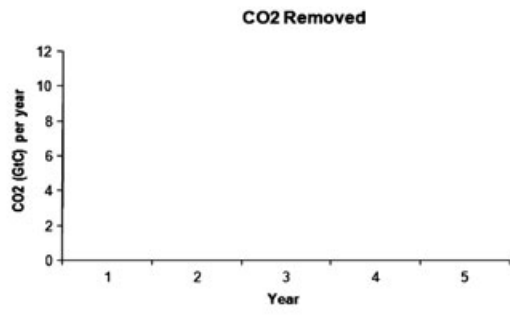
That is, the accumulation at time T (S_T) is the sum of the accumulation at time $T - 1$ (S_{T-1}) and the net inflow (=inflow (I_T) - outflow (O_T), at time T). For example, given an initial accumulation of 10 units and an inflow and outflow of two units each in the first period, the accumulation at the end of Period 1 will remain at 10 units (=10+2-2). Similarly, the accumulation in the second period (with four units inflow and two units outflow) will become 12 units. Calculated in the same way, the accumulation in the Periods 3, 4, and 5 will be 16, 22, and 30 units, respectively. Given that the accumulation increases nonlinearly over time, the effect proposed in H1 would become more prominent in the later periods than in the first few periods. In the first few periods, participants should appear to be more accurate than in later periods.

The sketched accumulation values averaged over all participants in each of the five periods were used as the main source of analyses. Each average accumulation value was compared with the correct corresponding values. Participants underestimated the correct accumulation if their sketched accumulation values were less than the correct values. This comparison of values in each period allows us to test participants' reliance on CH using one-sample t -tests.

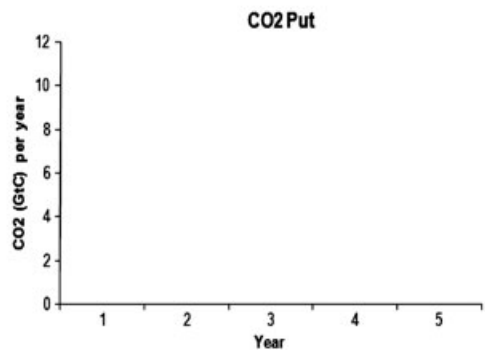
The figures given below shows carbon-dioxide (CO₂) put and removed from the atmosphere (shown in black) in each of 5 years (GtC = Giga or 10⁹ tons of carbon).



In the space provided below draw the CO₂ removed from the atmosphere per year:



In the space provided below draw the CO₂ put in the atmosphere per year:



In the space provided below draw the total amount of CO₂ in the atmosphere in each of the 5 years. The initial amount of CO₂ in the atmosphere has been shown as 10 GtC in the first year.

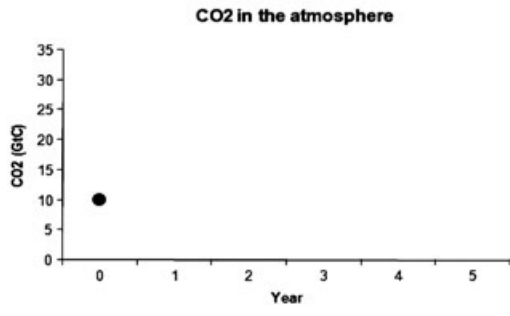


Figure 3. The task presented to the participants in the physical representation for the climate context in Experiments 1 and 2. Participants were provided a set of opaque containers that depicted the changed in the inflow and the outflow over time. Participants were asked to sketch the accumulation value over the five periods because of the changes in the inflow and the outflow

Correlation coefficients were calculated between participants' sketched inflow and outflow values in each of the five periods and the correct corresponding values given to them.

If any of these two correlation coefficients were different from 1.00 between the sketched and correct inflow and outflow, then the sketches were marked as incorrect. Moreover,

to classify participants as relying on CH, we correlated their sketched accumulation in each period to the correct inflow values. If this score was 1.00, then the sketched accumulation was classified as relying on CH. The coefficient value of 1.00 is a conservative grading scheme because CO₂ emissions are linearly increasing, and any nonlinear increasing sketch of CO₂ accumulation would result in a high positive correlation coefficient. The conservative grading scheme ensures that the relationship between the CO₂ emissions and concentration is none other than perfectly linear and one that follows CH. According to CH reliance, the difference between the correct accumulation values and corresponding human responses would increase over the time because of the increase in the nonlinearity of the accumulation.

Results

The percent of correct inflow and outflow responses in different treatments was more than 93%. Therefore, the participants' understanding of the inflows and the outflows was similar and highly accurate across different treatments.

To test H1, we compared the participants' average sketched accumulation value in each of the five periods with the correct corresponding values in the physical and graphical representations (Table 1). As expected, there was no difference between the correct and the corresponding average accumulation in the graphical and physical representation in the first three periods. For the last two periods, the average accumulation in the physical representation was closer to the correct value, compared with in the graphical representation. This result is illustrated in Table 1 by the weaker *p* values and smaller standard deviations and effect sizes in the physical representation compared with that in the graphical representation for the last two periods. Accumulation estimates in the physical representation were more accurate than those in the graphical representation (H1). The graphical representation showed more linear responses than the physical representation, as demonstrated by significantly larger underestimations in the last periods.

Table 2 presents the correct accumulation in different periods and the corresponding accumulation in the four treatments to test the effects of context: climate or marble.

Table 1. The correct accumulation in different periods and their corresponding average accumulation in the graphical and physical representations in Experiment 1

Representation	Time 1	Time 2	Time 3	Time 4	Time 5
Correct	10.0 (0.0) ^a	12.0 (0.0)	16.0 (0.0)	22.0 (0.0)	30.0 (0.0)
Graphical (G) Statistics (compared with correct)	10.4 (2.2) <i>t</i> (66) = 1.5, <i>ns</i> , <i>r</i> = 0.18 ^b	12.6 (3.4) <i>t</i> (66) = 1.5, <i>ns</i> , <i>r</i> = 0.18	15.7 (4.3) <i>t</i> (66) = -0.5, <i>ns</i> , <i>r</i> = 0.06	19.6 (5.6) <i>t</i>(66) = -3.4, <i>p</i> < .001, <i>r</i> = 0.39^c	24.7 (7.7) <i>t</i>(66) = -5.7, <i>p</i> < .001, <i>r</i> = 0.57
Physical (P) Statistics (compared with correct)	10.0 (1.1) <i>t</i> (63) = -0.3, <i>ns</i> , <i>r</i> = 0.04	11.9 (1.2) <i>t</i> (63) = -0.5, <i>ns</i> , <i>r</i> = 0.06	15.8 (1.5) <i>t</i> (63) = -0.4, <i>ns</i> , <i>r</i> = 0.05	21.4 (2.4) <i>t</i>(63) = -2.1, <i>p</i> < .05, <i>r</i> = 0.26	28.8 (4.0) <i>t</i>(63) = -2.4, <i>p</i> < .05, <i>r</i> = 0.29

^aNote: The values in bracket represent the standard deviation about the mean.

^bThe value indicates the effect size.

^cNumbers in boldface indicate results that are significant at *p* less than or equal to .05.

Table 2. The correct accumulation in different periods and their corresponding average accumulation in the four treatments in Experiment 1

