Title
Climate Change and Conceptual Change

Author
Clark, David Joseph

Publication Date
2013

Peer reviewed|Thesis/dissertation

Climate Change and Conceptual Change

By

David J. Clark

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Michael A. Ranney, Co-chair
Professor Richard B. Ivry, Co-chair
Professor Sonia Bishop
Professor John Canny

Fall 2013

Climate Change and Conceptual Change

Abstract

Climate Change and Conceptual Change

by

David J. Clark

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Michael A. Ranney, Co-chair
Professor Richard B. Ivry, Co-chair

Global Warming[1] ("GW") is easily one of the most pressing concerns of our time, and its solution will come about only through a change in human behavior. Compared to the residents of most other nations worldwide, Americans report lower acceptance of the realities of GW. In order to address this concern in a free society, U.S. residents must be convinced or coerced to take the necessary actions. In spite of the democratic appeal of education, however, many climate communicators appear to be settling on the notion that emotional persuasion is superior to education.

We'll set an empirical foundation in Chapter 2, reviewing an experiment in the Numerically Driven Inferencing (NDI) paradigm that sheds some light on the cognitive processes involved in learning and attitude shifts in response to surprising policy-relevant information. Chapters 3–6 contain results from a comprehensive program of research specifically targeting climate-related attitudes and beliefs in the United States. As alluded to above, there have been many surveys of American attitudes. Chapter 3 provides an overview of our approach to assessing climate-related beliefs and attitudes. In particular, we note relationships observed in one survey between scientific literacy regarding the GW mechanism on one hand and attitudes, including "willingness to sacrifice" on the other. As with some other empirical approaches, our results suggest that U.S. residents generally accept anthropogenic (i.e., "human caused") climate change, and support action on this issue. But even if this is the case, Chapter 4 describes an experiment demonstrating that these beliefs and attitudes are disturbingly fragile in the face of cherry-picked, misleading numerical facts. Chapter 5 then describes a pair of experiments evaluating the effects of representative

[1] I (and many others) prefer the term "Climate Change." Even though changes in atmospheric chemistry will force the mean temperature of the globe higher, the effects will complicated by climate systems, resulting in—at least in the short term—local cooling in some places and a variety of other changes in weather patterns. However, as we'll be focusing primarily on the physical mechanisms by which the earth is warming, we'll use the more colloquial term "Global Warming" in this dissertation.

numerical facts. Chapter 5's Study 1 (Section 5.1) demonstrates that even when students report strong psychological effects after receiving a set of surprising numbers, their beliefs and attitudes will not necessarily be affected. Chapter 5's Study 2 (Section 5.2) improves upon the clarity of materials used in Study 1 and demonstrates that such materials *can* effectively increase climate change acceptance and concern. In both of these studies, as with the study presented in Chapter 4, this relatively uncontextualized, surprising numerical information undermines students' confidence in their own knowledge. Chapter 6 reports on three successful experiments (spanning four samples) that provide a coherent explanation of the mechanism of climate change that includes relevant numerical facts. As with Study 2 in Chapter 5, this intervention shifts participant attitudes towards the scientific consensus. Unlike uncontextualized numerical information, however, this mechanism intervention additionally leaves participants feeling that they know more than they did prior to instruction. Chapter 6's Study 1 (Section 6.1) establishes this effect in classroom-based settings at two culturally distinct universities. Chapter 6's Study 2 (Section 6.2) provides an initial evaluation of the time-course of retention for the cognitive shifts that followed our mechanism intervention, and Chapter 6's Study 3 (Section 6.3) provides a successful demonstration of durable shifts with the general population online.

Taken together, these experiments point the way towards effective curricula and on-line materials that can help bolster support to combat climate change. While we must certainly be sensitive to the needs, values, and interests of our target audiences, we should not reflexively steer away from science education. Indeed, the experiments in this dissertation provide empirical support for the notion that science education materials can have a meaningful and lasting impact on GW attitudes and beliefs. While this may not provide the complete behavioral solution we need for the United States (and the world), it seems likely that such shifts will make behavioral and policy changes far more tractable in the coming years.

With love and respect for all that is good—in particular the unwavering support of my grandma (even if she doesn't fully accept climate change *or* evolution).

Contents

List of Figures

List of Tables

Acknowledgments

The tone of the following dissertation is one of success and hope. This, however, is largely due to my attempts to satisfy the (perhaps imagined) constraints of getting my dissertation submitted. At this moment, I still find myself struggling to grapple with the daunting challenges of two systems that place little value on genuine long-term thinking: the global economic system, and the university system. Dear reader, I know you have to eat, but please do your best to pay attention to the long-term consequences of your actions. You might also ask your elected officials to do the same. And golly, do educators need our help. And that person in the chicken suit who just shut down the highway while you're trying to get to work? They're grappling with similar issues too. I don't think any of us have even one complete answer. So, dear reader, I'm glad you're joining me in the effort to find one.

Looking back, I must say that the brightest star in this particular journey was Tawny Tsang. Her continued cheer and critical evaluation of my sometimes muddled direction were more than I could ask for.

I have also had many teachers. I call particular attention to Larry Goldfarb, director for my training in the method developed by Moshe Feldenkrais. I have great appreciation for all of the teachers that assisted Larry as well. This work continues to provide me with inspiration for helping myself and others develop healthier habits. Thanks also to Connie Cronin for being my longish-term student—this allowed me to learn a lot. Looking back over my formal education, I continue to believe that I was almost unfortunately lucky in the guidance and attention I received as an undergraduate at the University of Maryland, College Park. Between Bonnie Dorr, Amy Weinberg, Phil Resnik, Eric Slud, David Lightfoot, and the other members of the CLIP lab, I managed to recieve an incredibly rich introduction to the formal study of human language. It took me a while to realize that this experience, sadly, wasn't normal! At MIT, I benefited from the instruction of Ken Wexler and Anthony Wagner, and at NYU, I learned from Lila Davachi and Yadin Dudai. I feel it's incumbent to mention my first conversation with Tom Wickens when I was trying to get out of taking his intro to statistics for psychology graduate students. He was quite reasonable about it, saying that one could always benefit from a different treatment of the material, but I wouldn't have any of it—I was still a punk kid at age 29. Fortunately, I had to take his class in the end, and it was great. Thanks to Rich Ivry and Michael Ranney for serving as co-chairs on my dissertation, and to Sonia Bishop for her helpful (and early!) comments.

Extra special thanks to John Canny, who gives me some small hope that perhaps one can do simultaneously useful *and* interesting things within the university context.

I was of course helped enormously by my friends and colleagues along the way. Lila Davachi gets a double mention here. Particular thanks to Tim Anderson, Meghan Ballog, Derek Clark, Sailor Neal, Dustin Stansbury, Emily Cooper, Boris Fain, Dan Walsh, David Grefrath, and the rest of the gang at Ft. Awesome. And you certainly can't make it through grad school without your fellow research group members. Thanks to John Schlerf, Flavio Oliviera and Jing Xu for shepherding me into the CognAc lab, and thanks to Becca Elenzil, Ryan Morehead, and Alex Carstensen for seeing me through. Daniel Reinholz was not only a stalwart of the Reasoning group, but also directly assisted with much of the research that follows. And a general thanks to members of the CognAc and Reasoning groups. Thanks also for the timely help of Eti Valdez-Kaminsky. And special appreciation for Cynthia Rauschert, who I feel has been on a parallel struggle with me for many years now. Hopefully, we're both coming out the other end!

And of course I owe a great debt to my family. My brother (Derek) gets the second double mention here, and you've met my grandma, Anne Eccleston, in the dedication. Then of course, there is my mother, Joanne Schmidt, and my father, Mark Clark. Much love also to my sister, Stephanie Clark, my grandpa, Joe Eccleston, and all my aunts, uncles, and cousins. Then there's the extended family I've gained through various marriages. Thanks in particular to Darcy Clark for welcoming me into her home to get me through some of these final struggles. Thanks also to John Schmidt and Linda Clark.

More recently, I am fortunate to have made the acquaintance of The team at Oroeco, and other climate crusaders. People have been trying and failing to shift human behavior towards greater sustainability for decades. It is at this point that I must simply check my modesty at the door and say that *this time* we're going to get it right.

Lastly, I am grateful beyond words for the arrival of Madoka Wada in my life. I had no idea someone could be wonderful like that.

This dissertation was based on the ucastrothesis template. It was typeset with LuaLaTeX using fonts from the TeX Gyre project (Pagella for main text and Heros for sans serif) and Inconsolata for monospaced text. The research reported here was supported in part by the Institute of Education Sciences pre-doctoral training grant R305B090026 to the University of California, Berkeley and the University of California, Berkeley Committee on Research. The opinions expressed are those of the author and do not necessarily represent views of the Institute of Education Sciences, the U.S. Department of Education, or the University of California.

Chapter 1

Introduction

Anthropogenic climate change is real, and it's a behavioral problem. Behavioral change is challenging, and can proceed along any of a number of routes. An increasingly vocal set of climate communications researchers argues that science education is not effective, and some even argue that it is *counter-productive*. However, as with many failures to observe a result in the behavioral sciences, it may simply be that effective climate-related science education, is simply very hard. Indeed, as we'll review below, there are ready examples of the effectiveness of science education—both in terms of knowledge gains and in terms of policy-relevant preferences. Therefore, the groundwork is well-laid for a careful consideration of science and numeracy interventions targeting climate change.

1.1 The Problem of Global Climate Change

Scientific controversy has been a major element of public discourse for the last half-century (for a comprehensive review, see Oreskes & Conway, 2010). Perhaps no issue sparks as much controversy today as global climate change (commonly referred to as global warming, from which we'll derive the abbreviation "GW"). As such, a psychological consideration of the problem of climate change entails some basic background in the fundamental climate science and a dash of moral reasoning. In the chapters that follow, we'll see empirical results supporting the utility of these basic science materials for the development of successful interventions.

1.1.1 We're All Going to Die

Apart from those lucky few deathless microbes who might catch a ride out of this solar system before the sun goes extinct, all organisms on earth today (or descended from such) will ultimately find themselves dead—even us. This is, in my opinion, no great cause for alarm, though arguments to this effect fall outside of the scope of this dissertation. (So, on this point, I must refer you to other sources of your own choosing.) I hope I needn't argue this basic fact—a fact that is necessary for clarity in what we might (or should) possibly hope to accomplish in mitigating climate change.

So, we cannot keep the peoples on Earth from dying or (barring the interstellar travel option) keep existing species from going extinct. These are certain eventualities. This may provide for some moral ambiguity, but for the purposes of this dissertation, we might adopt what I hope will be a relatively uncontroversial notion. Specifically, our actions today may lead to more or less suffering for ourselves, and (as we'll see below) particularly for children and their descendants. Likewise, our actions stand to have a marked impact on the potential richness of biological diversity on this planet. Howsoever our moral compass may direct us in the face of such realities, that compass cannot function without a basic orientation to the likely—or, in some cases, certain—consequences of our actions.

1.1.2 The Stark Reality of Climate Change

Many are familiar with the United Nations' International Panel on Climate Change (IPCC). This organization produces a thorough summary of the current science, although only every few years. Thus, for more up-to-date information, one can turn to organizations such as the World Bank. We find, for example, that our atmosphere's carbon dioxide (CO_2) concentration is higher now than in any of the past 15 million years (World Bank, 2012). A more recent report commissioned by the World Bank (Potsdam Institute for Climate Impact Research and Climate Analytics, 2013) finds that there is now a 40% chance that our world will be 4°C warmer by 2100. This warming would have harsh consequences for all people, but would disproportionately affect the poorest individuals. Some of these effects are already present and require adaptation or preparedness

Global warming akin to recent and projected trends last occurred over 17 million years ago, when a 3-4°C gain occurred over 1,500,000 years. Thus, the projected timescale of 100 years is over 10,000 times faster than previous timescales (Barnosky, 2009). In previous warming periods of this magnitude, widespread extinctions occurred (Mayhew, Jenkins, & Benton, 2008). We must therefore assume that we stand to lose a staggering number of species in the timescale of our children's or grandchildren's lifetimes. This would include species comprising the biological systems we all depend upon (e.g., for food).

Given the timecourse of carbon dioxide and other greenhouse gases in the atmosphere, the primary variable of interest will be total emissions until we reach a carbon neutral economy. Plausible best-case and worst-case scenarios are outlined by Archer and Brovkin (2008), and these predictions are stark: we have likely already "baked" a certain amount of warming into the climate system that will last for at least hundreds of thousands of years. If we do not aggressively change our current behavior, these changes will be on the order of 6°C or more (which, will obviously have effects worse than those from 4°C increases detailed in the World Bank reports mentioned above).

It is also necessary to point out the fact that there has been vocal opposition to the above arguments for action. Some of the more credible have attacked the notion that climate action is appropriate on economic grounds (e.g., Lomborg, 2007). Other scholars, like Richard

Muller, have claimed that the science is "uncertain" (although this view was recanted in a report produced by his own research initiative in Rohde et al., 2013). While a full rebuttal of these arguments is beyond the scope of this dissertation, it is also fortunately not necessary—citizen scientists have created publicly available resources that do this already (most notably the Skeptical Science website). Such opposition to science by entrenched interests is nothing new, and is well-chronicled by Oreskes and Conway (2010).

1.2 Climate Change as a Behavioral Problem

Nearly all climate researchers have concluded that this problem is urgent and anthropogenic (i.e., essentially 100% triggered and exacerbated by human behavior). As such, it will be "solved" only by changes in human behavior. J. Harte and M. E. Harte (2008) provide one such plan for action, detailing methods to rapidly reduce emissions. Schrag (2011) provides an alternative perspective on what will be required, underscoring the importance of considering total emissions before arriving at a carbon-neutral energy economy. While both focus on emissions, these plans have markedly different foci. Given such examples, it is clear that a large part of the necessary behavior will be discourse leading to an agreed and appropriate plan of action. Below, we'll see some arguments that greater public acceptance and political will beyond current levels will almost certainly be required to achieve followthrough on a reasonable plan.