Treatment	Time 1	Time 2	Time 3	Time 4	Time 5
Correct	10.0 (0.0) ^a	12.0 (0.0)	16.0 (0.0)	22.0 (0.0)	30.0 (0.0)
Climate-physical Statistics (compared with correct)	10.2 (0.8) <i>t</i> (24) = 1.0, <i>ns</i> , <i>r</i> = 0.20	12.1 (1.1) <i>t</i> (24) = 0.6, <i>ns</i> , <i>r</i> = 0.12	15.9 (1.2) <i>t</i> (24) = -0.5, <i>ns</i> , <i>r</i> = 0.10	21.2 (2.3) <i>t</i> (24) = -1.7, <i>ns</i> , <i>r</i> = 0.33	28.4 (4.0) <i>t</i> (24) = -2.0, <i>ns</i> , <i>r</i> = 0.38
Marble-physical Statistics (compared with correct)	09.8 (1.3) <i>t</i> (38) = -0.9, <i>ns</i> , <i>r</i> = 0.14	11.8 (1.3) <i>t</i> (38) = -1.0, <i>ns</i> , <i>r</i> = 0.16	15.7 (1.7) <i>t</i> (38) = -0.4, <i>ns</i> , <i>r</i> = 0.03	21.4 (2.6) <i>t</i> (38) = -1.4, <i>ns</i> , <i>r</i> = 0.22	29.1 (4.0) <i>t</i> (38) = -1.4, <i>ns</i> , <i>r</i> = 0.22
Climate-graphical Statistics (compared with correct)	10.3 (2.4) <i>t</i> (24) = 0.6, <i>ns</i> , <i>r</i> = 0.12 ^b	12.3 (2.6) <i>t</i> (24) = 0.6, <i>ns</i> , <i>r</i> = 0.12	15.0 (3.1) <i>t</i> (24) = -1.5, <i>ns</i> , <i>r</i> = 0.29	18.4 (4.2) <i>t</i>(24) = -4.3, <i>p</i> < .001, <i>r</i> = 0.66^c	22.6 (6.5) <i>t</i>(24) = -5.7, <i>p</i> < .001, <i>r</i> = 0.76
Marble-graphical Statistics (compared with correct)	10.5 (2.0) <i>t</i> (41) = 1.6, <i>ns</i> , <i>r</i> = 0.24	12.8 (3.8) <i>t</i> (41) = 1.4, <i>ns</i> , <i>r</i> = 0.21	16.1 (4.9) <i>t</i> (41) = 0.2, <i>ns</i> , <i>r</i> = 0.03	20.4 (6.2) <i>t</i> (41) = -1.7, <i>ns</i> , <i>r</i> = 0.26	25.9 (8.1) <i>t</i>(41) = -3.3, <i>p</i> < .01, <i>r</i> = 0.46

^aNote: The values in bracket represent the standard deviation about the mean.

^bThe value indicates the effect size.

^cNumbers in boldface indicate results that are significant at *p* less than or equal to .05.

There was no effect of context in the physical representation. The accumulation values of the physical representations were no different than the correct accumulation values for all periods in both contexts. In contrast, there was an effect of context for the graphical representation, where people were more accurate in Period 4 in the marble context than in the climate context. But participants were similarly inaccurate in Period 5. In any case, regardless of the context, graphical representations led to underestimating the accumulation, whereas physical representations did not, supporting H2.

We analyzed the proportion of responses that relied on CH in the graphical compared with physical representations. Figure 4 shows a typical participant's response in the marble-graphical treatment that was classified as relying on CH (based upon a 1.0 correlation coefficient between the average accumulation and the inflow given in the treatment). The participant's accumulation in the fifth period (=18 marbles) underestimated the correct accumulation (=32 marbles) as a result of his relying on CH.

Table 3 shows that the proportion of responses that were classified as relying on CH was significantly greater in the graphical representation, regardless of context (climate or marble). That is because of the weaker *p* value and the effect

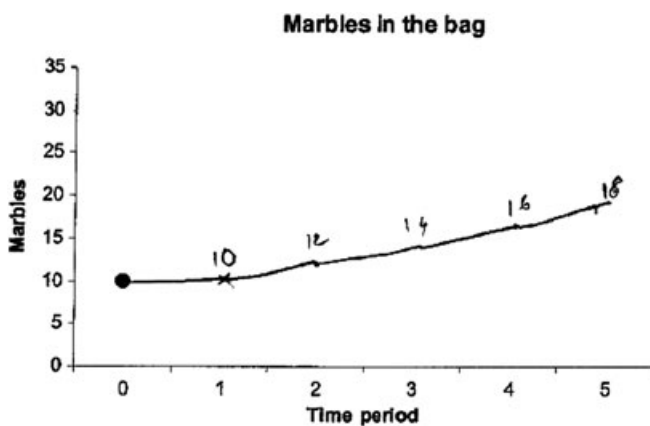


Figure 4. A typical participant's response in the marble-graphical treatment that shows the participant's sketched accumulation that was classified as relying on correlation heuristic. The marbles in the bag (i.e., accumulation) follows a linear trend over periods (with a constant slope of two marbles per period). The correlation coefficient of the marble accumulation with the inflow (i.e., marbles put) is 1.0. Also, the accumulation in the fifth period (=18 marbles) underestimates its correct value (=30 marbles)

Table 3. Proportion of responses classified as relying on the correlation heuristic (CH) in different treatments in Experiment 1. Comparison statistics with the correct accumulation's CH value (=0%) are also shown

Treatment	CH (%)	Statistics (compared with correct)
Correct	00	—
Graphical (G)	52	$t(66) = 8.5, p < .001, r = 0.72^a$
Climate-graphical	72	$t(24) = 7.9, p < .001, r = 0.85$
Marble-graphical	40	$t(41) = 5.3, p < .001, r = 0.64$
Physical (P)	09	$t(63) = 2.6, p < .05, r = 0.31$
Climate-physical	16	$t(24) = 2.1, p < .05, r = 0.39$
Marble-physical	05	$t(38) = 1.4, ns, r = 0.22$

^aNote: Numbers in boldface indicate results that are significant at *p* less than or equal to .05.

size in the physical representation compared with the graphical representation while comparing with a correct response.

Finally, Table A1 in the Appendix presents the results regarding the non-STEM and STEM backgrounds. In the graphical representation, participants' average accumulation underestimated the correct accumulation in the last period for non-STEM background and in the last two periods for STEM background. In the physical representation, there was no difference between the average and the correct accumulation, regardless of the background. A similar pattern was found for undergraduates, who underestimated the correct accumulation in the last period and for graduates, who underestimated in last three periods (see Table A2 in Appendix). For both backgrounds and levels of education, a smaller proportion of responses were classified as relying on CH in the physical representation than in the graphical representation (for detailed statistics, see Table A3 in the Appendix).

Discussion

In agreement with previous literature in mathematical education (Evangelidou et al., 2004; Leinhardt, Zaslavsky & Stein, 1990; Van Deyck, 2001) and on accumulation problems (Cronin et al., 2009), results show that people underestimate nonlinear accumulations and rely on the CH in graphical representations. There seems to be an effect of context present in the graphical representation: A familiar marble context produces better response compared with a less familiar climate context. This result, therefore, is in agreement with prior research that has reported on the effects of contexts (Gigerenzer & Hug, 1992; Brunstein, Gonzalez & Kanter, 2010).

In contrast, a physical representation reduces participants' underestimations as well as their reliance on the CH, regardless of the context, education background, and levels of education. Although we can only speculate, it is likely that the physical representation helps people to understand the accumulation's basic dynamics: Accumulation rises when the inflow is greater than the outflow, falls when the inflow is less than the outflow, and stabilizes when the inflow equals the outflow. Because of metaphor use in the physical representation, participants might better visualize the processes that govern the changes in accumulation, a visualization that is not possible in the graphical representation. Clearly, the physical representation helped people understand the nonlinear growth: As the accumulation increases nonlinearly, performance in the physical representation is better than that in the graphical representation.

Furthermore, we found that accumulation estimations were more accurate in the physical representation compared with the graphical representation as the number of periods increased from one to five. As people rely on CH, the difference between the correct accumulation values and the corresponding human responses would increase over time because of the increase in the nonlinearity of the accumulation. In the next section, we report an experiment extending analysis of the physical representation to text and graphical representations in two different climate problems.

EXPERIMENT 2: PHYSICAL, GRAPHICAL, AND TEXT REPRESENTATIONS IN INCREASING AND DECREASING PROBLEMS

In addition to the graphical representation, prior research has also used text representations to investigate people's reliance on CH in CO₂ accumulation problems (Serman, 2008; Serman & Booth Sweeney, 2007). Reliance on CH has been found to be equally strong in the text representation (Cronin et al., 2009). Also, research has reported the physical representation to be effective against text descriptions in assessments of risk (Garcia-Retamero & Dhami, 2011). These authors examined to what extent Polish immigrants to the UK had difficulties in understanding treatment risk reductions expressed as text ratios (in native or non-native languages) and further whether this population could be aided by a physical representation using visual displays to enhance comprehension. When assessing treatment risk reduction, participants often paid too much attention to the number of treated and non-treated patients who died (i.e., numerators) and paid insufficient attention to the overall number of patients (i.e., denominators) - known as "denominator neglect". However, providing visual aids (physical representation) in addition to the textual numerical information proved to be an effective method for eliminating denominator neglect. According to Garcia-Retamero and Dhami (2011), the visual aids drew participants' attention to the overall number of treated and non-treated patients and helped them make more accurate risk estimates.