1.2.1 Behavior is Embedded Within our Economic System

Hoffman (2010) suggests that while tobacco smoking provides a useful (and recent) example of a successful shift in science-related policies and attitudes, he goes on to suggest that slavery may provide a more apt model. A central differentiating feature between these two examples is that slavery was central to the ante-bellum economy, much as carbon-intensive energy production is now. One cannot simply substitute another vice, as one could with tobacco consumption. Rather, dramatic and deep changes must be made to the manner in which our current economic system functions, including the way we produce everything from food to luxury goods. Indeed, the availability of fossil fuels and other energy sources was likely critical to the shift away from slavery.

It is unlikely that a declaration of a carbon-neutral policy will result in anything approaching the conflict of the American civil war. It is clear, however, that there will continue to be strong opposition to such a policy from a wide variety of economic sectors. Thus, from a policy point of view, broad (and ultimately international) public agreement stands to play a key role in overcoming such opposition.

1.2.2 Americans are Weird

While humans continue to increase our understanding of the world, issues like climate change are not readily comprehended by non-specialists. The IPCC (Intergovernmental Panel on Climate Change) and Skeptical Science have assembled and disseminated the scientific consensus on GW, but, sadly, the U.S. public remains divided on both the existence and the cause of climate change (cf. Hoffman, 2011). Indeed, while much of the developed world accepts anthropogenic climate change as a reality, as of January 2010 only 57% of individuals surveyed in the United States think global warming is happening at all(Q47 in Leiserowitz, Maibach, & Roser-Renouf, 2010). When asked to assume that global warming *is* happening, only 47% of the same group of respondents indicated that they thought it was "caused mostly by human activities" (Q50 in Leiserowitz et al., 2010). Presumably, the number of Americans accepting *anthropogenic* climate change is somewhat less than this figure.

While some argue that Americans' acceptance and concern is somewhat higher than indicated above (e.g., Krosnick & MacInnis, 2013, report that 77% of Americans "found human activity responsible for warming" as of 2012), it seems indisputable that Americans are consistently outliers in this and many other dimensions. Multiple sources illustrate that Americans lag almost all other surveyed nations on GW acceptance. Ray and Pugliese (2010) report that the United States lags all of the 110 other nations in a Gallup telephone survey, with 47% of respondents reporting that "rising temperatures are a result of natural causes" (i.e., not the result of human causes). This sample included both peer nations and developing countries. Leiserowitz (2007) describes how even at the height of American GW acceptance (i.e., around the time of release of An Inconvenient Truth in 2006), we lagged much of the rest of the world.

Ranney (2012) provides a more comprehensive overview of American exceptionalism. A general problem identified within Ranney's Reinforced Theistic Manifest Destiny (RTMD) framework is the fact that individuals often reject scientific ideas when they are in conflict with their other attitudes and beliefs. It is as if we are endowed with something of a conceptual immune system, comprising religious and nationalistic beliefs in some, and more scientifically grounded beliefs in others. When an individual is "exposed" to ideas in opposition to their currently held beliefs, these beliefs serve as a defense against any potential conceptual changes (as also discussed by Shepherd & Kay, 2012).

Ranney mentions a number of similarities between evolution-related and climate-change-related cognition, but there is a sharp distinction between the kinds of "emotional responses" that might be experienced in response to these scientific ideas. Specifically, climate change is something that involves the ethical status of actions that we do every day, both individually and as a society. Evolution, on the other hand, tends to incohere with personally held religious beliefs, most directly divine creation, but then by extension, deity and afterlife (Ranney, 2012). Thus, Americans appear to require more convincing than citizens of other nations when it comes to scientific assertions, but there are likely nuances particular to each scientific domain—in particular, we should understand precisely where conceptual conflicts may arise. So, we may find some useful examples in successful evolution education interventions, but we must be mindful that the two domains are quite distinct in some ways.

1.3 Climate Communication Strategies

A group of climate communication researchers, oddly, suggests that educational ventures would be of little or no help. Kahan et al. (2012) found that, for the U.S. (a high per-capita carbon user), numeracy and scientific literacy were correlated with more biased views on GW. Specifically, amongst "hierarchical individualists" (Kahan's term that approximates "conservatives") science literacy was negatively correlated with acceptance of the reality of GW ($r = -0.12$, $p = 0.03$; note, this correlation was positive amongst "egalitarian communitarians"). Similarly, McCright and Dunlap (2011) highlight data indicating that the "education level" effect on climate belief is moderated by conservatism/affiliation (with conservative or "Republican" GW denial being slightly positively related, if at all, with education).[1] This (also correlational) evidence, they claim, disproves a naïve "knowledge deficit" view—that is, the view that more education can shift the public's beliefs toward the scientific consensus about climate change.

The above position harkens back to a classic social psychology report by Lord, Ross, and Lepper (1979), which reported that people with a strong position tended to polarize further after receiving information that was contrary to their views (though this information was not particularly factual). Interestingly, research from our own lab has refuted the applicability of this result to educational policy. The above study removed the middle third of individuals from the population, leaving only those with relatively strong views. However, Nelson (2007) shows that a persuasive intervention will in fact have the expected net effect if the population as a whole is considered. Certainly, a shift in the middle third of the United States on the issue of climate change would be sufficient to amass the kind of political will alluded to above.

We can even find hints for the development of successful science education interventions in the observations of McCright and Dunlap (2011). Specifically, this work indicates a bifurcation in the kinds of information that liberals and conservatives tend to receive. This split leaves open the possibility that well-constructed interventions may indeed induce conservatives to accept the scientific consensus (with little challenge to their core values).

It cannot be emphasized enough that while overreaching one's results may be close to required in today's academic publishing milieu, such a practice borders on irresponsible with respect to critical policy-relevant science. Certainly, one should be cautious in stretching a result from supporting "science education approaches are difficult" to "science education approaches *don't work*." The results of such publications are difficult to quantify, but one

[1] The reported statistical effects are significant interaction terms in a model predicting GW acceptance. In particular, significant interactions were observed between political ideology and political party on one hand, and educational attainment on the other.

can turn to the mainstream press to see how such results are reported. In a recent article in *Mother Jones*, Mooney (2011) provides a summary of much of the above research, arguing against "the standard notion that the way to persuade people is via evidence and argument." I can also offer an anecdote: I met with an individual working at a prominent science museum who was in the process of managing a four-city deployment of a climate-change exhibit. Based on her reading of the literature, they decided to err on the side of avoiding polarization, and excluded scientific descriptions of anthropogenic climate change. Thus, based on a reading of the current literature, *science museums* are deciding to *withhold science education* for fear of polarizing the public.

This kind of problem is not limited to science education, but rather seems to be part of a general trend. For example, Sachs et al. (2012), after evaluating hard-to-use versus easy-to-use programmable thermostats, report that their provision of easier-to-use thermostats had no significant effect on heating energy usage. Here, I must again resort to anecdote: this result was then presented at a conference I attended as definitive evidence that "programmable thermostats don't work." This is akin to exploring two iterations in an engineering design process, and then declaring that the problem is insoluble. We would have never succeeded with the moon shot, or with mobile handheld computing with such thinking![2] As outlined above, climate-relevant behavior change is a hard problem in general. Thus, a successful intervention is likely going to require much careful thought and iteration. In this document, we'll see an example of such a process.

1.4 Science and Numeracy Education for Climate Change

Taken more critically, the above-cited work supports the claim that *absent proper guidance*, individuals will not arrive at a complete understanding or acceptance of the current scientific consensus. Indeed, Lewandowsky, Gignac, and Vaughan (2013) show that offering climate scientists' consensus boosts anthropogenic climate change acceptance. Thus, as we'll argue below, a more sensible position is likely that science education is hard, but if done correctly, it can have useful effects. It's not hard to *want* to take an educational approach to the climate problem—it's likely the most democratic approach we can devise! In the face of the rhetoric described above, educators may already be withholding scientific materials for fear that they will be polarizing. In the studies presented in the following chapters, we'll see that even a small amount of true information can quickly act as a cognitive "lever" to enhance one's understanding and perspective on climate change.

Note both that new knowledge often facilitates societal shifts and that science "education" has historically driven major social changes—from heliocentrism replacing church doctrine to the acceptance of a tobacco-cancer link in spite of industry obfuscations. (We

[2]The Apple Newton, for example, was released in 1993—almost 10 years after the Psion Organizer was released in 1984. Even with Apple's legendary engineering capacity and existing commercial models to learn from, the Newton was a flop. Using such failures to argue that handheld computing "doesn't work" clearly leads to a logical contradiction with today's reality!

offer more such germane evidence below.) These data-driven shifts demonstrate how sociologists and social psychologists who hold the stasis view must be incorrect or overly pessimistic. Whether or not they realize it, theorists are haggling over speed, and some nations learn (e.g., to accept evolution or climate change; Ranney, 2012) faster than others. Of course, learning or acting too slowly can exacerbate existing problems.

We partially agree, though, with those who critique a "knowledge deficit" view of public attitudes (cf. Dickson, 2005). Arbitrary or propaganda-like information need not drive one toward a more empirically supported view. Similarly, as argued by both Kahan et al. (2012) and Ranney (2012), individuals' assessment of the risks of GW will be based on an interaction between individuals' knowledge and values. We see the problem as a *wisdom* deficit, for which cognitively sophisticated educators can provide the tools that help the public better evaluate the evidence and make choices that match their values. (See Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012 for a fine discussion of such tools, particularly the correction of misinformation.) We believe that the findings described here will demonstrate that a well considered educational approach can avoid triggering a "conceptual immune response" by reducing the likelihood of a clash with individuals' values—focusing on surprising but non-controversial information. Such interventions may prove critical for public engagement.

Clear successes have been observed with education regarding controversial science. Evolution is one oft-cited example of a polarizing science topic, but Shtulman and Calabi (2008) report that increases in evolution knowledge do indeed boost acceptance. It may be that education must specifically cover critical gaps in students' knowledge. While we should act with caution in comparing GW and evolution education, it is relevant to notice here that both topics require vastly larger views of time, as compared to many other sciences. The longest sample of atmospheric greenhouse gasses and global temperature goes back an impressive 800,000 years. Impressive, that is, until you consider that most major phyla emerged around 530 million years ago. Thus, arguments regarding time scale or system complexity can be readily addressed with successful evolution education in addition to the kinds of experiments we'll see below. Sinatra, Kardash, Taasoobshirazi, and Lombardi (2012) also demonstrate that GW attitudes can be shifted with textual materials, though in this case with a more "persuasive" essay. As we'll see in Chapter 4, however, even individuals with high acceptance levels may be dramatically swayed by only a dash of misleading factual information.

1.4.1 The Numerically Driven Inferencing (NDI) Paradigm

In addition to the arguments offered above, our laboratory has provided arguments and many experimental findings that run counter to either "polarization" or stasis in response to scientific or numerical information. Our group has observed policy shifts regarding many other social issues (e.g., abortion and immigration) with as little as a single number/statistic (Garcia de Osuna, Ranney, & Nelson, 2004; Munnich, Ranney, Nelson, Garcia de Osuna, & Brazil, 2003; Ranney et al., 2008). Below, we offer more experimental results that counter the stasis view, and we explain the different results, in part, by noting that we include a full spectrum of participants, rather than filtering for those who are already relatively extreme (cf. Lord et al., 1979).

NDI procedures (introduced by Ranney, Cheng, Garcia de Osuna, & Nelson, 2001) provide an approach to changing conceptions, attitudes, and even behaviors with quite minimalist interventions (e.g., providing estimators with a single, critical, highly germane, feedback statistic, cf. Rinne, Ranney, & Lurie, 2006). As with the climate change literature reviewed above, the education and social psychology literature provide multiple examples of failures to elicit conceptual change. For example, Chi (2005) describes an intervention in which only 1 in 100 eighth-graders were able to shift to a correct conceptual model of diffusion. Similar examples are available in a variety of literatures (cf. diSessa & Sherin, 1998; Lord et al., 1979). Certainly, there are marked differences between the above mentioned approaches to conceptual change. For the purposes of the current effort, we will focus our attention on those approaches that have been successful (namely, NDI and targeted science education approaches).

One of the elements of the NDI program, the EPIC procedure, represents an intervention that is relatively compact and well specified. More importantly, EPIC has been shown to induce long-lasting conceptual change (e.g., Ranney et al., 2008), as evidenced by increased accuracy on estimations up to 12 weeks later (Munnich, Ranney, & Bachman, 2005). In the EPIC procedure, participants engage with real-world numerical facts that bear on a societal issue, such as abortion, criminal justice, the environment, etc. (e.g., Garcia de Osuna et al., 2004; Munnich et al., 2003). People often poorly estimate these quantities, such that the true values are surprising (even shocking) to many individuals, and experimental research on NDI has provided the basis for successful classroom curricula for both high school students and graduate students in journalism (Munnich, Ranney, & Appel, 2004; Ranney et al., 2008). During the EPIC procedure, participants:

1. Provide an **Estimate** for each policy-relevant quantity,
2. State what they would **Prefer** each quantity to be,
3. Receive actual quantities as feedback to **Incorporate** (as new "Information"), and
4. Indicate whether their preferences have **Changed** upon receiving feedback.

Work that we'll see in Chapter 2 sheds light on the cognitive components of a simpler "Estimate-Inform" procedure. Moving forwards, expanding into an exploration of Preference (in the form of climate-relevant attitudes) allows for a more complete exploration of the effects of such interventions.

1.5 Conceptual and "Less Conceptual" Cognition

Above, we have seen that an unfortunate status quo seems to be arising in the climate change communication literature. Specifically, it is claimed that one should focus on "less conceptual" processing in devising interventions for the public. While there is little data available on the cognition of these scientists, I would hazard to guess that part of the appeal of "less conceptual" cognition (as with the presentation of GW relevant risks) is that individuals can interface with such materials much more readily than with more complex materials such as with science education. However, while less complex or conceptual materials may be faster to comprehend than science-related materials, the total amount of *learning* or *conceptual change* can be far greater when that individual engages with concepts with rich connections to their understandings.

In its limit, the conceptual domain is the space of cognitive processes in which everything is connected to everything. Strong examples would include Whorfian (or neo-Whorfian) theories in which language constrains visual perception (Boroditsky, 2001), or the notion of embodied cognition claims in which our emotional preferences for spatially arranged items may be guided by our fluency with our own right or left sides (Casasanto, 2009). A more prosaic example illustrating the difference between more and less conceptual processing is provided in Clark and Wagner (2003), in which learning with pre-existing knowledge (specifically, encoding known words versus plausible pseudo-words) lowered demands on prefrontal and parietal working memory structures.