Furthermore, we used an increasing accumulation problem in Experiment 1, where the inflow increased linearly and the outflow was constant. This inflow-outflow behavior caused the accumulation to increase nonlinearly over time. However, if one considers a decreasing problem where the trend of the inflow *decreases* over time and is greater than a constant outflow, then the accumulation will still increase nonlinearly over time. This increase, however, will be in direct opposition to the linearly decreasing inflow, allowing us to test people's reliance on CH more convincingly. If participants rely on CH in this decreasing problem, they would also sketch a linearly decreasing accumulation over time, and their sketch will greatly underestimate the actual nonlinearly increasing accumulation. Previous research has shown that a decreasing inflow that results in an increasing accumulation is more challenging than an increasing problem (Dutt, 2011; Dutt & Gonzalez, 2007; Gonzalez & Dutt, 2007, 2011; Lebiere, Gonzalez & Warwick, 2009). Finally, the decreasing problem also seems to be a realistic case for Earth's climate because policymakers could potentially decide upon interventions that decrease CO₂ emissions in the near future, rather than leave them unregulated and increasing over time.

According to these facts and our results of Experiment 1, we expect that

H3 : Estimations of accumulation will be more accurate in a physical representation compared with either graphical or text representation, regardless of whether the inflow linearly increases or decreases over time.

To test H3, we take two nonlinear problems in the climate context, increasing and decreasing inflow, in three different representations: text (i.e., using a written description), graphical (i.e., using mathematical graphs), and physical (i.e., using pictures).

Participants

One hundred and thirty-two adults from Pittsburgh, PA were recruited through a website advertisement and participated in the experiment. Forty-eight percent were graduate students enrolled in a MS or PhD program or had previously completed one of these degrees. The rest were either enrolled in an undergraduate program or had previously completed a bachelor's degree. Forty-seven participants were women. Ages ranged from 18 to 70 years ($M=27$ years, $SD=10$ years). Sixty-nine percent of the participants self-reported having degrees in STEM, and the rest had a non-STEM background. All participants received a flat \$3 compensation for the study.

Experimental design

Participants were randomly assigned to one of six between-subjects treatments: 3 (text, graphical, and physical representations) \times 2 (increasing and decreasing) problems. In each treatment, participants attempted a nonlinear accumulation problem concerning a CO₂ accumulation in the atmosphere with different representations and increasing or decreasing inflows (where inflows were greater than a constant outflow): text increasing ($N=26$), text decreasing ($N=18$), graphical increasing ($N=26$), graphical decreasing ($N=18$), physical increasing ($N=26$), and physical decreasing ($N=18$). Like in Experiment 1, all six problems depicted the changes in the inflow and the outflow over five periods (or years) and asked participants to sketch the curve of the inflow, the outflow, and the resulting accumulation over those five periods. The dependent variables used were identical to those used in Experiment 1.

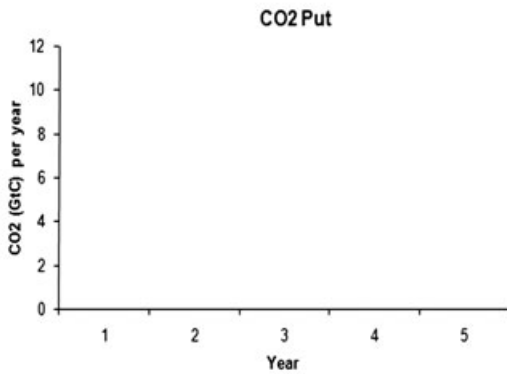
Figure 5 shows the decreasing problem presented to participants in the text representation. The tabulated inflow (CO₂ emissions) decreased linearly over time from 10 giga or 10⁹ tons of carbon (GtC) per year to 2 GtC per year, whereas the outflow remained constant (=2 GtC per year). The increasing problem in the text representation was like the decreasing problem, but now, the inflow increased linearly over time from 2 GtC per year to 10 GtC per year (outflow was constant at 2 GtC per year). The increasing problem in the graphical and physical representations was identical to that used for the climate context in Experiment 1 (except that the initial accumulation was set at 20 GtC instead of 10 GtC so as to keep the initial accumulation equidistant from the two endpoints of the y-axis).¹ The decreasing problem in the

¹In Experiment 1, we placed the starting accumulation ($t=0$) below the mid-point of the y-axis. This non-equidistant placement of initial accumulation along with the climate context (where real-world CO₂ accumulations are increasing) could have hinted to participants that the accumulation was going to increase over the next five periods. Furthermore, this could have been a possible reason for their relying on CH in different treatments in this experiment. Therefore, in Experiment 2, we correct this methodological issue in our design.

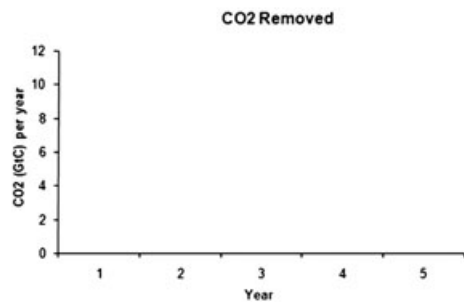
The table given below shows carbon-dioxide (CO₂) put and removed from the atmosphere in each of 5 years (GtC = Giga or 10⁹ tons of carbon).

Year	CO ₂ Put (GtC per year)	CO ₂ Removed (GtC per year)
1	10	2
2	8	2
3	6	2
4	4	2
5	2	2

In the space provided below draw the CO₂ put in the atmosphere per year:



In the space provided below draw the CO₂ removed from the atmosphere per year:



In the space provided below draw the total amount of CO₂ in the atmosphere in each of the 5 years. The initial amount of CO₂ in the atmosphere has been shown as 20 GtC.

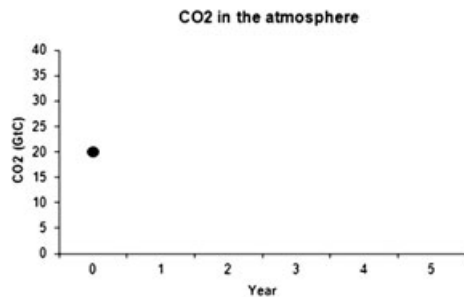


Figure 5. The decreasing problem presented to participants in the text representation in Experiment 2

graphical and physical representations followed the same structure as the increasing problem; however, the inflow decreased linearly in the problem from 10 GtC per year to 2 GtC per year. As in Experiment 1, the correct accumulation in the increasing and decreasing problems could be derived by repeatedly using Equation 1.

Results

In both the increasing and decreasing problems for different representations, more than 96% of participants correctly sketched the inflow and outflow shapes over time. Thus, most participants understood the inflow and outflow well and similarly across conditions.

To test H3, we compared participants’ responses in the text, graphical, and physical representations against the correct responses in both the increasing and decreasing problems. Table 4 presents the correct accumulation in different periods and the corresponding average accumulation participants gave in the graphical, text, and physical representations for each problem. In the increasing problem, the average accumulation in the physical representation was much closer to the correct accumulation in all periods compared with that in the graphical or text representations. Upon comparing the results for graphical and physical representations, we find that these results are in the same direction as reported in Experiment 1.²

In the decreasing problem, the accumulation in the physical representation was again closer to the correct accumulation than in the graphical or text representations, as it is observed by the weaker *p* values and effect sizes of the treatments involving the physical representation. The decreasing problem, however, was clearly more challenging for participants (Dutt, 2011; Gonzalez & Dutt, 2011; Lebiere et al., 2009). Participants relying on CH think that the CO₂ accumulation decreases linearly over the five periods, whereas doing the same in the increasing problem has the accumulation increasing linearly over time. Therefore, it is likely to become counterintuitive for participants in the decreasing problem to sketch accumulations that increase over the five periods. More underestimation of the correct CO₂ accumulation is found in the decreasing problem, regardless of the representation, but the accumulation reported in the graphical and text representations underestimated the correct accumulation more than in the physical representation. Taken together, these confirm our expectation in H3: the estimation of the accumulation is more accurate in the physical representation compared with in the graphical and text representations for both increasing and decreasing problems.

As illustrated in Table 5, greater underestimation in text and graphical representations compared with the physical representation is likely due to reliance on CH. In both increasing and decreasing problems, the proportion of responses classified as relying on CH was lower for the physical representation than for graphical and text representations.