Our mind is also endowed with a number of special-purpose, relatively stable, fast, local ("encapsulated") or "hard-wired" capacities. The "motor system" is an excellent example of this.[3] Conceptually, our motor experience is simple—we desire an object and simply reach for it. Under the hood, an enormous number of degrees of freedom are resolved, satisfying multiple complex constraints all without our awareness. Clark and Ivry (2010) construct a set of features that roughly describe the nature of cognitive processing in more or less conceptual modes. I adapt the table given there for Table 1.1. Depending on the needs of a given behavior, learning (or performance) might be better handled by cognition of one sort or the other. These criteria echo what is discussed in the decision making literature (Kahneman, 2003) and cognitive development literature (Sloman, 1996; Carey, 2011).

1.5.1 NDI: A Successful Model for Conceptual Change

A fundamental question in cognition concerns the nature of what is learned. Some well-established psychological learning and memory models (e.g., Nadel & Moscovitch, 1997) might predict that changes in estimation accuracy must ultimately be mediated by the consolidation of episodic memory. In this case, we would expect participants' reports of explicit memory for feedback (the "I" in EPIC) to correlate well with improvements in estimation accuracy at subsequent testing. This would clearly be learning of a conceptual type.

Recent evidence suggests, however, that pre-existing conceptual structures can be re-modeled in a highly efficient manner that may not rely as heavily on the brain structures implicated in episodic memory formation (Tse et al., 2007; Clark & Wagner, 2003). In this

[3]There may be more than one motor system, but at least one of them should serve to illustrate this point.

Table 1.1: Features of more or less conceptual processing. Adapted (liberally) from Clark and Ivry (2010)

More conceptual	**Less conceptual**
Large amount of learning per trial that saturates quickly (high gain)	Small, incremental amount of learning per trial (low gain)
Requires extra time, cognitive resources for processing	Learns automatically without effort
Required for contextual learning	Unimodal or modular learning
Accessible to awareness and conscious intention	Impenetrable to awareness, operates independent of conscious strategies
Consolidation processes are enhanced during sleep	Consolidates off-line with the simple passage of time
Ready transfer to related tasks	Task-specific and inflexible
Rational and recollective	Emotional and intuitive

case, we might expect increases in estimation accuracy even when participants report no (episodic) memory whatsoever for the quantity provided as feedback—particularly if participants had pre-existing knowledge to support such learning. There may be multiple routes to re-modeling our conceptual stores, even when our learning experience is of a somewhat less conceptual flavor.

Evidence of pre-existing knowledge is indicated by surprise upon receiving feedback, which implies an incorrect prior expectation regarding the true value. However, subsequent learning that correlates with surprise might also be explained by an account involving the emotional impact of the information (Munnich, Ranney, & Song, 2007; Thagard, 2006). Therefore, it is important to assess not only surprise, but also whether the surprise had an emotional (i.e., less conceptual) character. It may be the case that surprise mediates improved episodic memory. Alternatively, surprise and the existence of prior knowledge may operate partly or wholly in parallel—mediating direct changes in semantic memory.

Most generally, learning may be driven by the actual experience of surprise (e.g., Munnich et al., 2005; Kang et al., 2009). In addition, improvements in estimation could be driven by a direct (potentially approximate) episodic memory of feedback. Thus, it seems useful to query participants about their surprise, and whether it is of a more emotional or conceptual sort (and we'll see examples of this below). In addition, we can probe participants' memories in an attempt to assess conscious recollective ability. In the end, these processes are likely overlapping, but it may be possible to differentially drive some aspects of learning and not others, etc.

1.6 Summary and Document Plan

There is a clear need for the development of educational interventions targeting climate change acceptance and attitudes. Above, we have seen that there is some indication that compact, evidence-based interventions may provide notable, durable shifts in policy-relevant attitudes. In the chapters that follow, we will more closely examine a set of experiments regarding these sorts of approaches to climate change cognition. Chapter 2 reports on an experiment that illuminates some aspects of the psychological processing of such information, particularly with respect to the role of surprise and conscious subsequent memory. Such notions may be central to understanding how a successful intervention would work. Subsequent chapters will discuss a variety of interventions specifically applied to GW understanding, and assess impacts on beliefs and attitudes.

Chapter 2

Two Routes to Improved Numerical Estimates

As described in Section 1.5, cognition can occur in a relatively local (or special-purpose manner), or alternatively in an integrated fashion that (among other things) affords volitional or conscious access. In Clark and Ranney (2010), we provide some evidence and argumentation for separable learning processes that might be differentially involved when learning numerical information. Below, I present an overview of two experiments in abbreviated form, noting those details that are relevant to the issues posed above. Specifically, these results demonstrate our ability to observe, and potentially guide, cognition between heavily conceptual *episodic* or *semantic* modes of operation on one hand, and less conceptual *emotional* processing on the other.

2.1 Overview

In short, participants saw textual descriptions of numeric items and provided their best estimates. After this, they received the true value, and indicated the degree to which they found the true value *surprising*. After a period including at least one night's sleep, participants were presented with the previously shown textual descriptions of the quantities. Here, participants indicated their *metacognitive assessment of their memory* for the item (including the correct feedback they received from the day before), in addition to re-estimating (or potentially recalling) the value.

2.2 Experimental Methods

The following experiment was designed to assess whether estimative improvement occurs even with respect to items for which no feedback was received—as was found in curricular NDI studies (e.g., Munnich et al., 2004; Ranney et al., 2008). The experiment (1) addresses the correlates of surprise and the timing of feedback on subsequent improvements in numerical estimation—as well as (2) probes whether these improvements are necessarily mediated by explicit recollection. A subset of the EPIC procedure (introduced in Section 1.4.1) was used to explore these issues; participants engaged only in estimation ("E") and feedback ("I"), leaving aside personal preference ("P" and "C").

2.2.1 Participants

Twelve people (seven female) participated, including UC Berkeley undergraduates and members of the general public recruited via the UC Berkeley Psychology department's online recruitment systems (RPP and RSVP). They received either course credit or $20 for their participation in two one-hour sessions over two consecutive days. Ages ranged from 18-56 years.

2.2.2 Materials

Numerical facts (106 of them) were selected from Ranney et al. (2008). An example quantity is "The current percentage of deaths in the U.S. that are caused by lung cancer." Three statistical facts were set aside for the basis of example items (namely US population, world population, and US Gross National Income). Items ranged over a number of topics, and included politics, population dynamics, economics, the environment, education, crime etc. Most items were expressed in percentage form, with the rest being counts of dollars, people, events, or things. For numbers above 999, a comma was used, as in "13,600." For numbers in the millions, billions, or trillions, the appropriate word was used to indicate the order of magnitude (e.g., "300 million"). This was intended to minimize possible confusions about the exact value of the number.

2.2.3 Procedure

Custom software utilizing Vision Egg (Straw, 2008) presented all materials and collected responses (source code available upon request). Descriptions of numerical facts were presented in 1-4 lines of text (with less than 55 characters per line). A prompt for numeric entry was located below the description. Feedback concerning the veridical value was provided in a third location, between the description and the text-entry area.

Blocks of items

Items were randomly distributed into the following four kinds of blocks. Each of these blocks was involved in two or more runs over the course of the experiment. E: Participants only provided Estimates in a single run. EI: Participants provided Estimates followed immediately by correct numerical Information as feedback (i.e., feedback was provided in the same run as the initial estimation). E_I: Participants provided Estimates, then received correct numerical Information in a run that was well-separated from the run in which they provided their Estimate (i.e., "_" signifies a temporal delay). New: A block of items was reserved in both experiments to provide a gauge of false recognition or false recollection.

Experimental runs

Participants engaged in a number of self-paced runs on each of the two consecutive days, as Figure 2.1 depicts. The presentation of stimuli and responses made were uniform across a given run. During the first day, analogous to a "study" phase, participants completed three partially similar runs of numerical estimation and/or informative feedback. The second day was analogous to a "test" phase, in which participants' learning was assessed.

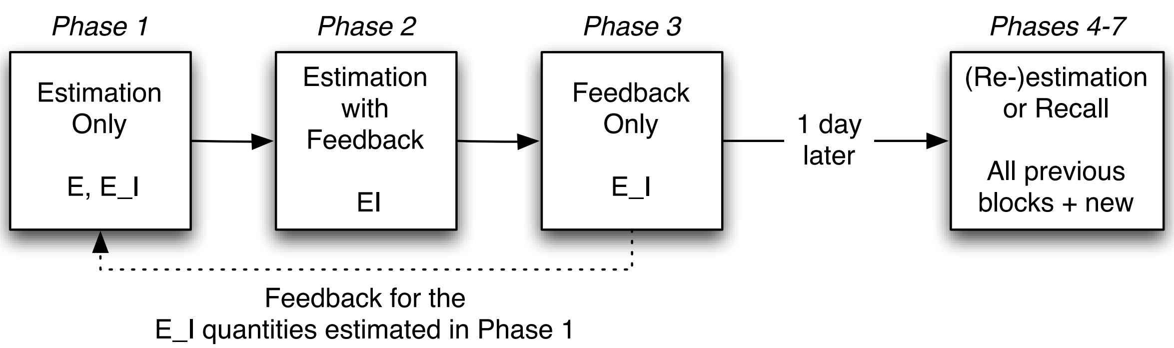


Figure 2.1: A schematic of the experiment's seven runs. Run 1: Estimates were obtained for the E and E_I blocks of items (23 each), randomly intermixed in one run. Run 2: Participants provided 23 Estimates that were immediately followed by Informing the participant of the correct value. Run 3: Feedback (I) was provided for the 23 items from the E_I block that had been Estimated in Run 1. Runs 4-7: Subjects estimated (or recalled) quantities and provided explicit memory ratings for all previous items as well as 34 new items.

During estimation (Runs 1 and 2, with 23 items each), subjects were given a textual description of an item's quantity, followed by a prompt to provide an estimate. For Run 2, feedback was provided 500 milliseconds after each estimate was entered. For Run 3 (with 23 items), the correct numerical value was provided prior to the textual description in order to minimize covert estimation.

In Runs 2 and 3 (thus, for blocks including "I"), surprise ratings were elicited regarding the values given as feedback. Three possible levels of surprise were collected:

1. Little or no surprise
2. Genuine surprise
3. "Visceral" or intense surprise

On day 2, trials were similar to the estimation-only trials in Run 1 described above—no additional feedback was provided. An additional 34 items from the "new" block were randomly intermixed with the items presented during study. Additionally, participants rated their memory for the item according to the following four levels:

1. "The item is new to me"
2. "The item was presented yesterday, but I have no sense of the value provided as feedback"
3. "The item was presented yesterday, and I have some sense of the correct value"
4. "The item was presented yesterday, and I have a fairly accurate recollection of the value."

Choice 1 indicates no recognition or recollection. This is equivalent to labeling the item as "new," and it is the correct response for items from the new block. Choices 2-4 as a group indicate that the item is "old," but with varying levels of familiarity and/or recall. These are correct responses for the E, EI and E_I blocks (although choices 3 and 4 entail a belief that the participant actually received feedback at study, and so might also be considered incorrect for the E block). Choices 2 and 3 indicate perceived recognition, but at least a partial failure in recall. Choice 4 indicates a subjective sense of fairly complete recall.

Note that the estimation task used here is somewhat different than item recognition or cued recall tasks used in many learning and memory studies. The closest point of comparison is likely the notion of *source* memory, in which details surrounding the initial experience of the experimental item are well correlated with hippocampal activity at encoding (Davachi, Mitchell, & Wagner, 2003). In particular, we are *not* asking participants to attempt to recall a particular item from memory. Indeed, these memory ratings can be viewed as a form of metacognition regarding the estimation process and its relationship to the participants' previous experience with the item (i.e., source memory for the item).

2.2.4 Analysis

We modeled improvement as a binomial outcome (as did Munnich et al., 2005). This allows for the treatment of items that have differing distributions within a unified framework (e.g., a linear model would have difficulty modeling both percentages and values in the billions, particularly given our sample size). Items were labeled as to whether estimates improved or not. These labels were fit with a binomial generalized linear model, using the lme4 package in the R statistical environment (R Development Core Team, 2009). This treatment allows for a full multi-factorial mixed-effects analysis. Below, participants are always included as a random effect, and other factors are treated as fixed effects. Linear contrasts were evaluated using the multcomp package, which controls for family-wise error rate (Hothorn, Bretz, & Westfall, 2008).

Unless otherwise noted, data were pre-processed to remove ties. This was done to allow for a null hypothesis that 50% of the remaining items randomly improved and 50%

randomly worsened. If we counted ties as failures to improve, then random drift would end up spuriously suggesting the lack of an effect. Removing ties allowed for tests of whether estimates improved, on average, more than they worsened—both formally and when examining graphs. Otherwise, the removal had little effect on the results, except where explicitly noted below.

2.3 Results

2.3.1 Improvements in Accuracy of Estimation

We can easily reject a null model (in which the presence and timing of feedback had no effect) in favor of a model predicting different improvements across "E," "EI," and "E_I" feedback conditions ($\chi^2(2) = 25.9, p < 10^{-6}$). Post-hoc comparisons between each condition and chance levels, as well as between condition comparisons (as in a Tukey HSD test) were performed simultaneously. In the no-feedback case (E), estimation improvement did not differ significantly from chance ($p = 0.39$), although improvement with Immediate (EI) and Delayed (E_I) feedback were clearly above chance ($p < 10^{-4}$). This may seem unsurprising, but it might have been the case that improvements were at least partially driven by general improvements in estimation skill, and this would have led to at least some modest improvements even without feedback on test items. Indeed, this kind of skill development was the successfully accomplished goal of various EPIC-based curricula (e.g., Munnich et al., 2004; Ranney et al., 2008). In this less extensive experimental manipulation, though, we understandably elicit no such skill improvements. Thus, we assume that these improvements are driven almost entirely by item-specific learning.

2.3.2 Predicting Learning from Surprise and Meta-Cognitive Memory Assessment

As is often the case, the participants' forced familiarity judgments appeared to be superior to their own assessment of their memory (see Aggleton & Brown, 2006 for an overview of relevant theories and experimental paradigms). In participant debriefings, several individuals claimed to be uncertain whether items were old even from Run 1 to Run 3 for items in the E_I block—that is, over an interval of less than 30 minutes! However, participants were excellent at discriminating between old and new items a day later when given a forced choice; 76% of new items were correctly identified as new on Day 2, while erroneous "new" responses to previously seen items had a mean prevalence of less than 9%. This level of recognition accuracy is not surprising given the considerable depth of processing involved, and the rich pre-existing memory structures available for scaffolding these episodes.

As the lack of feedback yielded non-significant changes in estimation accuracy, here we consider only items from conditions including feedback ("I"). Many of these effects are

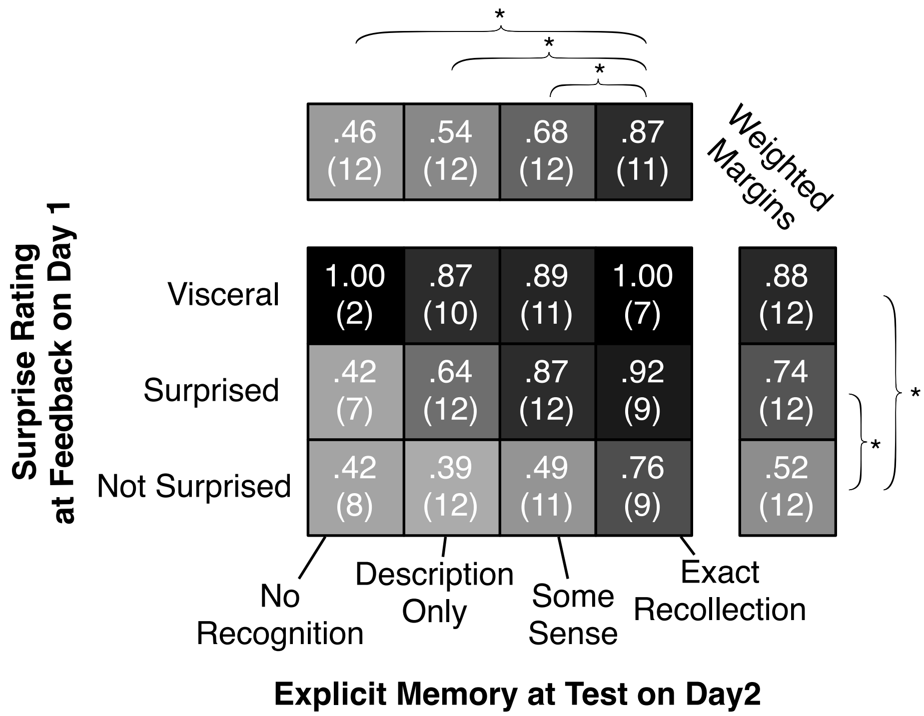


Figure 2.2: Fraction of items improving across different levels of surprise rating and metacognitive memory assessment. The number is parentheses represents the number of subjects (out of 12) contributing to that cell. Margins are appropriately weighted according to the number of items in each bin, and as such are not the simple mean of the row or column in the central table. Significant differences between levels of individual factors are marked with an asterisk.

depicted in Figure 2.2. A model that predicts estimation improvements on the basis of both surprise and declarative memory responses is well-supported by the data. We readily reject a reduced model excluding memory ($\chi^2(3) = 34.8, p < 10^{-7}$), as well as one excluding surprise ($\chi^2(2) = 295.22, p < 10^{-16}$). An inclusion of an interaction term does not yield a significantly better model ($\chi^2(6) = 2.85, p = 0.8$).

It should be noted, that there is a small (but non-significant) difference in surprise ratings between EI and E_I blocks: subjects rated 64% of EI block items as surprising ("2" or "3") versus 59% for E_I (although no straightforward effect was observed with metacognition on memory). This result mirrors the results obtained in, e.g., the study in Section 6.1 regarding climate change cognition, in which prior estimation increased participant reports of surprise (cf. Rinne et al., 2006). Thus, the timing of feedback may have an effect on estimation improvement that is mediated by surprise; these issues seem best addressed in a subsequent study, though—likely one that modulates surprise to a greater degree.

Recall of the exact value (memory response "4"), as compared to other memory, was a highly significant predictor of improved estimation (all p's < 0.001 for the lower two ratings, $p = 0.01$ when compared with response "3"). No other comparisons between memory levels are significant. For surprise, both moderate and visceral ratings yielded significantly greater improvement than for not-surprised rated items ($p < 0.002$ in both cases), but did not differ significantly from one another. Note that participants provided the exact numerical figure given as feedback only 35% of the time when selecting choice 4. Even if we broaden this liberally to items where participants are within 15% of the true value, they were only correct about 74% of the time.

Finally, if we consider the relation between surprise and metacognition on memory, there appears to be almost no correlation. The correlation of fixed effects between memory and surprise terms in our model was consistently smaller in magnitude than 0.1. This, combined with the lack of significance of an interaction term, provides some evidence for independent learning processes.

While both *surprise* and *metacognitive memory assessment* were predictive of improved estimation from the first to the second session, these measures did not interact significantly, and moreover were uncorrelated with one another (i.e., progressive darkening from the lower-left to upper-right corner in Fig. 2.2).

On its own, this result would be insufficient to make strong claims about multiple cognitive routes for learning. But, this result fits well with an ever increasing literature (an overview of which was provided in section 1.5).

2.3.3 Exclusions

As many as three items lacked estimates from some subjects or exhibited a clear lack of understanding (e.g., a number such as 10 million for a question asking for percentage) and these items were excluded from the analyses above. Due to a technical issue, participant 01 was not run on the standard E manipulation, but was included in memory and surprise-related analyses, as these analyses did not include E trials.

2.4 Discussion

Given the overall improvements in estimation ability evidenced in curricular studies by Munnich et al. (2004) and Ranney et al. (2008), it is of interest that we see no statistically significant improvement in items that didn't receive feedback (the "E" block). In other words, it appears that participants did not improve their estimation skills in the absence of feedback particular to a given item. Nonetheless, it seems that learning in this considerably shorter experiment was largely item-specific and related to the integration of feedback. This lack of improvement in the present experiment may be due to a lack of time for reflection or development of strategies—which were highlighted, taught, and fostered in the curricular studies. (Munnich et al., 2004 and Ranney et al., 2008, also focused, to a fair degree, on preferences and personalized policies which may engage a web of related concepts.)

2.4.1 Learning Without Metacognitive Report of Recall

From the point of view of a memory theory, the most interesting result is perhaps the existence of learning even when participants claimed “no sense” of the numerical value provided at feedback—rather like a memorial analog to blindsight. This argues against the notion that improvements in estimation are simply the result of explicit episodic memory. The result is reminiscent of extant dual-process memory models. For example, Davachi et al. (2003) suggest that successful recognition could occur through a process of recollection and/or a sense of familiarity. These processes moreover appear to be subserved by distinct sub-regions in the medial temporal lobe. In the present study, though, we see improvement in numerical estimation—which is perhaps most akin to a cued recall task for EI and E_I items—without full recall of the number presented on the previous day. Thus, the task here is perhaps more naturally expressed in the language of the remember/know distinction (Knowlton, 1998). That is, while participants appear not to remember a specific (usually multi-digit) number from the previous day, there is still a sense in which they know the number better than they knew it the day before.

Based on the significance of the existing results, however, it seems reasonable to posit that a non-episodic form of learning undergirds some of the improvement in participants’ abilities to estimate accurately. Further, the learning for improved estimation (or memory) seems to occur often without an explicit, precise recollection of the feedback from the prior day. This argues for some implicit and/or rapidly semanticized learning in support of these improvements. In particular, this appears to have something of a “less conceptual” flavor. This line of reasoning is reminiscent of studies of children applying abstract mathematical rules before they are aware of doing so (Siegler, 2000).

Of course there is also a clear role for explicit episodic recall in learning numerical information. In particular, how well participants believed they could recall the number was indeed predictive of improved estimation. But instructional materials that elicit surprise in students may allow students to learn without conscious awareness that they have learned anything—at least in domains that are scaffolded by nontrivial preexisting knowledge. If the material is unsurprising, it appears that episodic encoding may be a critical step in successful improvement. It should be noted that surprise might be too specific a notion. It may be that the relevant feature has more to do with general emotional salience, or how interesting the material is to students. Certainly, however, it seems that there are multiple routes to learning even relatively concise facts. Thus, our development of climate change interventions might usefully engage factors such as surprise and engagement with pre-existing knowledge to bolster more rote forms of learning.

Acknowledgements

The work reported in this chapter has been previously published, in part, in Clark and Ranney (2010). All such material is re-used here with the permission of my co-authors, the publishers, and the Graduate Division at the University of California, Berkeley. Thanks to Luke Miratrix for his input during the development of Figure 2.2

Chapter 3

Introducing our Survey Methods

Chapters 4, 5, and 6 all utilize similar survey methods to assess climate-related beliefs and attitudes, in addition to a number of related constructs relevant to Ranney's (2012) RTMD theory. For reference, the full list of survey items used in this body of research is included in Appendix A. Note that codes for each item are given in sans serif typeface, and that convention will indicate these codes in what follows. For example, one question has the code knwgbl. Codes which include numbers, such as gw1_2, indicate that they are one of several questions probing a particular construct (in this case, global warming, or "GW" beliefs and attitudes). In our experimental chapters, we'll examine the way our interventions are able to shift these beliefs and attitudes (primarily those related to climate change), as well as noting how these beliefs and attitudes relate to one another. I hasten to note that the number of potential relationships between the many variables we have measured would require an enormous amount of data to test fully. As such, we will restrict ourselves primarily to the exploration of *a priori* relationships of interest.

3.1 Clarifying "Beliefs" and "Attitudes"

Survey methods in the social sciences may use the terms "belief" and "attitude" in a quite technical fashion. For example, in Fishbein's Theory of Reasoned Action, an "attitude" is essentially the weighted sum of a set of beliefs and norms (see Montaño & Kasprzyk, 2008, for an overview of such theories). While we will not contradict this formulation, below I take a more common-language approach. Specifically, in the text that follows, a "belief" should be taken as a measure of agreement with an objectively verifiable fact about the world. For example, the reality of anthropogenic climate change (assessed primarily by item gw1_2) may be difficult to ascertain, but in the end, it is something that could be settled by observation. An "attitude," on the other hand, indicates a measure of agreement with an emotional or evaluative stance towards some aspect of the world. For example,

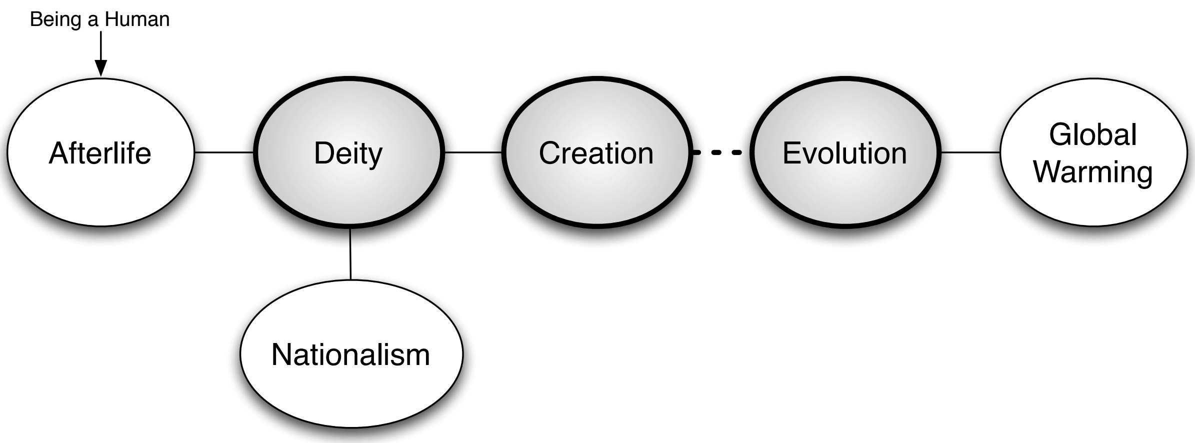


Figure 3.1: Core RTMD relationships (adapted from Ranney, 2012). Conceptual coherence or association (operationalized as positive correlation) is represented by a solid line, and conceptual incompatability by a dashed line. A coherent cluster representing devotion to "god and country" is on the left, and a science-accepting cluster is on the right. The most direct conflict is captured by the explanatory competition between creation and evolution. Note that while all constructs were surveyed, GW items are the ones primarily reported on in this document.

worry about global warming (gw2_3).[1]

3.2 An Overview of Survey Items

Survey items were primarily sourced from Martinez (2009), and were chosen on the basis of both observed quality (from the results of Martinez, 2009) and conceptual fit for our interventions. The first page in particular (consisting of the six items with a "1" prior to the underscore) was selected as the most ideal set of six questions for targeting the six RTMD constructs, depicted in Figure 3.1 (Ranney, 2012). Notably, gw1_2 was selected for its focus on acceptance of *anthropogenic* climate change. Note that while we included a variety of measures in our surveys, in what follows, we will focus solely on those items assessing GW attitudes and beliefs.

[1]Certainly, one can objectively and even reliably *observe* signs of worry in the behavior or physiology of an individual. But here we are talking about a spectrum of stances an individual could take ranging from the assignment of truth or falsity on one end, to the assignment of emotional value on the other. In other words, the notions of "belief" and "attitude," as I use them, have nothing to do with how we are observing the individual.

3.2.1 Naïve Survey Results

Most of the climate-related interventions that follow include some measure of participant attitudes and beliefs prior to the intervention. In this dissertation, we are primarily seeking insight into different forms of conceptual *change* (and thus, one hopes, behavioral change). Given this, a detailed consideration of these naïve results is beyond our scope. However, some of the relationships obtained seem relevant to understanding the mind of our potential students. I therefore note such results below. For a fuller treatment of survey material, please consult the relevant publications of the Reasoning group (notably, S. Cohen, 2012, Ranney, Clark, Reinholz, & Cohen, 2012a, Study 1).