We also tested the effectiveness of the physical representation compared with the graphical and text representations for each problem among different educational backgrounds and levels. For STEM and non-STEM backgrounds in both problems, estimates of the accumulation were much closer to the correct in the physical representation compared with

²This observation is after the fact that Experiments 1 and 2 were run at different times and contained different number of participants in the graphical representation (*N* = 67 in Experiment 1 and *N* = 25 in Experiment 2). Given these differences, even if we make a close comparison between the two experiments, we find that participants underestimated the accumulation in Periods 1–5 in Experiment 2 and in Periods 4–5 in Experiment 1. This difference between the two experiments is likely due to the fact that in Experiment 2, four out of 25 participants drew the accumulation shape exactly like the shape of inflow, whereas only one out of 67 participants showed this behavior in Experiment 1. As the values of inflow are much smaller than those of accumulation, copying the inflow as the accumulation by more participants causes underestimation over all five periods, 1–5, in Experiment 2.

Table 4. The correct accumulation in different periods and their corresponding average accumulation in the graphical, text, and physical representations for the increasing and decreasing problems in Experiment 2

Representation	Time 1	Time 2	Time 3	Time 4	Time 5
Increasing problem					
Correct	20.0 (0.0) ^a	22.0 (0.0)	26.0 (0.0)	32.0 (0.0)	40.0 (0.0)
Graphical (G) Statistics (compared with correct)	18.5 (5.4) $t(25) = -1.4$, <i>ns</i> , $r = 0.27^b$	19.9 (5.3) $t(25) = -2.0$, $p < .05$, $r = 0.37^c$	21.9 (5.6) $t(25) = -3.7$, $p < .001$, $r = 0.59$	24.5 (6.7) $t(25) = -5.7$, $p < .001$, $r = 0.75$	27.6 (9.0) $t(25) = -7.0$, $p < .001$, $r = 0.81$
Text (T) Statistics (compared with correct)	18.5 (5.4) $t(25) = -1.4$, <i>ns</i> , $r = 0.27$	20.0 (5.3) $t(25) = -1.9$, <i>ns</i> , $r = 0.36$	22.1 (5.6) $t(25) = -3.6$, $p < .01$, $r = 0.58$	24.7 (6.7) $t(25) = -5.5$, $p < .001$, $r = 0.74$	27.9 (9.0) $t(25) = -6.9$, $p < .001$, $r = 0.81$
Physical (P) Statistics (compared with correct)	20.0 (0.0) $t(25) = 0.0$, <i>ns</i> , $r = 0.00$	22.0 (0.0) $t(25) = 0.0$, <i>ns</i> , $r = 0.00$	25.8 (0.7) $t(25) = -1.8$, <i>ns</i> , $r = 0.34$	31.3 (2.0) $t(25) = -1.8$, <i>ns</i> , $r = 0.34$	38.6 (3.9) $t(25) = -1.8$, <i>ns</i> , $r = 0.34$
Decreasing problem					
Correct	28.0 (0.0)	34.0 (0.0)	38.0 (0.0)	40.0 (0.0)	40.0 (0.0)
Graphical (G) Statistics (compared with correct)	19.0 (8.9) $t(17) = -4.3$, $p < .001$, $r = 0.72$	20.1 (12.4) $t(17) = -4.7$, $p < .001$, $r = 0.75$	20.4 (15.2) $t(17) = -4.9$, $p < .001$, $r = 0.77$	20.0 (17.1) $t(17) = -5.0$, $p < .001$, $r = 0.77$	18.8 (18.1) $t(17) = -5.0$, $p < .001$, $r = 0.77$
Text (T) Statistics (compared with correct)	18.0 (9.2) $t(17) = -4.6$, $p < .001$, $r = 0.74$	19.4 (13.2) $t(17) = -4.7$, $p < .001$, $r = 0.75$	19.9 (16.0) $t(17) = -4.8$, $p < .001$, $r = 0.76$	19.3 (17.8) $t(17) = -4.9$, $p < .001$, $r = 0.77$	17.8 (18.7) $t(17) = -5.0$, $p < .001$, $r = 0.77$
Physical (P) Statistics (compared with correct)	22.0 (8.2) $t(17) = -3.1$, $p < .01$, $r = 0.60$	25.1 (11.9) $t(17) = -3.2$, $p < .01$, $r = 0.61$	27.0 (14.8) $t(17) = -3.2$, $p < .01$, $r = 0.61$	27.7 (16.7) $t(17) = -3.1$, $p < .01$, $r = 0.60$	27.1 (17.6) $t(17) = -3.1$, $p < .01$, $r = 0.60$

^aNote: The values in bracket represent the standard deviation about the mean.

^bThe value indicates the effect size.

^cNumbers in boldface indicate results that are significant at p less than or equal to .05.

Table 5. Proportion of responses classified as relying on the correlation heuristic (CH) in different representations and problems in Experiment 2. Comparison statistics with the correct accumulation's CH value (=0%) are also shown

Increasing problem		
Representation and education	CH (%)	Statistics (compared with correct)
Correct	00	—
Graphical (G)	73	$t(25) = 8.2$, $p < .001$, $r = 0.85^a$
Text (T)	69	$t(25) = 7.5$, $p < .001$, $r = 0.83^b$
Physical (P)	12	$t(25) = 1.8$, <i>ns</i> , $r = 0.34$
Decreasing problem		
Treatment	CH (%)	Statistics (compared with correct)
Correct	00	—
Graphical (G)	61	$t(17) = 5.2$, $p < .001$, $r = 0.78$
Text (T)	56	$t(17) = 4.6$, $p < .001$, $r = 0.74$
Physical (P)	33	$t(17) = 2.9$, $p < .01$, $r = 0.58$

^aNote: The value indicates the effect size.

^bNumbers in boldface indicate results that are significant at p less than or equal to .05.

the graphical and text representations (see Table A4 in the Appendix). Again, in both problems, both the undergraduates' and graduates' accumulation were generally more accurate in the physical representation (see Table A5 in the Appendix).

Similarly, the proportion of responses relying on CH among STEM and non-STEM backgrounds in the physical representation was lower (see Table A6 in the Appendix). There appears to be a similar relationship of differences among graduate and undergraduate levels of education (see Table A6 in the Appendix).

Discussion

Our results indicate that the physical representation is more effective compared with both text and graphical representations across different problem types and is effective in reducing participants' reliance on CH. The decrease in CH reliance in the physical representation was despite the fact that the relationship between the accumulation and the inflow is counterintuitive in the decreasing problem: Emissions decrease linearly, although the actual CO₂ accumulation continues to increase nonlinearly over time. This is observed in both the graphical and text representations; the average sketched accumulation over the five periods had a negative slope. In the physical representation, however, the slope of the average sketched accumulation over the five periods was positive. Taken together, these findings reinforce the effectiveness of the physical representation in different problems. Its effectiveness makes it suitable for use in different kinds of dynamic problems, where the resulting shape of the inflow might be different over time.

GENERAL DISCUSSION

People underestimate the correct value of nonlinear accumulation given their reliance on the CH for different contexts and problems. The use of a physical representation, however, can help reduce their underestimation and reliance on the CH in nonlinear accumulation problems, regardless of the problem’s dynamics or context and participants’ educational backgrounds and levels.

A physical representation may motivate participants to be involved in the problem they are solving and may help them construct a correct representation of the problem. Research that has investigated physical representation in the past also has similar conclusions about the representation’s effectiveness (Cooper & Harries, 2002, 2003; De Lange, 1987; Freudenthal, 1983; Palm, 2002). Given its effectiveness here and the documented benefits in mathematical education and judgment and decision-making literature (Aprea & Ebner, 1999; De Corte et al., 1996; Galesic et al., 2009; Garcia-Retamero & Dhimi, 2011; Garcia-Retamero & Galesic, 2010; Garcia-Retamero et al., 2010), physical representations will also be effective in reducing CH reliance in a wider range of nonlinear problems and contexts.

Generally, one could imagine many different physical representations in nonlinear problems that depend upon what needs to be communicated to a decision maker. For example, an icon array has been used to effectively communicate fractions, probability, or risk (Galesic et al., 2009; Garcia-Retamero & Dhimi, 2011; Garcia-Retamero & Galesic, 2010; Garcia-Retamero et al., 2010). The physical representation used here provided a picture snapshot of the inflows and the outflows at any instance of time. A possible limitation of our representation is that it can be impractical for problems with a larger time scale. One could depict the changes in flows only for the first few and last few periods in problems with larger time scales but may still reap the benefits of using a physical representation. By doing so, one is more likely to be able to communicate a correct understanding of accumulation in a problem.

In this paper, the physical representation is shown to be effective in both climate and non-climate contexts. Therefore, its use has an immense potential to improving public understanding in these contexts. Because the current wait-

and-see policies are likely to cause catastrophic changes in the near future (Alley et al., 2003; Scheffer et al., 2001), however, our findings are particularly relevant to research on wait-and-see behavior for climate change. With wait-and-see behavior, people prefer to delay policy actions that mitigate climate change to a future time. Changes in the atmospheric CO₂ accumulation (which is one of the main contributors for climate change) have been increasing nonlinearly for many years since the Industrial Revolution (IPCC, 2007). Furthermore, the accumulation’s nonlinear increase is predicted to intensify over the next 50–60 years (IPCC, 2007), with adverse consequences such as temperature change, melting of polar ice caps, and rising sea levels (Alley et al., 2003; Scheffer et al., 2001). As seen in our findings, people tend to underestimate nonlinear changes in the CO₂ accumulation when inferring from text and graphical representations. Because text and graphical representations are commonly used (IPCC, 2001a; EPA, 2010; Schiermeier, 2010), people are likely to underestimate the CO₂ accumulation’s nonlinear change and inaccurately infer much less CO₂ in the atmosphere than predicted by climate scientists. Underestimating the accumulation is likely to undermine the importance of the climate problem and could result in people’s wait-and-see behavior. A physical representation could be used to represent the CO₂ accumulation information and communicate the dynamics of Earth’s climate. Using the physical representation over other forms such as text or graphs could improve our estimation on atmospheric CO₂ and its associated climate change, ultimately reducing people’s wait-and-see behavior.

Let us return to the fable about the inventor of chess and the king. By now, it becomes easy to explain why the king thought that the inventor’s request was *modest*: He was unable to foresee how the accumulation of rice was nonlinear. He underestimated the rice accumulation, much like our participants did on nonlinear accumulation problems. But the king recognized the cleverness of the inventor’s request halfway through fulfilling it. The actual “physical exercise” of putting the rice grains on the chess board’s squares was an example of a physical representation. As shown by the current research, the use of physical representation enables people to overcome underestimation of accumulation in nonlinear problems.

APPENDICES

APPENDIX A1: THE CORRECT ACCUMULATION IN DIFFERENT PERIODS AND THE CORRESPONDING AVERAGE ACCUMULATION IN THE TWO REPRESENTATIONS SPLIT BY STEM AND NON-STEM BACKGROUNDS IN EXPERIMENT 1

Representation and backgrounds	Time 1	Time 2	Time 3	Time 4	Time 5
Correct	10.0 (0.0)	12.0 (0.0)	16.0 (0.0)	22.0 (0.0)	30.0 (0.0)
Graphical–non-STEM ¹ Statistics (compared with correct)	11.3 (2.3)² <i>t</i> (22) = 2.3 , <i>p</i> < .05 , <i>r</i> = 0.44^{3, 4}	13.7 (4.7) <i>t</i> (22) = 1.7, <i>ns</i> , <i>r</i> = 0.34	16.7 (6.2) <i>t</i> (22) = 0.5, <i>ns</i> , <i>r</i> = 0.11	20.0 (7.6) <i>t</i> (22) = –1.3, <i>ns</i> , <i>r</i> = 0.27	24.3 (9.0) <i>t</i> (22) = – 3.0 , <i>p</i> < .01 , <i>r</i> = 0.54

(Continues)

APPENDIX A1: (Continued)

Representation and backgrounds	Time 1	Time 2	Time 3	Time 4	Time 5
Graphical-STEM Statistics (compared with correct)	10.0 (2.0) $t(43)=0.1,$ <i>ns, r=0.02</i>	12.1 (2.3) $t(43)=0.2,$ <i>ns, r=0.03</i>	15.2 (2.9) $t(43)=-1.8,$ <i>ns, r=0.26</i>	19.5 (5.6) $t(43)=-3.9,$ $p < .001, r=0.51$	24.9 (7.7) $t(43)=-4.9,$ $p < .001, r=0.60$
Physical-non-STEM Statistics (compared with correct)	09.7 (1.6) $t(24)=-0.9,$ <i>ns, r=0.18</i>	11.7 (1.6) $t(24)=-1.0,$ <i>ns, r=0.20</i>	15.5 (2.1) $t(24)=-1.3,$ <i>ns, r=0.26</i>	21.1 (3.2) $t(24)=-1.4,$ <i>ns, r=0.27</i>	28.6 (4.9) $t(24)=-1.4,$ <i>ns, r=0.27</i>
Physical-STEM Statistics (compared with correct)	10.1 (0.7) $t(36)=1.0,$ <i>ns, r=0.16</i>	12.1 (0.8) $t(36)=0.8,$ <i>ns, r=0.13</i>	16.0 (0.8) $t(36)=0.0,$ <i>ns, r=0.00</i>	21.7 (1.4) $t(36)=-1.3,$ <i>ns, r=0.21</i>	29.3 (2.6) $t(36)=-1.7,$ <i>ns, r=0.57</i>

Note: ¹STEM-science, technology, engineering, and management.
²The values in bracket represent the standard deviation about the mean.
³The value indicates the effect size.
⁴Numbers in boldface indicate results that are significant at p less than or equal to .05.

APPENDIX A2: THE CORRECT ACCUMULATION IN DIFFERENT PERIODS AND THE CORRESPONDING AVERAGE ACCUMULATION IN THE TWO REPRESENTATIONS SPLIT BY GRADUATE AND UNDERGRADUATE LEVELS OF EDUCATION IN EXPERIMENT 1

Representation and level of education	Time 1	Time 2	Time 3	Time 4	Time 5
Correct	10.0 (0.0)	12.0 (0.0)	16.0 (0.0)	22.0 (0.0)	30.0 (0.0)
Graphical-undergraduate Statistics (compared with correct)	10.7 (2.8) ¹ $t(36)=1.6,$ <i>ns, r=0.04</i> ²	13.3 (4.2) $t(36)=1.8,$ <i>ns, r=0.05</i>	16.7 (5.1) $t(36)=0.8,$ <i>ns, r=0.02</i>	20.8 (5.8) $t(36)=-1.2,$ <i>ns, r=0.03</i>	26.0 (7.2) $t(36)=-3.4,$ $p < .001, r=0.09$ ³
Graphical-graduate Statistics (compared with correct)	10.0 (0.8) $t(29)=0.0,$ <i>ns, r=0.00</i>	11.8 (1.6) $t(29)=-0.6,$ <i>ns, r=0.02</i>	14.6 (2.8) $t(29)=-2.8,$ $p < .01, r=0.10$	18.2 (5.1) $t(29)=-4.1,$ $p < .001, r=0.14$	23.1 (8.0) $t(29)=-4.7,$ $p < .001, r=0.16$
Physical-undergraduate Statistics (compared with correct)	09.8 (1.4) $t(34)=-0.9,$ <i>ns, r=0.03</i>	11.7 (1.4) $t(34)=-1.1,$ <i>ns, r=0.03</i>	15.6 (5.1) $t(34)=-1.4,$ <i>ns, r=0.04</i>	21.2 (2.8) $t(34)=-1.7,$ <i>ns, r=0.05</i>	28.7 (4.4) $t(34)=-1.7,$ <i>ns, r=0.05</i>
Physical-graduate Statistics (compared with correct)	10.1 (0.7) $t(28)=1.0,$ <i>ns, r=0.04</i>	12.1 (1.0) $t(28)=0.8,$ <i>ns, r=0.03</i>	16.0 (1.0) $t(28)=-0.2,$ <i>ns, r=0.01</i>	21.6 (1.8) $t(28)=-1.3,$ <i>ns, r=0.05</i>	29.0 (3.4) $t(28)=-1.6,$ <i>ns, r=0.06</i>

Note:¹The values in bracket represent the standard deviation about the mean.
²The value indicates the effect size.
³Numbers in boldface indicate results that are significant at p less than or equal to .05.

APPENDIX A3: PROPORTION OF RESPONSES CLASSIFIED AS RELYING ON THE CORRELATION HEURISTIC (CH) IN DIFFERENT TREATMENTS FOR DIFFERENT EDUCATIONAL BACKGROUNDS AND LEVELS OF EDUCATION IN EXPERIMENT 1. COMPARISON STATISTICS WITH THE CORRECT ACCUMULATION'S CH VALUE (=0%) ARE ALSO SHOWN

Treatment	CH (%)	Statistics (compared with correct)
Correct	00	—
Graphical-non-STEM ¹	65	$t(22)=6.4, p < .001, r=0.81$ ^{2,3}
Graphical-STEM	45	$t(43)=6.0, p < .001, r=0.68$
Physical-non-STEM	08	$t(24)=1.4, ns, r=0.27$

(Continues)

APPENDIX A3: (Continued)

Treatment	CH (%)	Statistics (compared with correct)
Physical–STEM	08	$t(36) = 1.8, ns, r = 0.29$
Graphical–undergraduate	54	$t(36) = 6.5, p < .001, r = 0.18$
Graphical–graduate	50	$t(29) = 5.4, p < .001, r = 0.18$
Physical–undergraduate	08	$t(34) = 1.8, ns, r = 0.05$
Physical–graduate	10	$t(28) = 1.4, ns, r = 0.05$

Note: ¹STEM-science, technology, engineering, and management.