S. Cohen (2012) reports a variety of results from a sample of 270 San Diego residents (201 park visitors and 69 community college students). Notably, some important relationships were obtained using Cohen's data contrary to the "knowledge deficit" and polarization arguments referenced in Section 1.3. First, we observe a robust correlation between mechanistic climate change knowledge and attitude toward climate change. This result was maintained even when taking political party into account. Specifically, mechanistic knowledge correlates with acceptance that global warming is occurring ($r = 0.22$, $p = 0.0002$) and is anthropogenic ($r = 0.17$, $p = 0.005$). Anthropogenic climate change acceptance also predicted financial "willingness to sacrifice" ($\chi^2(4) > 32$, $p < 0.001$ for each of four items), and one's knowledge score predicted two of these items ($\chi^2(1) > 3.8$, $p < 0.05$ for both). Further, acceptance of biological evolution was found to predict beliefs and attitudes toward climate change (as RTMD hypothesizes, and, e.g., Ranney, 2012 found).

We might infer, then, that "acceptance of controversial science" is a problem above and beyond political ideology. These findings suggest that the effects of well-chosen aspects of education are both significant and somewhat independent of political affiliation. Indeed, evolution acceptance was a significant predictor of climate change acceptance even in a model including the two major political parties ($\chi^2(4) = 12.3$, $p < 0.02$; N.B., including other parties dramatically reduces quality of fit for any model, likely due to small bin sizes). Given these results with a representative sample of the american public, we considered ourselves justified in focusing primarily on attitude and belief questions in the interventions that follow. Below, as appropriate, we will see how such relationships maintain in our various populations in the context of a number of interventions.

3.3 Ensuring Quality of Survey Responses

A central concern in large survey studies is determining if the surveys were completed in good faith. We want to avoid including surveys that were filled out in a random or incoherent way. Below, I explain a graphical method for identifying individuals who fall outside of the normal variation in responses across survey items.

3.3.1 A Graphical Method for Checking Survey Quality

We can represent participant responses in a raster plot, where each response from an individual is represeted by a particular shade or color in a grid. Each column represents a specific question and each row represents an individual. We can then easily sort rows by the mean response of that participant, and columns by mean response for that question. Plots are created using only items that should be relatively coherent (e.g., all items dealing with religion, or climate change). An example of this is given in Figure 3.2.

Participants who respond markedly differently from those with similar mean scores pop out visually, and can be inspected manually. In many cases, after inspection, such "abnormal" participants were retained—largely on the basis of inspecting their written answers. Only if they appeared to be truly negligent or outright non-cooperative in responding were they excluded. This method of graphical inspection was applied for all samples in the following empirical chapters, but may not be mentioned in cases where all subjects were considered acceptable, as was normally the case.

Acknowledgements

Particular thanks to Luke Miratrix for his input in developing the method described in Section 3.3.

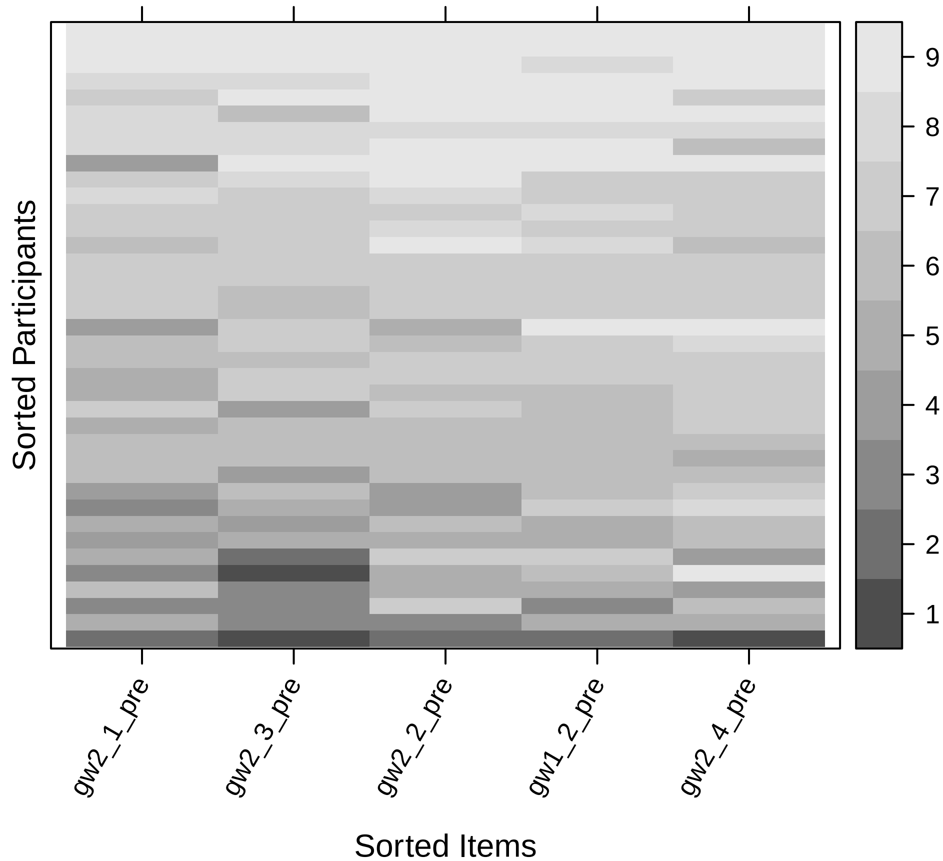


Figure 3.2: An example of a plot allowing quick identification of individuals who answer markedly differently than those with similar mean attitudes. In this case, the data are from the study detailed in Section 6.3. Here, `gw` items from an example pretest are plotted (and reverse coded items are flipped prior to plotting). The item with the lowest mean rating is leftmost, and the item with the highest mean rating is rightmost. Likewise, the 38 participants are sorted in increasing mean order from bottom to top. The fifth individual from the bottom is a bit suspicious, with a "1" for `gw2_3` and a "9" for `gw2_4`, and thus was carefully inspected to ensure honest engagement with the survey. There are also two participants with straight "9"s at the top of the graph. Upon inspection, these participants simply provided extreme answers to all questions, but otherwise appear to be taking the survey seriously.

Chapter 4

Learning "Evil" Non-Representative Climate-Related Numbers

The Numerically Driven Inferencing (NDI) paradigm is introduced in Section 1.4.1. In Chapter 2, we see that surprising numerical information can have a lasting impact on an individual's conception of a politically relevant number—sometimes even without conscious recollection of the experience that catalyzed this change. Below, we will see that misleading, cherry-picked numerical facts can have a marked shift on individuals' beliefs and policy preferences. This result serves as a call-to-arms for climate educators, as even relatively well-educated, liberal, Global Warming (GW) accepting students are highly susceptible to the kinds of information currently being used to undermine acceptance of climate change.

4.1 Overview

As described in Section 1.3, there is some debate surrounding the value of scientific or numeric information regarding climate change. Indeed, some have claimed that such interventions will only serve to polarize individuals, thus making the situation worse than it already is. However, some organizations publish out-of-context facts to try to undercut the reality or gravity of human-caused climate change. Such numbers are often blatantly cherry picked. For example, one can locate GW deniers who note that the Earth's temperature decreased (by 0.2°F) from 1940 to 1975 (Jastrow, Nierenberg, & Seitz, 1991). This surprising fact, though, hardly contradicts the ever more obvious warming trend over the last 125+ years—one can pluck many "trends" in noisy time series by picking endpoints that are oddly high or low. Given this rather clear intent to mislead (corroborated by Oreskes & Conway, 2010), we (partly tongue-in-cheek) label these numbers "evil."

Our three hypotheses were that misleading facts would reduce:

1. participants' climate change acceptance,
2. ratings of their knowledge of the issue, and

3. their climate-change-related funding preferences.

Of course, lest we would have eroded participants' acceptance of anthropogenic climate change more than fleetingly, we debriefed them right afterward with more complete information—including a large dose of representative numerical facts as well as the basic physical mechanism of GW (the mechanism is detailed in full in Chapter 6). It should be clear that we're not interested in perfecting an approach to eroding acceptance of the overwhelming scientific consensus on climate change!

In this Chapter, we'll see easily obtained and dramatic effects with an NDI intervention using cherry-picked numbers that appear to contradict the veracity of anthropogenic climate change (when considered in isolation). There is only one experiment reported, with the intent of determining whether these materials would drive meaningful changes. In particular, we administered this experiment only to undergraduate seminar courses in which we could be sure to administer a proper debriefing.

4.2 Study: UCB Lecture Intervention with "Evil" Numbers

4.2.1 Methods

Participants

Two lecture courses of UC Berkeley undergraduates (spanning cognitive science and "Behavioral Change") were engaged in this intervention (N = 104). 59 students completed the 8-item "no pretest" version of the experiment and 45 completed the 2-item full EPIC intervention. All participants were retained after examining the coherence of survey responses.

In the 8-item intervention, 34 participants were female and mean conservativism of 3.64 (on a 1–9 scale; $sd = 1.59$) was observed. In the 2-item intervention, 31 participants were female and mean conservativism was 3.68 ($sd = 1.43$). Breakdown of political party is given in Table 4.1.

Materials and procedure

Participants were engaged in one of two misleading numeracy interventions (depicted in Figure 4.1). In both versions, survey methods were as described in Chapter 3. This study utilized a somewhat compact version of a pre and post-intervention test using only the 14 items in Table A.2 in the Appendix (up through `engage`, and in the order given in that table), plus a self-rating of climate-change knowledge (for a total of 15). After the intervention, all participants also completed a brief demographic survey. A reproduction of the actual format of the survey is provided in Appendix D. Notable demographic results from this survey are reported directly above, under "Participants."

In the "no pretest blast" version of the intervention, participants estimated each of eight items prior to receiving the feedback values, with an emphasis on maximizing the

Table 4.1: Stated party affiliations for participants in the "evil" NDI study (UC Berkeley undergraduates).

	8-item	2-item
democrat	20	21
republican	4	2
green	1	0
libertarian	1	3
independent	6	2
none	21	13
other	1	1
decline to state	5	3

quantity of feedback numbers presented to the participant. To this end, this eight-item survey included only a posttest (i.e., no pretest), and lacked a policy component (thus, it was an EI intervention, lacking "P" or "C," similar to the approach used in Chapter 2).

A more comprehensive engagement containing only two items was administered to the rest of the class. This version included a pretest and additional questions about each item. In addition, we asked students about their surprise level after each feedback value and requested both their climate-change funding Policies and post-feedback policy Changes versus various United Nations Development Project (UNDP) millennium goals. Thus, this latter variant was a full EPIC intervention. The same set of alternatives was used across the four variants of the 2-item intervention, and these are listed along with policy-relevant instructions in Appendix B. (The format of this part of the intervention is given in Appendix D, but the specific pages with the "evil" numbers have been redacted, as we wish to withhold the text of the numerical materials to avoid spreading misleading information.)

Note that while this experiment is presented first as a motivation for the following chapters, it was actually carried out *after* a number of experiments in Chapters 5 and 6. Thus, a number of experimental design choices made here are motivated by findings from experiments in those chapters.

4.2.2 Results

Overall, these numbers had a profound impact, the details of which are described below. As with other NDI interventions (e.g., Ranney et al., 2008), and consistent with our pilot testing, individuals generally found each of these items surprising, ranging from surprise ratings of 5.83 to 8.53 across both interventions. Mean surprise ratings were 6.03 for the 2-item intervention and 6.62 for the 8-item intervention. Ratings were on a 1–9 scale, with all ratings above "1" indicating some level of surprise.

Shifts away from GW policy preferences

As hypothesized, policy preferences for funding goals related to climate change dropped ($\chi^2(1) = 22$, $p < 0.01$) for both funding questions across all eight numbers (the funding policies are described in Appendix B). Unfortunately for global warming as a social priority, the highest mean pretest preference for funding climate change initiatives reached only a 50-50 split of available funds. These results are depicted in Figure 4.2. While, due to time constraints, we did not include funding qeustions to check for a similar result in our 8-item intervention, it seems likely that similar (or greater) shifts would occur along with the much more drastic GW attitude shifts we’ll see below.

GW acceptance eroded by misleading numbers

Also, as hypothesized, mean climate change acceptance dropped significantly, from 6.5 on the pretest to 6.2 on the posttest for the two-item condition (6% of available room, for a 9-point scale, $t(42) = -4.3$, $p < 0.001$), and significantly to 5.9 for the eight-item condition (12% of available room, $t(88.6) = 2.61$, $p < 0.005$). Note that these shifts were also in the direction of ambivalence (a “5” rating), and may reflect confusion rather than disagreement. Mean ratings are depicted in Figure 4.3.

Self-confidence in GW knowledge eroded by misleading numbers

Our third hypothesis was also supported, as self-rated knowledge dropped from a mean of 5.0 on the pretest to 4.5 for the two-item condition (12% of available room, $t(44) = -2.5$, $p < 0.01$), and plummeted to 2.9 on eight-item survey ($t(87.2) = -5.3$, $p < 0.001$). This latter decrease, 2.1, represents 53% of the available room to drop on a 9-point scale, which is exceptionally large. These ratings are depicted in Figure 4.4.

4.2.3 Discussion

In stark contrast to arguments that numeracy is polarizing (e.g., Kahan et al., 2012), we have provided an existence proof that appropriately selected scientific facts can have a profound effect in eroding the existing beliefs of a population (i.e., “liberals” can be pushed in a more “conservative” direction). In particular, we have demonstrated marked erosion of self-confidence in one’s own knowledge, as well as acceptance and concern regarding anthropogenic climate change—even in our relatively liberal and anthropogenic-climate-change-accepting sample of UC Berkeley undergraduates. Such results were observed with as little as *two* numbers. Consider the effect of Richard Muller’s writings prior to Rohde et al. (2013). A prominent professor at our ostensibly liberal institution wrote extensively on why we should doubt the veracity of climate change—including in the mainstream media. We must assume that educated, liberal, GW-accepting individuals may be easily swayed by a small dose of factual (but non-representative) numerical or scientific information from such a source.

The primary point illustrated by this study is that individuals' understandings are demonstrably fragile. Even an intervention of a few minutes can massively undercut individuals' confidence in their own knowledge, along with overall belief and concern about global climate change. An additional point is that—as noted by Kahan et al. (2012) and McCright and Dunlap (2011)—if we survey individuals with high scientific literacy, who are likely to have been self-guided in their climate-relevant education, such education appears to be polarizing. Thus, one might conclude that climate change accepters are unlikely to come into contact with numbers like the ones in this study on their own. However, there are concerted efforts to distribute such numbers on the internet and elsewhere. A final point is that, as shown above and as noted by McCright and Dunlap (2011), scientific information might push individuals both towards scientific consensus, as well as *away* from it. Thus, overall, it seems wise to build a solid foundation of climate-change-relevant knowledge in the American populace.

It is clear that even relatively educated members of the public (e.g., undergraduates at a top-tier university) are highly susceptible to misleading, cherry picked facts. Such facts are clearly known to organizations attempting to undermine the overwhelming scientific consensus about climate change. Thus, climate educators and communicators must counter the increasing sophistication with which such organizations distribute misleading information.

Acknowledgements

The work reported in this chapter has been previously published, in part, in Clark, Ranney, and Felipe (2013). All such material is re-used here with the permission of my co-authors, the publishers, and the Graduate Division at the University of California, Berkeley.

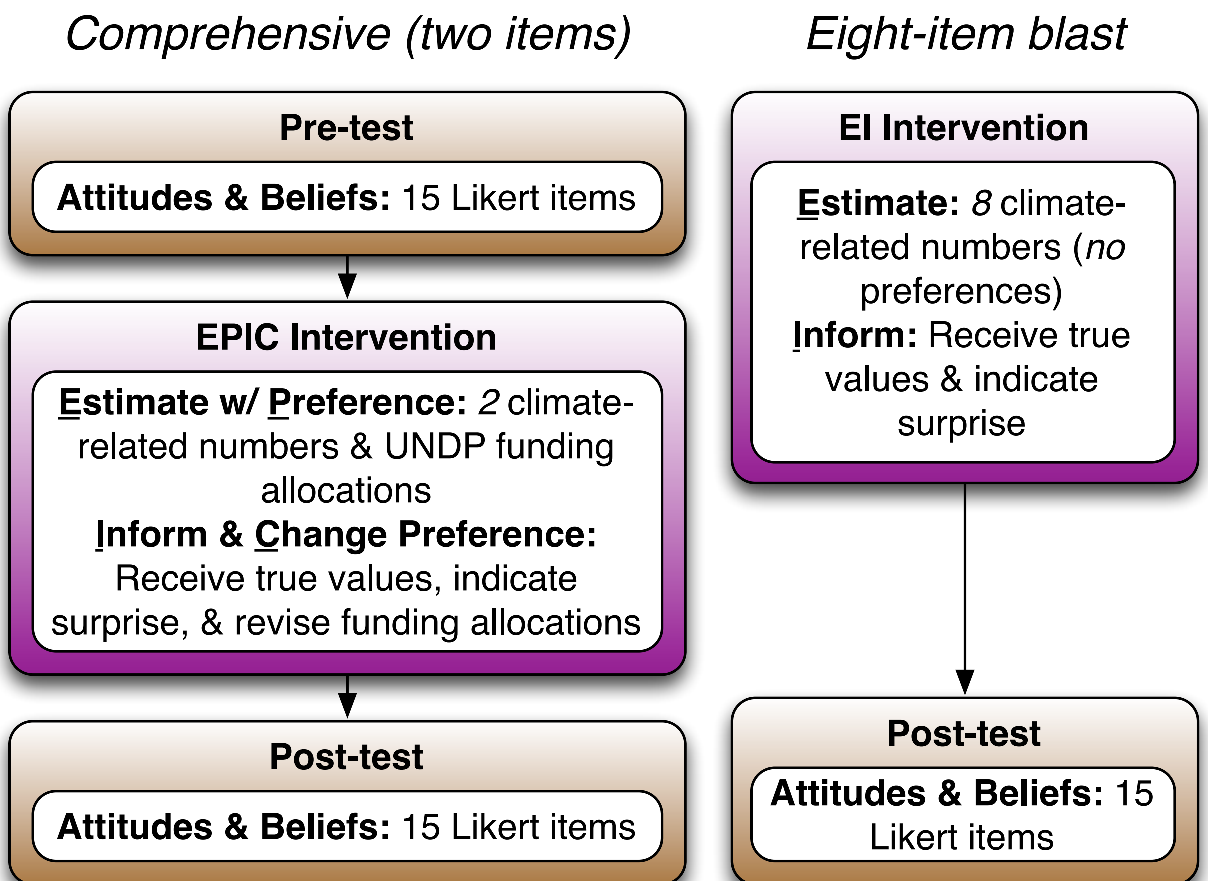


Figure 4.1: An overview of experimental procedures for our "evil" NDI experiment. If we consider our EPIC or EI interventions to be "jam," and our surveys to be slices of bread, we see that the participants in the 2-item condition received a full "sandwich." (Participants in the 8-item condition received no pretest). Flow for NDI experiments in Chapter 5 had a full "sandwich" structure, but used the condensed EI intervention style from the 8-item intervention (based on the results we'll see below, in which the 8-item intervention elicited much larger changes in our measures).

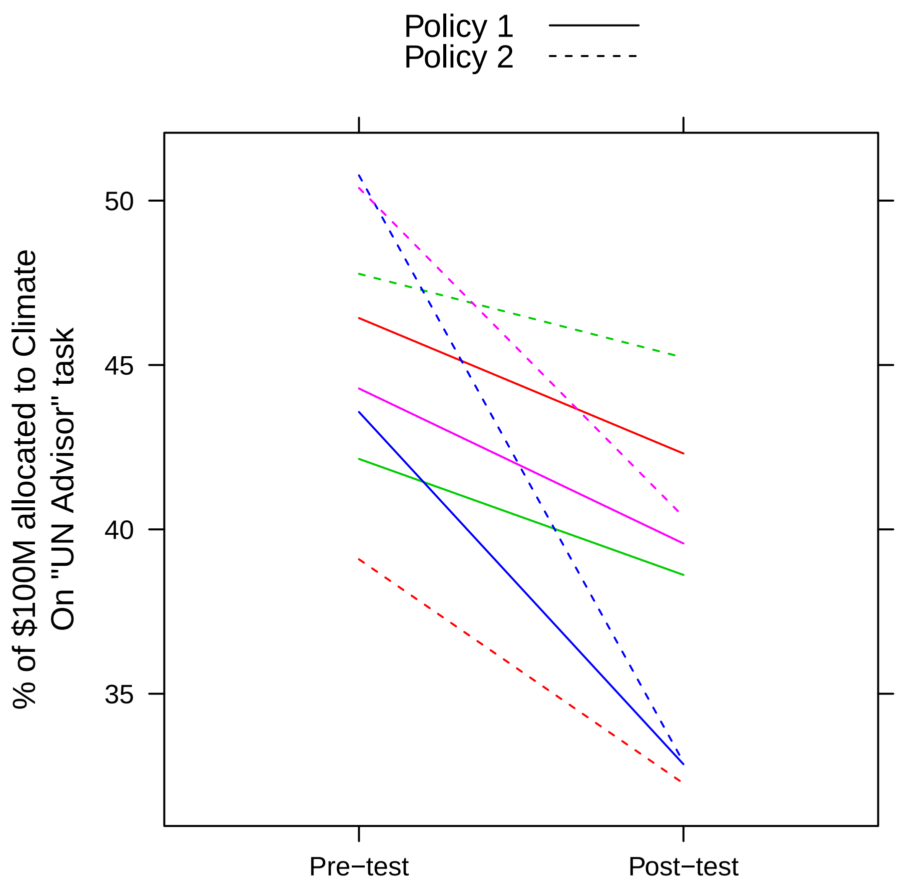


Figure 4.2: Significant drops in preference for allocation of $100M towards climate-related projects on "UN Advisor" task ($p < 5 \times 10^{-6}$). The two policy decisions are detailed in Appendix B. Each of the four survey variants is represented by a different color (each variant utilized two of the eight numerical estimation items used in this study). The two policy choices remained the same in each variant (and are indicated by solid versus dashed lines). Values are expressed as a percentage of funds that were allocated to climate-relevant projects versus projects supporting an alternative UNDP millennium goal.

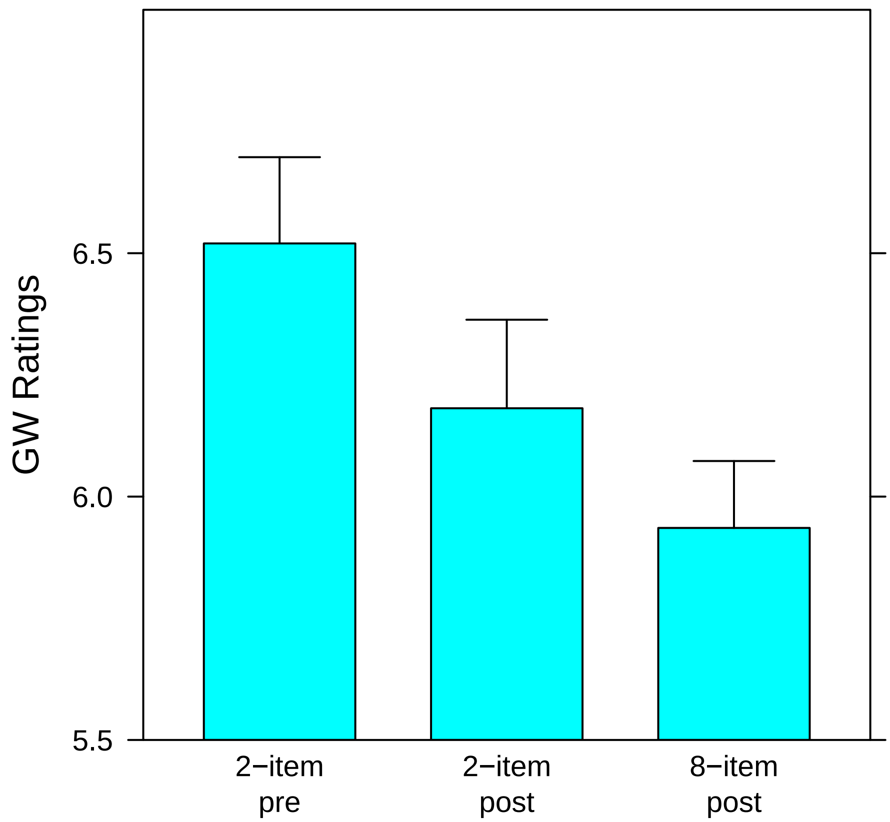


Figure 4.3: Mean ratings for GW survey items on pre and posttest. Pretest surveys were only administered for the 2-item condition, but should be indicative of population responses. Error bars represent standard errors for each measure.

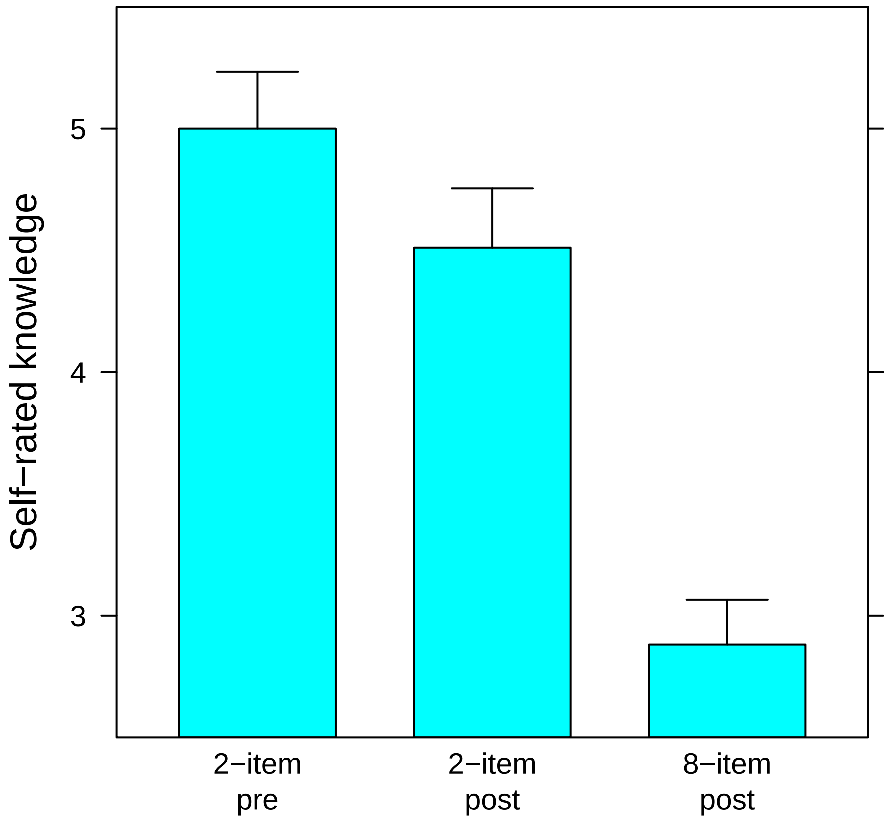


Figure 4.4: Self-rated knowledge for individuals on pre and posttests. Again, pretests were only administered to individuals in the 2-item experimental variant.

Chapter 5

Learning Representative Climate-Relevant Numbers

The Numerically-Driven Inferencing (NDI) paradigm is introduced in Section 1.4.1. In Chapters 2 and 4, we have seen further demonstrations of the kinds of marked attitudinal and conceptual shifts one can obtain with quite minimalist interventions. In particular, we've seen the striking effects of the EPIC procedure (both introduced by Ranney et al., 2001), as well as more minimal interventions involving only the "E" (Estimate) and "I" (Inform) portions of the EPIC intervention. (Rinne et al., 2006, also explore the primary importance of committing to an estimation prior to receiving the correct information.)

As with Chapter 4, we again present a collection of NDI interventions in a somewhat different order than they were actually performed in. This way, we are able illustrate a potential problem—specifically, even these more representative numbers yield erosion of confidence in one's own knowledge, just as with the "evil" numbers in Chapter 4. We'll proceed to demonstrate that this problem is solved in Chapter 6. In the studies below, we presented participants in the different conditions with numerical information that is relevant to global climate change acceptance. Unlike in Chapter 4, we used numbers that were likely to *boost* acceptance. As before, survey methods employed in the following studies are described in detail in Chapter 3.

5.1 Study 1: Online Experiment with UC Undergrads

Given the efficacy of "evil" numbers and previous successes of the NDI paradigm, this study assessed the efficacy of numbers that support the claim(s) of global climate change. Again partly tongue-in-cheek, we call these "saintly" numbers. Given prior NDI studies of similarly "shocking" magnitudes (e.g., Garcia de Osuna et al., 2004), our hypothesis was that the accurate feedback would increase participants' climate change acceptance, but diminish self-confidence in their knowledge of the issue.