²The value indicates the effect size.

³Numbers in boldface indicate results that are significant at p less than or equal to .05.

APPENDIX A4: THE CORRECT ACCUMULATION IN DIFFERENT PERIODS AND THE CORRESPONDING AVERAGE ACCUMULATION IN DIFFERENT REPRESENTATIONS AND PROBLEMS SPLIT BY EDUCATION BACKGROUND IN EXPERIMENT 2

Representation and backgrounds	Time 1	Time 2	Time 3	Time 4	Time 5
	Increasing problem				
Correct	20.0 (0.0)	22.0 (0.0)	26.0 (0.0)	32.0 (0.0)	40.0 (0.0)
Graphical–non-STEM	20.0 (0.0)	21.4 (0.5)	23.1 (1.6)	25.1 (3.3)	27.4 (5.9)
Statistics (compared with correct)	$t(06) = 0.0, ns, r = 0.00$	$t(06) = -2.8, p < .05, r = 0.75$	$t(06) = -4.8, p < .01, r = 0.89$	$t(06) = -5.4, p < .01, r = 0.91$	$t(06) = -5.7, p < .001, r = 0.92$
Graphical–STEM ¹	17.9 (6.3) ²	19.3 (6.1)	21.5 (6.4)	24.3 (7.6)	27.7 (10.0)
Statistics (compared with correct)	$t(18) = -1.5, ns, r = 0.33^3$	$t(18) = -1.9, ns, r = 0.41$	$t(18) = -3.1, p < .01, r = 0.59^4$	$t(18) = -4.4, p < .001, r = 0.72$	$t(18) = -5.4, p < .001, r = 0.79$
Text–non-STEM	20.0 (0.0)	21.0 (0.7)	22.2 (1.4)	23.0 (2.1)	24.0 (2.8)
Statistics (compared with correct)	$t(04) = 0.0, ns, r = 0.00$	$t(04) = -3.2, p < .05, r = 0.85$	$t(04) = -6.3, p < .01, r = 0.95$	$t(04) = -9.5, p < .001, r = 0.98$	$t(04) = -12.6, p < .001, r = 0.99$
Text–STEM	18.1 (6.0)	19.8 (5.9)	22.1 (6.2)	25.1 (7.4)	28.8 (9.8)
Statistics (compared with correct)	$t(20) = -1.5, ns, r = 0.32$	$t(20) = -1.7, ns, r = 0.36$	$t(20) = -2.9, p < .01, r = 0.54$	$t(20) = -4.3, p < .001, r = 0.69$	$t(20) = -5.3, p < .001, r = 0.76$
Physical–non-STEM	20.0 (0.0)	22.0 (0.0)	25.9 (0.6)	31.5 (1.7)	39.1 (3.3)
Statistics (compared with correct)	$t(12) = 0.0, ns, r = 0.00$	$t(12) = 0.0, ns, r = 0.00$	$t(12) = -1.0, ns, r = 0.28$	$t(12) = -1.0, ns, r = 0.28$	$t(12) = -1.0, ns, r = 0.28$
Physical–STEM	20.0 (0.0)	22.0 (0.0)	25.7 (0.8)	31.1 (2.3)	38.2 (4.5)
Statistics (compared with correct)	$t(12) = 0.0, ns, r = 0.00$	$t(12) = 0.0, ns, r = 0.00$	$t(12) = -1.5, ns, r = 0.40$	$t(12) = -1.5, ns, r = 0.40$	$t(12) = -1.5, ns, r = 0.40$
	Decreasing problem				
Correct	28.0 (0.0)	34.0 (0.0)	38.0 (0.0)	40.0 (0.0)	40.0 (0.0)
Graphical–non-STEM	15.7 (7.2)	15.0 (10.1)	14.0 (12.3)	12.7 (14.0)	11.0 (14.7)
Statistics (compared with correct)	$t(05) = -4.2, p < .01, r = 0.88$	$t(05) = -4.6, p < .01, r = 0.90$	$t(05) = -4.7, p < .01, r = 0.90$	$t(05) = -4.8, p < .01, r = 0.91$	$t(05) = -4.8, p < .01, r = 0.91$
Graphical–STEM	20.7 (9.4)	22.7 (13.1)	23.7 (16.0)	23.7 (17.9)	22.7 (18.9)
Statistics (compared with correct)	$t(11) = -2.7, p < .05, r = 0.63$	$t(11) = -3.0, p < .05, r = 0.67$	$t(11) = -3.1, p < .05, r = 0.68$	$t(11) = -3.2, p < .01, r = 0.69$	$t(11) = -3.2, p < .01, r = 0.69$

(Continues)

APPENDIX A4: (Continued)

Representation and backgrounds	Time 1	Time 2	Time 3	Time 4	Time 5
Text–non-STEM	13.0 (7.3)	12.3 (10.6)	11.3 (13.1)	10.0 (14.7)	8.3 (15.5)
Statistics (compared with correct)	$t(05) = -5.0$, $p < .01$, $r = 0.91$	$t(05) = -5.0$, $p < .01$, $r = 0.91$	$t(05) = -5.0$, $p < .01$, $r = 0.91$	$t(05) = -5.0$, $p < .01$, $r = 0.91$	$t(05) = -5.0$, $p < .01$, $r = 0.91$
Text–STEM	20.5 (9.3)	23.0 (13.3)	24.2 (16.1)	24.0 (18.0)	22.5 (18.9)
Statistics (compared with correct)	$t(11) = -2.8$, $p < .05$, $r = 0.65$	$t(11) = -2.9$, $p < .05$, $r = 0.66$	$t(11) = -3.0$, $p < .05$, $r = 0.67$	$t(11) = -3.1$, $p < .01$, $r = 0.68$	$t(11) = -3.2$, $p < .01$, $r = 0.69$
Physical–non-STEM	19.0 (7.3)	20.0 (11.0)	20.5 (13.8)	20.5 (15.9)	20.0 (17.2)
Statistics (compared with correct)	$t(03) = -2.4$, ns , $r = 0.81$	$t(03) = -2.5$, ns , $r = 0.82$	$t(03) = -2.5$, ns , $r = 0.82$	$t(03) = -2.5$, ns , $r = 0.82$	$t(03) = -2.3$, ns , $r = 0.80$
Physical–STEM	22.9 (8.4)	26.6 (12.2)	28.9 (15.0)	29.7 (16.9)	29.1 (18.9)
Statistics (compared with correct)	$t(13) = -2.3$, $p < .05$, $r = 0.54$	$t(13) = -2.3$, $p < .05$, $r = 0.54$	$t(13) = -2.3$, $p < .05$, $r = 0.54$	$t(13) = -2.3$, $p < .05$, $r = 0.54$	$t(13) = -2.3$, $p < .05$, $r = 0.54$

Note: ¹STEM-science, technology, engineering, and management.

²The values in bracket represent the standard deviation about the mean.

³The value indicates the effect size.

⁴Numbers in boldface indicate results that are significant at p less than or equal to .05.