5.1.1 Methods

Participants

UC Berkeley undergrads (N = 60) were recruited via the Research Participation Pool (RPP). Prior to engaging in experiments in the RPP program, many participants complete a pretest screening survey containing demographics and other items contributed by numerous experimenters. This RPP pretest was completed by 30 of our participants.

Materials and procedure

This study used an on-line version of materials otherwise similar to those used in our "evil" NDI study (reported in Section 4.2.1), and used a pretest survey included in the RPP screening survey. In particular, the core intervention in which all participants were engaged was analagous to the eight-item blast shown in Figure 4.1. Some participants, however, also completed a pretest (at an earlier time during RPP prescreening). Thus, 30 participants completed a full "sandwich" while the other 30 were in a no-pretest condition.

For those sandwich participants that completed the pretest, it was completed, on average, 18 days prior to the intervention. In the main intervention, we queried individuals about eight quantities (listed in Table E.1 in the Appendix). The eight items were accompanied by questions directed at participants' surprise and their reactions to each number. Fictitious monetary policies were left out of this version, as simple attitude shifts were readily observed in the simplified 8-item "evil" intervention, and these shifts are more directly comparable across experiments. An added feature of the online intervention is that we could remind individuals of the estimates they gave on the same page on which they incorporated numerical feedback, ensuring that they contrasted the two. As with online surveys in Chapter 6, a posttest regarding participant attitudes and beliefs was administered both immediately after our intervention and after a retention interval.

5.1.2 Results

Self-rated knowledge and surprise

These items were, as anticipated and as with the "evil" items, able to significantly erode self-rated knowledge (5.3 to 4.0 on a 1–9 scale for our full sandwich condition, $t(29) = -3.6$, $p < 0.01$). This erosion was comparable to that found with the "evil" numbers. The numerical items ranked relatively high on participant surprise compared to the (significantly effective) 400-word mechanism intervention from Chapter 6. The mean surprise rating across items was 4.8 (on a 1–9 scale; all ratings above "1" indicate some level of surprise). While this is a bit lower than surprise ratings for "evil" numbers in Chapter 4, it markedly exceeds the mean 2.9 surprise rating for the 400 words obtained using an analogous online intervention with UC Berkeley undergrads (reported in Section 6.2.2). It is also very much in line with the 4.7 mean surprise rating noted in Study 2, which *did* obtain attitude shifts using very similar estimation items (reported under "Participants" in Section 5.2.1). Thus, the immediate affective impact of these numbers was reported as *higher* than an intervention that (as we'll see in Chapter 6) effectively supported significant shifts in attitudes for both students and the general public.

One of the most surprising numbers (with a mean surprise rating of 5.2) was the percentage of active researchers who support the tenets of anthropogenic climate change, reflective of the strong relationship between perceived scientific consensus and the acceptance of climate change reported in Lewandowsky et al. (2013). The two numbers most comparable to the statistics in the 400 words (and in the video adaptation available at http://HowGlobalWarmingWorks.org) were similarly surprising, with the rises in atmospheric methane and atmospheric CO2 ranking at 5.9 and 5.1, respectively—both higher than the mean surprise ratings from the 400 words in Chapter 6.

Global warming (GW) attitutudes

In spite of the powerful impacts described above, attitudes, acceptance, and beliefs regarding climate change remained stable after this intervention with "saintly" numbers (6.71 pre versus 6.67 posttest). This lack of effect is counter to prior NDI studies (as well as the results reported in Chapter 6), in which individuals' preferences and beliefs were often markedly shifted by even a single number.

5.1.3 Discussion

An experimental silver lining here is the demonstration that participants will not report greater climate change acceptance merely by dint of experimenter demand. In both this *and* previous NDI and RTMD studies, participants were explicitly told that all feedback statistics and other information were fully accurate, that the study involved *no* deceptions. One possible explanation is due to a methodological change: prior studies also provided the particular scientific/literature source both for each statistic that was sought and each provided as feedback. Sources were not uniformly provided in this study. This is one difference that may partially account for our lack-of-effect.

So, it is possible that participants were less compelled by the authority of this study's statistics, compared to those in Chapter 4. Another possibility is that, as in Chapter 4, participants were left feeling less knowledgeable—perhaps weakening any boost these surprising numbers could have on climate change acceptance.

A final possiblity is that the effect of this numerical intervention would be strengthened by an appropriate context for integrating this information. That is, perhaps we could not simply present our numerical information as it was in isolation with the expectation of an effect. Indeed, as we report in (Clark et al., 2013), similar numbers had little immediate effect on high-school students as a part of a global warming mechanism curriculum. However, students exposed to numbers like those in this study retained the effects of the curriculum to a greater extent than students in a control condition.

A possibility that will disconfirmed in the introduction to the following section is that this lack of result was due to the delay between attitudes assessed during during RPP prescreening and our core intervention. As we'll shortly see, this lack-of-result was replicated within a single session intervention.

5.2 Study 2: Online intervention with Amazon Mechanical Turk

After the difficulty obtaining shifts in GW attitudes and beliefs above, I was able to replicate this difficulty by presenting the same materials to a more general population on Amazon Mechanical Turk. I won't go through the exercise of reporting another null result here, apart from mentioning that unlike the above, this intervention contained a full "sandwich" in a single survey/session. Thus, it seems unlikely that our difficulty in observing a result with these particular materials depends on the timing of the pretest.

I then engaged in a thorough examination of the wording of the items (also discovering that the informational feedback for one item was off by an order of magnitude due to an error converting from the metric system). This process was relatively informal, and consisted of showing the items to naïve individuals and asking them if they had any difficulty understanding them. Descriptions were iterated until they appeared to make sense to non-experts.

5.2.1 Methods

Participants

"Workers" on the Amazon Mechanical Turk platform (N = 40) were recruited to complete a two-part survey. Language used in recruitment made no mention of climate change, and was titled "Politics, Numbers, and Your Attitudes." After removal of problematic participants above, we were left with 38 participants in the intervention. Eighteen of the total retained participants were female. One participant reported being born outside the United States, but residing here for 20 years. Stated party affiliations are listed in Table 6.13. Mean conservativism was 4.0 on a 1–9 scale ($sd = 2.1$), which is comparable with our college students in other studies in this dissertation. However, participants in this study were more likely to declare being a Democrat or Republican. While the sample is still clearly biased towards the Democratic/liberal end of the political spectrum, the ratio between Republicans and Democrats is less extreme than our undergraduate samples. This is in line with the results reported by Richey and Taylor (2012), in which approximately 73% of polled workers on Mechanical Turk reported voting for Obama (versus approximately 51% in the election results). The complete distribution of declared political party affiliation is reported in Table 5.1.

Table 5.1: Number (and percentage) of individuals declaring a given party affiliation.

	party (percentage)
democrat	17 (44.7)
independent	12 (31.5)
republican	4 (10.5)
libertarian	3 (7.8)
other	1 (2.6)
decline to state	1 (2.6)
total	38 (100.0)

Materials and procedure

Materials are largely identical with those in Study 1, apart from the improvements in comprehensibility described above. An additional improvement in terms of making the survey more clear was replacing a fill-in-the-blank question for specifying units and "increase" or "decrease" for their estimate. In this version of the experiment, one or two choices including the unit and direction were provided as appropriate to each estimate. For example, "% of researchers" or "feet increase." The item reproduced in Appendix D is an exact copy of the format used for this study. Note that participants were further asked to indicate their experience of an item on three different scales—one asking about "surprise," one about "embarrassment at lack of knowledge," and one about the familiarity of the item.

On inspecting the materials used in Study 1, numerous individuals remarked that the instructions were too long, and the materials difficult to understand. Thus, for this study, I endeavored to eliminate unnecessary instructions and generally simplify language. This was a somewhat subjective process, but feedback from non-experts indicated that the materials used in this study were indeed easier to understand. Note that in the process of streamlining instructions, I removed instructions to the effect that *no* deceptions were used. This is in contrast to the previous (ineffective) study that did include such an instruction. In addition, in this study even more than in the above study, the inclusion of information about authority was scant (e.g., the more accessible phrase "journal article" replaced phrases such as "article published in PNAS"). The correct trade-off between simplicity and completeness (e.g., inclusion of proper references) is a worthy topic of further consideration. The exact wording used in this study is shown in Table E.2 in the Appendix.

A final change was the removal of an item on sea-level rise, the feedback for which had previously been incorrectly reported as 10 times higher than the true amount. It is a small possibility that this item undermined individual's belief in our other numbers in Study 1 above, but based on comments, only a few individuals appeared to doubt the number.

Data quality on Amazon Mechanical Turk

The use of an anonymous, on-line labor pool raises concerns about data quality. For example, people may try to take the survey again or they may lie about their demographics (i.e., claiming they are U.S. residents so that they may gain the credit). Re-taking is one of the most easily guarded against concerns on Amazon Mechanical Turk, as Amazon will attempt to enforce this if requested, as was done in this study. However, in addition, no IP addresses were repeated in the participant pool for this study.

Amazon will attempt to restrict individuals to the U.S. if requested, and this was done for this experiment. An additional layer of verification of location is straightforward using "geo IP" databases. In this case, geographical locations were retrieved using the GeoLite data created by MaxMind (available from http://www.maxmind.com). On our survey, participants indicated the state they reside in. Participant IP addresses were subsequently checked against this reported location. Here, two individuals IP's appeared to be located in Germany and Guatemala, and so these participants were excluded. Most other participants had IP addresses that resolved to the state they claimed to be from. One participant's IP address was not listed in the MaxMind database, and was traced to either Hughes Net or Bright Home, both U.S.-only satellite internet providers.

5.2.2 Results

Surprise and related measures

The numerical feedback was again ranked as surprising, and in a similar range to that observed in the above Study (which failed to shift attitudes). Here, we have a total of 21 measures of something like surprise, three for each of our seven estimation items. The number of potential relationships is large here, so we must be cautious in over-interpreting *post hoc* observation. However, some clear structure appears to exist amongst these correlations, as can be seen in Figure 5.1. Specifically, we can observe a clear block structure near the diagonal for the "embarrassment" and familiarity ratings.[1] It seems that individuals had a tendency to rate these items relatively low or high across items (reflected by larger off-diagonal correlations), while such correlation was not present with "surprise" (i.e., ratings of surprise tended not to predict one another within a given participant). This can be summarized statistically with the averages of the off-diagonal correlations in these blocks. The surprise items had a mean correlation with each other of 0.14, while the embarrassment and familiarity items had means of 0.41 and 0.40, respectively.

Another clear structure can be seen in the smaller diagonals visible off the main diagonal. Where these diagonals appear, this indicates that for a given item, the relevant measures tended to covary. The mean of the correlations along these minor diagonals were

[1] "Block structure" is a common notion used in mathematics when describing a square region of a grid, where the elements of that region are more similar to each other than to some other portion of the grid. Here, the grid is displayed as a bitmap in Figure 5.1, but it could just as easily be represented as a matrix of numbers. In either case, the term "block structure" would be appropriate

as follows: surprised with embarrassed was 0.57, surprised with familiar was -0.31, and embarrassed with familiar was -0.09. Thus, embarrassment and familiarity do a poor job of (anti-)predicting one another. Surprise, however, is well correlated with the embarrassment rating, and still reasonably (anti-)correlated with familiarity.

Please note that above, we are not engaged in a process of hypothesis testing, as these correlations are of little import for our hypothesis-driven conclusions. Rather, these correlations should be interpreted as descriptive measures to assist with the comprehension of the structure in this data (and potentially inspire future hypothesis-driven research).

Across items, surprise ranged from 3.2 to 6.3 ($\mu = 4.7$), embarrassment from 2.6 to 4.3 ($\mu = 3.4$), and familiarity from 2.9 to 4.1 ($\mu = 3.6$). (All items were rated on a 1–9 scale.) Some floor effects may have occurred for embarrassment and familiarity (the opposite of what we'll see in Chapter 6. If we order our items by one of these ratings, the order remains the same if we use surprise or embarrassment (but the order changes for familiarity).

Overall, it appears that our surprise question is the least likely to be uniformly high or low across items for a given subject, thus the variation in responses to this item is more likely to be reflective of an individual's response to the item itself. For these numbers, surprise also suffers from less of a floor effect. That said, no combination of mean or maximum surprise (and related) scores were significantly predictive of shifts in GW acceptance and concern.

GW acceptance supported by clear numerical information

This version of the intervention did significantly boost mean GW acceptance and concern by about 1/3 of a point from a pretest mean of 6.4 to a posttest mean of 6.8, as depicted in Figure 5.2. The shift was significant ($t(37) = 2.74$, $p < 0.01$). Thus, it appears that while we (and others) experience some failures of numeracy to achieve shifts in the direction of scientific consensus, it appears that a carefully crafted intervention can have useful effects.

Self-confidence in GW knowledge is still eroded

While we see shifts above towards the scientific consensus on global warming, participants still report feeling about a point less knowledgeable (dropping from a pretest mean of 5.2 to a posttest mean of 4.2) as depicted in Figure 5.3. This drop is significant ($t(37) = -3.38$, $p < 0.001$).