APPENDIX A5: THE CORRECT ACCUMULATION IN DIFFERENT PERIODS AND THE CORRESPONDING AVERAGE ACCUMULATION IN DIFFERENT REPRESENTATIONS AND PROBLEMS SPLIT BY LEVELS OF EDUCATION IN EXPERIMENT 2

Representation and level of education	Time 1	Time 2	Time 3	Time 4	Time 5
		Increasing problem			
Correct	20.0 (0.0)	22.0 (0.0)	26.0 (0.0)	32.0 (0.0)	40.0 (0.0)
Graphical–undergraduate	20.0 (0.0) ¹	21.3 (0.6)	23.1 (1.7)	25.4 (3.7)	28.1 (6.6)
Statistics (compared with correct)	$t(13) = 0.0$, ns , $r = 0.00$	$t(13) = -4.4$, $p < .001$, $r = 0.77^{2,3}$	$t(13) = -6.3$, $p < .001$, $r = 0.87$	$t(13) = -6.6$, $p < .001$, $r = 0.88$	$t(13) = -6.7$, $p < .001$, $r = 0.88$
Graphical–graduate	16.7 (7.8)	18.3 (7.6)	20.5 (7.9)	23.4 (9.1)	27.0 (11.5)
Statistics (compared with correct)	$t(11) = -1.5$, ns , $r = 0.41$	$t(11) = -1.7$, ns , $r = 0.46$	$t(11) = -2.4$, $p < .05$, $r = 0.59$	$t(11) = -3.3$, $p < .01$, $r = 0.71$	$t(11) = -3.9$, $p < .01$, $r = 0.76$
Text–undergraduate	20.0 (0.0)	21.5 (0.5)	23.5 (1.8)	26.2 (4.0)	29.4 (7.2)
Statistics (compared with correct)	$t(16) = 0.0$, ns , $r = 0.00$	$t(16) = -4.2$, $p < .001$, $r = 0.72$	$t(16) = -5.6$, $p < .001$, $r = 0.81$	$t(16) = -6.0$, $p < .001$, $r = 0.83$	$t(16) = -6.1$, $p < .001$, $r = 0.84$
Text–graduate	15.6 (8.8)	17.2 (8.7)	19.3 (8.9)	21.9 (9.8)	24.9 (11.6)
Statistics (compared with correct)	$t(08) = -1.5$, ns , $r = 0.47$	$t(08) = -1.7$, ns , $r = 0.52$	$t(08) = -2.3$, $p < .05$, $r = 0.63$	$t(08) = -3.1$, $p < .05$, $r = 0.74$	$t(08) = -3.9$, $p < .01$, $r = 0.81$
Physical–undergraduate	20.0 (0.0)	22.0 (0.0)	26.0 (0.0)	32.0 (0.0)	40.0 (0.0)
Statistics (compared with correct)	$t(13) = 0.0$, ns , $r = 0.00$	$t(13) = 0.0$, ns , $r = 0.00$	$t(13) = 0.0$, ns , $r = 0.00$	$t(13) = 0.0$, ns , $r = 0.00$	$t(13) = 0.0$, ns , $r = 0.00$
Physical–graduate	20.0 (0.0)	22.0 (0.0)	25.5 (0.9)	30.5 (2.7)	37.0 (5.4)
Statistics (compared with correct)	$t(11) = 0.0$, ns , $r = 0.00$	$t(11) = 0.0$, ns , $r = 0.00$	$t(11) = -1.9$, ns , $r = 0.50$	$t(11) = -1.9$, ns , $r = 0.50$	$t(11) = -1.9$, ns , $r = 0.50$

(Continues)

APPENDIX A5: (Continued)

Representation and level of education	Time 1	Time 2	Time 3	Time 4	Time 5
		Decreasing problem			
Correct	28.0 (0.0)	34.0 (0.0)	38.0 (0.0)	40.0 (0.0)	40.0 (0.0)
Graphical–undergraduate Statistics (compared with correct)	16.5 (7.9) $t(07) = -4.1,$ $p < .01, r = 0.51$	16.5 (11.4) $t(07) = -4.4,$ $p < .01, r = 0.53$	16.0 (14.0) $t(07) = -4.4,$ $p < .01, r = 0.53$	15.0 (15.8) $t(07) = -4.5,$ $p < .01, r = 0.54$	13.5 (16.7) $t(07) = -4.5,$ $p < .01, r = 0.54$
Graphical–graduate Statistics (compared with correct)	21.0 (9.5) $t(09) = -2.5,$ $p < .05, r = 0.27$	23.0 (13.0) $t(09) = -2.7,$ $p < .05, r = 0.29$	24.0 (15.9) $t(09) = -2.8,$ $p < .05, r = 0.30$	24.0 (17.9) $t(09) = -2.8,$ $p < .05, r = 0.30$	23.0 (18.9) $t(09) = -2.8,$ $p < .05, r = 0.30$
Text–undergraduate Statistics (compared with correct)	14.5 (8.3) $t(07) = -4.6,$ $p < .01, r = 0.55$	14.5 (12.0) $t(07) = -4.6,$ $p < .01, r = 0.55$	14.0 (14.8) $t(07) = -4.6,$ $p < .01, r = 0.55$	13.0 (16.7) $t(07) = -4.6,$ $p < .01, r = 0.55$	11.5 (17.6) $t(07) = -4.6,$ $p < .01, r = 0.55$
Text–graduate Statistics (compared with correct)	20.8 (9.3) $t(09) = -2.4,$ $p < .05, r = 0.26$	23.4 (13.3) $t(09) = -2.4,$ $p < .05, r = 0.26$	24.6 (16.1) $t(09) = -2.6,$ $p < .05, r = 0.28$	24.4 (17.9) $t(09) = -2.8,$ $p < .05, r = 0.30$	22.8 (18.9) $t(09) = -2.9,$ $p < .05, r = 0.31$
Physical–undergraduate Statistics (compared with correct)	25.4 (4.4) $t(06) = -1.5,$ $ns, r = 0.24$	29.7 (7.5) $t(06) = -1.5,$ $ns, r = 0.24$	32.6 (9.7) $t(06) = -1.5,$ $ns, r = 0.24$	34.0 (11.0) $t(06) = -1.4,$ $ns, r = 0.23$	34.0 (11.5) $t(06) = -1.4,$ $ns, r = 0.23$
Physical–graduate Statistics (compared with correct)	19.8 (9.4) $t(10) = -2.9,$ $p < .05, r = 0.28$	22.2 (13.6) $t(10) = -2.9,$ $p < .05, r = 0.28$	23.5 (16.7) $t(10) = -2.9,$ $p < .05, r = 0.28$	23.6 (18.8) $t(10) = -2.9,$ $p < .05, r = 0.28$	22.7 (19.8) $t(10) = -2.9,$ $p < .05, r = 0.28$

Note:

¹The values in bracket represent the standard deviation about the mean.

²The value indicates the effect size.

³Numbers in boldface indicate results that are significant at p less than or equal to .05.

APPENDIX A6: PROPORTION OF RESPONSES CLASSIFIED AS RELYING ON CORRELATION HEURISTIC (CH) IN DIFFERENT REPRESENTATIONS AND PROBLEMS SPLIT BY EDUCATION BACKGROUND AND LEVELS OF EDUCATION IN EXPERIMENT 2. COMPARISON STATISTICS WITH THE CORRECT ACCUMULATION'S CH VALUE (=0%) ARE ALSO SHOWN

	Increasing problem	
Representation and education	CH (%)	Statistics (compared with correct)
Correct	00	—
Graphical–non-STEM ¹	86	$t(06) = 6.0, p < .001, r = 0.93^{2,3}$
Graphical–STEM	68	$t(18) = 6.2, p < .001, r = 0.83$
Text–non-STEM	80	$t(04) = 4.0, p < .05, r = 0.89$
Text–STEM	67	$t(20) = 6.3, p < .001, r = 0.82$
Physical–non-STEM	08	$t(12) = 1.0, ns, r = 0.28$
Physical–STEM	15	$t(12) = 1.5, ns, r = 0.40$
Graphical–undergraduate	79	$t(13) = 6.9, p < .001, r = 0.89$
Graphical–graduate	67	$t(11) = 4.7, p < .001, r = 0.82$
Text–undergraduate	71	$t(16) = 6.2, p < .001, r = 0.84$
Text–graduate	67	$t(08) = 4.0, p < .01, r = 0.82$
Physical–undergraduate	00	$t(13) = 0.0, ns, r = 0.00$
Physical–graduate	25	$t(11) = 1.9, ns, r = 0.50$

(Continues)

Treatment	Decreasing problem	
	CH (%)	Statistics (compared with correct)
Correct	00	
Graphical–non-STEM	83	$t(05) = 5.0, p < .01, r = 0.91$
Graphical–STEM	50	$t(11) = 3.3, p < .01, r = 0.71$
Text–non-STEM	83	$t(05) = 5.0, p < .01, r = 0.91$
Text–STEM	42	$t(11) = 2.8, p < .05, r = 0.65$
Physical–non-STEM	50	$t(03) = 1.7, ns, r = 0.70$
Physical–STEM	29	$t(13) = 2.3, p < .05, r = 0.54$
Graphical–undergraduate	75	$t(07) = 4.6, p < .01, r = 0.55$
Graphical–graduate	50	$t(09) = 3.0, p < .05, r = 0.32$
Text–undergraduate	75	$t(07) = 4.6, p < .01, r = 0.55$
Text–graduate	40	$t(09) = 2.4, p < .05, r = 0.26$
Physical–undergraduate	14	$t(06) = 1.0, ns, r = 0.16$
Physical–graduate	45	$t(10) = 2.9, p < .05, r = 0.28$

Note: ¹STEM-science, technology, engineering, and management.

²The value indicates the effect size.

³Numbers in boldface indicate results that are significant at p less than or equal to .05.

ACKNOWLEDGEMENTS

This research was partially supported by the National Science Foundation (Human and Social Dynamics: Decision, Risk, and Uncertainty, Award Number 0624228) award to Professor Cleotilde Gonzalez. We also thank Hau-yu Wong, Dynamic Decision Making Laboratory, Carnegie Mellon University, for providing editorial comments to an initial version of the manuscript.

REFERENCES

- Alley, R., Marotzke, J., Nordhaus, W., Overpeck, J., Peteet, D., Pielke, R., . . . Wallace, J. (2003). Abrupt climate change. *Science*, 299, 2005–2010. DOI: 10.1126/science.1081056
- Almor, A., & Sloman, S. A. (2000). Reasoning versus memory in the Wason selection task—A non-deontic perspective on the perspective effects. *Memory and Cognition*, 28, 1060–1069. DOI: 10.1.1.25.4190
- Apra, C., & Ebner, H. J. G. (1999, August). *The impact of active graphical representation on the acquisition and application of knowledge in the context of business education*. Paper presented at the Eighth European Conference for Research on Learning and Instruction, Göteborg, Sweden.
- 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
- Brunstein, A., Gonzalez, G., & Kanter, S. (2010). Effects of domain experience in the stock–flow failure. *System Dynamics Review*, 26(4), 347–354. DOI: 10.1002/sdr.448
- Cooper, B., & Harries, T. (2002). Children’s responses to contrasting ‘realistic’ mathematics problems: Just how realistic are children ready to be? *Educational Studies in Mathematics*, 49, 1–23. DOI: 10.1023/A:1016013332659
- Cooper, B., & Harries, T. (2003). Children’s use of realistic considerations in problem solving: Some English evidence. *The Journal of Mathematical Behavior*, 22(4), 449–463. DOI: 10.1016/j.jmathb.2003.09.004
- Cronin, M., & Gonzalez, C. (2007). Understanding the building blocks of system dynamics. *System Dynamics Review*, 23(1), 1–17. DOI: 10.1002/sdr.356
- 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(1), 116–130. DOI: 10.1016/j.obhdp.2008.03.003
- De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2007). *The illusion of linearity: From analysis to improvement*. New York: Springer.
- De Corte, E., Greer, B., & Verschaffel, L. (1996). Psychology of mathematics teaching and learning. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 491–549). New York: Macmillan.
- De Lange, J. (1987). *Mathematics, insight, and meaning: Teaching, learning and testing of mathematics for the life and social sciences*. Utrecht: OW and OC.
- Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Gaming*, 11(1), 87–106. DOI:10.1177/104687818001100108
- Dörner, D., Kimber, R., & Kimber, R. (1997). *The logic of failure: Recognizing and avoiding error in complex situations*. Cambridge, MA: Basic Books.
- Dutt, V. (2011). Explaining human behavior in dynamic tasks through reinforcement learning. *Journal of Advances in Information Technology*, 2(3), 177–188. DOI:10.4304/jait.2.3.177-188
- Dutt, V., & Gonzalez, C. (2007). Slope of inflow impacts dynamic decision making. In *Proceedings of the 25th International Conference of the System Dynamics Society* (pp. 79). Boston, MA: System Dynamics Society.
- Dutt V., & Gonzalez, C. (in press). Human Control of Climate. *Climatic Change*. DOI: 10.1007/s10584-011-0202-x.
- Dutt V., & Gonzalez, C. (2010). Decisions from experience reduces “wait-and-see” behavior in climate change. Unpublished manuscript under review.
- EPA (2010). Past climate change. Retrieved from the EPA website: http://www.grida.no/publications/other/ipcc_tar/?src=climate/ipcc_tar/wg1/index.htm
- Evangelidou, A., Spyrou, P., Elia, I., & Gagatsis, A. (2004). University students’ conceptions of function. In M. Johnsen Høines, & A. Berit Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics*

- Education* (Vol. 2, pp. 351–358). Bergen, Norway: Bergen University College.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.
- Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks to low-numeracy people. *Health Psychology, 28*, 210–216. DOI: 10.1037/a0014474
- Garcia-Retamero, R., & Dhami, M. K. (2011). Pictures speak louder than numbers: On communicating medical risks to immigrants with limited non-native language proficiency. *Health Expectations, 14*, 46–57. DOI: 10.1111/j.1369-7625.2011.00670.x
- Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: Overcoming challenges in people's understanding of risks. *Social Science & Medicine, 70*, 1019–1025. DOI: 10.1016/j.socscimed.2009.11.031
- Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making, 30*(6), 672–684. DOI: 10.1177/0272989X10369000
- Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127–171. DOI: 10.1016/0010-0277(92)90060-U
- Gonzalez, C., & Dutt, V. (2007). Learning to control a dynamic task: A system dynamics cognitive model of the slope effect. In Lewis, Polk, & Laird (Eds.), *8th International Conference on Cognitive Modeling* (pp. 61–66). Ann Arbor, MI: Taylor & Francis.
- 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
- Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words? *Journal of Educational Psychology, 89*(4), 667–681. DOI: 10.1037/0022-0663.89.4.677
- IPCC (2001a). Climate change 2001: Working group I: The scientific basis. Retrieved from the IPCC website: http://www.grida.no/publications/other/ipcc_tar/?src=/climate/ipcc_tar/wg1/index.htm
- IPCC (2001b). Climate change 2001: Synthesis report (summary for policymakers). Retrieved from the IPCC website: <http://www.ipcc.ch/pdf/climate-changes-2001/synthesis-spm/synthesis-spm-en.pdf>.
- IPCC. (2007). Climate Change 2007: Synthesis Report. Contribution of working groups I, II and III to the fourth assessment report of the intergovernmental panel on climate change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC. Geneva, Switzerland.
- 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
- Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99. DOI: 10.1111/j.1551-6708.1987.tb00863.x
- Lebiere, C., Gonzalez, C., & Warwick, W. (2009). Convergence and constraints revealed in a qualitative model comparison. *Journal of Cognitive Engineering and Decision Making, 3*(2), 131–155. DOI: 10.1518/155534309X441880
- Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of Educational Research, 60*(1), 1–64. DOI: 10.3102/00346543060001001
- Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A., Buja, L. E., . . . Teng, H. (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.
- Palm, T. (2002). *The realism of mathematical school tasks: Features and consequences*. (Doctoral dissertation). Sweden: University of Umeå.
- 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
- Pólya, G. (1945). *How to solve it*. Princeton, NJ: Princeton University Press.
- 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
- Schiermeier, Q. (2010). The real holes in climate science. *Nature, 463*, 284–287. DOI: 10.1038/463284a
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition and sense-making in mathematics. In D. A. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York: MacMillan Publishing Company.
- Sterman, J. D. (2008). Risk communication on climate: Mental models and mass balance. *Science, 24*(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
- Van Deyck, B. (2001). *Correlatie en regressie: Een lesmodule voor de behandeling van een statistisch probleem in de derde graad van het secundair onderwijs*. (Master's Thesis). University of Leuven, Belgium.
- Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2007). Pupils' over reliance on linearity: A scholastic effect? *British Journal of Educational Psychology, 77*, 307–321. DOI: 10.1348/000709906X115967
- Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *The Journal of Educational Research, 83*(6), 301–312. DOI: 10.1234/12345678
- Wigley, T. (2005). The climate change commitment. *Science, 307*, 1766–1769. DOI: 10.1126/science.1103934

Authors' biographies:

Dr. Varun Dutt is a Post-Doctoral Fellow of the Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University. He is also the Knowledge Editor of the English daily, *Financial Chronicle*. His current research interests focus on risk perceptions in environmental decision making, dynamic decision making, and modeling of human behavior.

Dr. Cleotilde Gonzalez is an Associate Research Professor and Director of the Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University. She is an affiliated faculty at HCII, CCBI, and CNBC. She is in the Editorial board of the *Human Factors Journal* and the Associate Editor of the *JCEDM*.

Authors' addresses:

Varun Dutt and **Cleotilde Gonzalez**, Dynamic Decision Making Laboratory, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, USA.