5.2.3 Discussion

As compared with Study 1, the primary changes were a different target population and a modification to improve the comprehensibility of our materials. While there were differences in declared party affiliation, conservativism was quite similar between this group and other student populations. It should be noted, however, that our undergraduates' mean GW ratings on the pretest were a bit high compared to our other interventions with

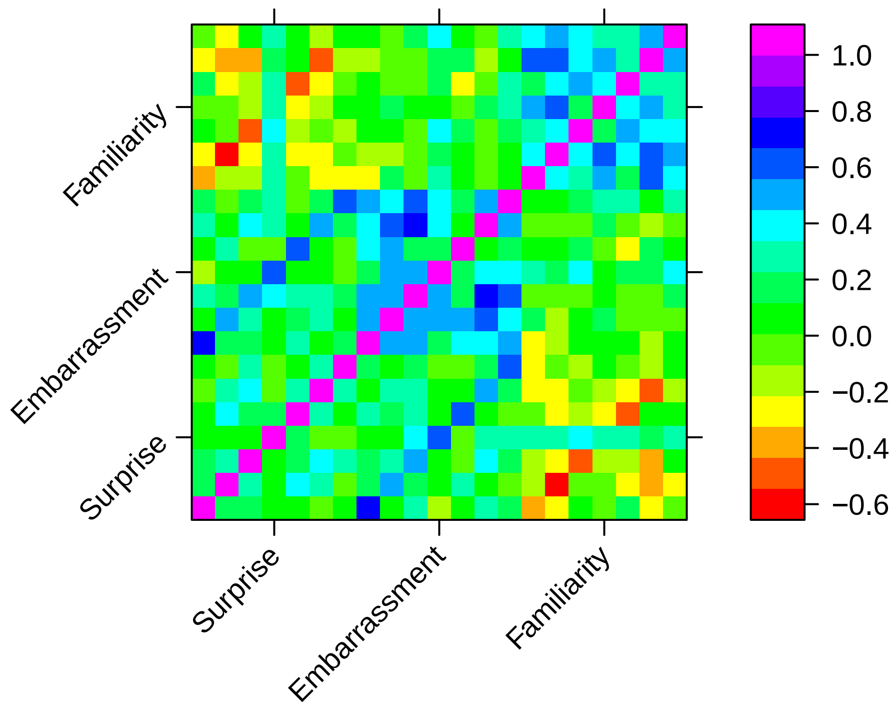


Figure 5.1: An image plot of the correlations between ratings for each of the 3 ratings across the 7 estimation items (note that the plot is symmetric, such that all off-diagonal correlations are included twice). Each rating type is presented in a seven-item block centered vertically and horizontally on one of the three labels (textual labels apply to all items within a seven-by-seven block). Relatively clear block structure is apparent surrounding the main diagonal for embarrassment and familiarity (specifically, there are more blue/high correlation values in these seven-by-seven blocks). Parallel to the main diagonal, a blue/high correlation diagonal of positive correlations can be seen within an item between embarrassment and surprise, and a red/negative correlation diagonal can be seen for surprise and familiarity.

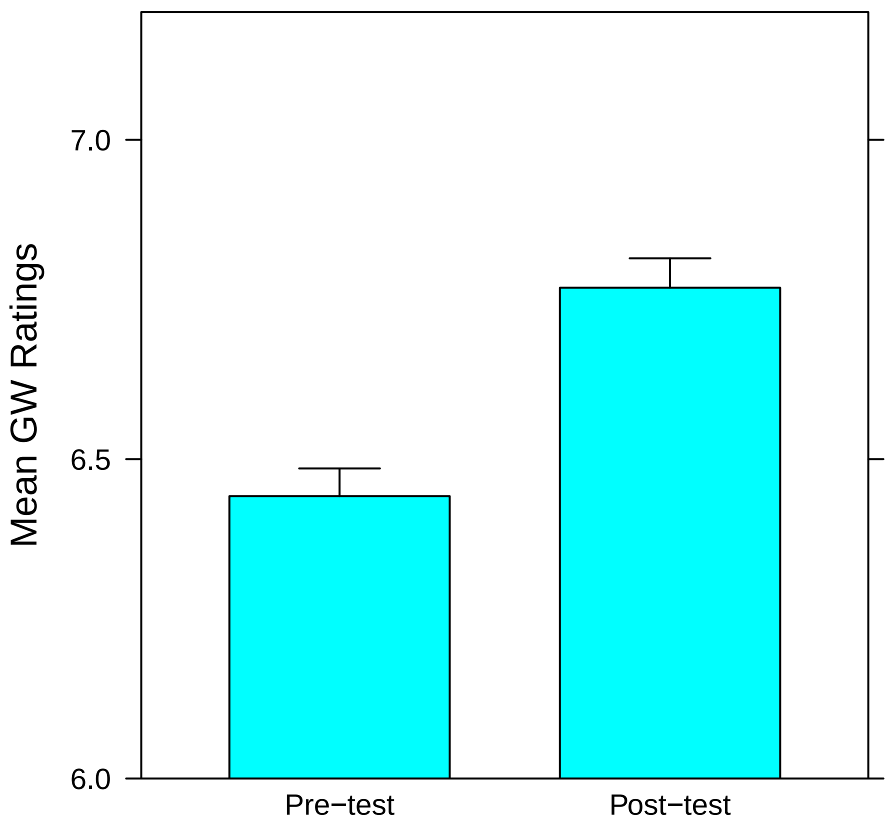


Figure 5.2: Shifts in GW ratings in Mechanical Turk intervention with climate-change-supporting numbers. The pretest mean is 6.4 and the posttest mean is 6.8.

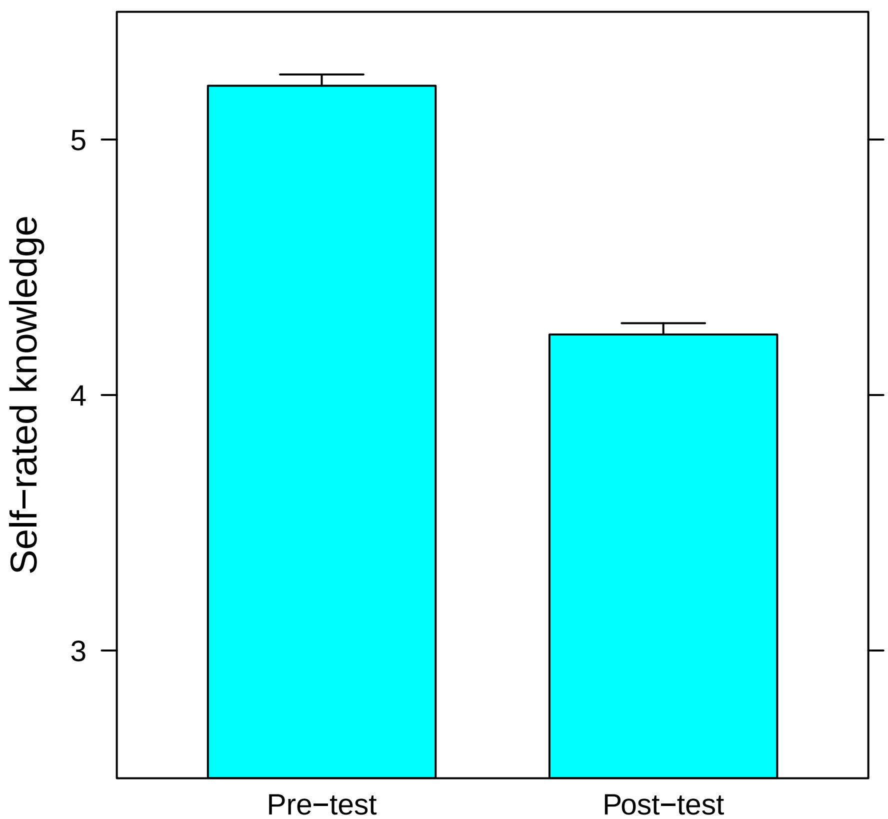


Figure 5.3: Erosion of confidence in self-ratings of GW knowledge in Mechanical Turk intervention with climate-change-supporting numbers. Means are 5.2 on the pretest and 4.2 on the posttest.

undergrads (and comparable to some posttest scores!). Thus, there may have been a ceiling effect. Regarding the materials used in this study, they seemed more comprehensible to non-experts (at least informally). As such, this intervention may have been more persuasive in shifting participants' acceptance of climate change and related attitudes. We should be careful, however, in making comparisons across populations with different interventions. If one were truly interested in the effect of materials, one should provide the materials as used in Studies 1 and 2 to the same target population. If one is uninterested in such questions about materials, I'd recommend the materials from Study 2, perhaps with the addition of a statement about "no deceptions being used". The basic "saintly" information communicated in both sets of materials was quite similar.

The similarity in the effects reported in Chapter 6 provide some evidence that similar interventions perform similarly across UC Berkeley undergrads and workers on Mechanical Turk. Thus, it seems reasonable to recommend that careful attention be given to materials like those used in this chapter. Items should be tested for comprehensibility with naïve individuals prior to attempting to use them in a belief change or behavioral change intervention. Again, the optimal balance of simplicity and completeness is a topic for future study.

Note also that, for reasons of time and simplicity, we did not include policy shifts in this study. However, we can still compare the size of belief and attitude shifts with the 2-item study described in Chapter 4 and infer that we are seeing attitudinal shifts of a similar magnitude.

5.3 Summary and Conclusions

Despite the lack of any observed (even numerical) shift in GW beliefs and attitudes in Study 1 above, it affords us a number of insights. Critically, we cannot simply throw a set of statistics at Americans and expect that to impact their beliefs and attitudes. While we cannot claim to know for sure what "went wrong" with Study 1, there are a few notable differences. In both studies, not all items had sources, but sources were further simplified (and sometimes omitted) in Study 2. Many were likely difficult to understand in Study 1 (wordings in Appendix E are reflective of the final wordings used in Study 2). Unlike Study 2, Study 1 *did* include the assertion that the study involved no deceptions. Thus, an obvious possible explanation is that a certain degree of comprehensibility may be necessary to effect shifts in GW attitudes and beliefs. A final possibility is that this lack-of-effect occurred because of a ceiling effect. But in any case a real silver lining here is support for the conclusion that shifts, when we do observe them, are *not* driven merely by experimenter demand.

Combined with Chapter 4, we have now witnessed numeracy-based interventions that push individuals towards and away from the scientific consensus on anthropogenic climate change. In addition, we have seen that even when students claim surprise regarding a set of numbers, they may not be influenced by these numbers—unless perhaps they are presented with the necessary clarity.

It should be noted also that, as in Chapter 4, participants were left feeling less knowledgeable than they reported prior to the intervention. It remains for future research to determine what impact this might have on behavior, but it seems likely that a lack of confidence would likely inhibit public statements or commitments regarding climate change.

Our research group has also integrated such numbers with more comprehensive interventions. For example, Clark et al. (2013) reports on the utility of such numbers in improving retention of a climate change curriculum described in Felipe (2012). In the following chapter, we'll see another relatively simple, but more comprehensive intervention that includes two of our more surprising numbers. This intervention leaves participants both more informed *and* more confident in their own knowledge.

Acknowledgements

The work reported in this chapter has been previously published, in part, in Clark et al. (2013). All such material is re-used here with the permission of my co-authors, the publishers, and the Graduate Division at the University of California, Berkeley.

Chapter 6

Teaching the Mechanism of the Greenhouse Effect

As described in Chapters 1 and 3, American's lag much of the world regarding acceptance of anthropogenic (i.e., "human caused") climate change. Informally, Michael Ranney, then other members of our group started questioning whether people were able to mechanistically explain how human activities cause an increase in global mean temperature. Almost no one could provide a satisfactory explanation, including most of us! As described in Section 3.2.1, no surveyed members of the general public were able to offer a credible description, but the quality of partially correct responses correlated significantly with global warming (GW) attitudes.

Our investigations revealed that almost no-one—only a few academic experts and *no* members of the general public—knew the basic concepts described in our 400 words. Moreover, variation along measured knowledge was correlated with individuals' attitudes. Thus, we developed and evaluated an science education intervention (arguments for this approach are also provided in Section 1.4). Prof. Ranney, Lloyd Goldwasser, and Daniel Reinholz (along with input from myself and Ronald Cohen) first developed a short 400-word description of the mechanism. This text is reproduced in full in Appendix G (and in a more condensed form in Ranney, Clark, Reinholz, & Cohen, 2012b). Before reading the text yourself, I would encourage you to spend 10 minutes describing *your* understanding either aloud or on paper.[1]

In these experiments, we sought to formally ask:

1. Is this lack of understanding for the mechanism of global climate change as pervasive as it seemed to be?
2. Does instruction regarding the mechanism of global climate change increase individuals' acceptance of the reality of anthropogenic climate change?

[1] If you personally doubt the veracity of anthropogenic climate change, then you may modify the exercise to describing the mechanism of the greenhouse effect, first described by Nobel laureate Svante Arrhenius in 1896.

Along the way, we additionally considered related aspects of learners' cognition (the details of which are described below).

The history of educational research would imply that it's quite difficult to arrive at definitive answers to big policy questions. For example, phonics versus whole-word reading has been debated at least since the dawn of the Common Era, as discussed in Compayré (1889). Below, however, I report on a series of experiments that argue strongly (if not definitively!) that instruction regarding the physical mechanism of the greenhouse effect appears to have some positive effect on public acceptance of anthropogenic climate change. As discussed above, such public acceptance seems central to any truly democratic approach to the problem of climate change.

As discussed in Chapter 1, others' studies detract from the utility of approaching climate change as a science education problem (Lord et al., 1979; Kahan et al., 2012, e.g.,). In that discussion, we noted numerous potential shortcomings of such studies, such as the exclusion of participants with moderate attitudes. In this intervention, we focus on tackling one notable shortcoming. Specifically, the interventions in this chapter focus on a fundamental, well-researched knowledge gap, and our assessment focuses on acceptance/belief. Such contrasts may explain the difference between observing instructional benefits (as we have) or polarization (as others occasionally have; Lundmark, 2007, cf.). We'll see further evidence below, however, that such interventions are applicable across a variety of settings, time-frames, and populations, and that global warming understandings and attitudes are far from static. Most importantly, such understandings seem to affect attitudes and beliefs in a meaningful way.

6.1 Study 1: Lecture Room Interventions at UCB and UTB

In this chapter, we are considering the efficacy of a mechanistic explanation of climate change in addressing the educational challenges laid out in Chapters 4 and 5. It should be noted, however, that this experiment was actually carried out prior to the those experiments. Thus, this experiment was a particularly thick, exploratory observation of individuals' beliefs, attitudes and knowledge. Here, "thick" means that we explored the same phenomenon through multiple routes—for example, doing keyword searches of textual responses, and examining coded responses though a number of metrics.

The methods are largely analogous to those in the previous 2 chapters (see in particular Section 4.2.1). Here, as opposed to a primarily numerical intervention, we sought to understand how a relatively brief 400-word mechanistic explanation, including two of the more surprising numbers from Chapter 5, might affect climate-relevant beliefs and concern, as well as how this might be modulated by prior commitment to one's own explanations and stated attitudes. The general flow of the experiment is given in Figure 6.1.

The primary goal here was a proof of concept. By assessing university students—some of our nation's most highly educated citizens—we provide a strong test of our hypothesis that most Americans are ignorant of the mechanism of global climate change. An additional

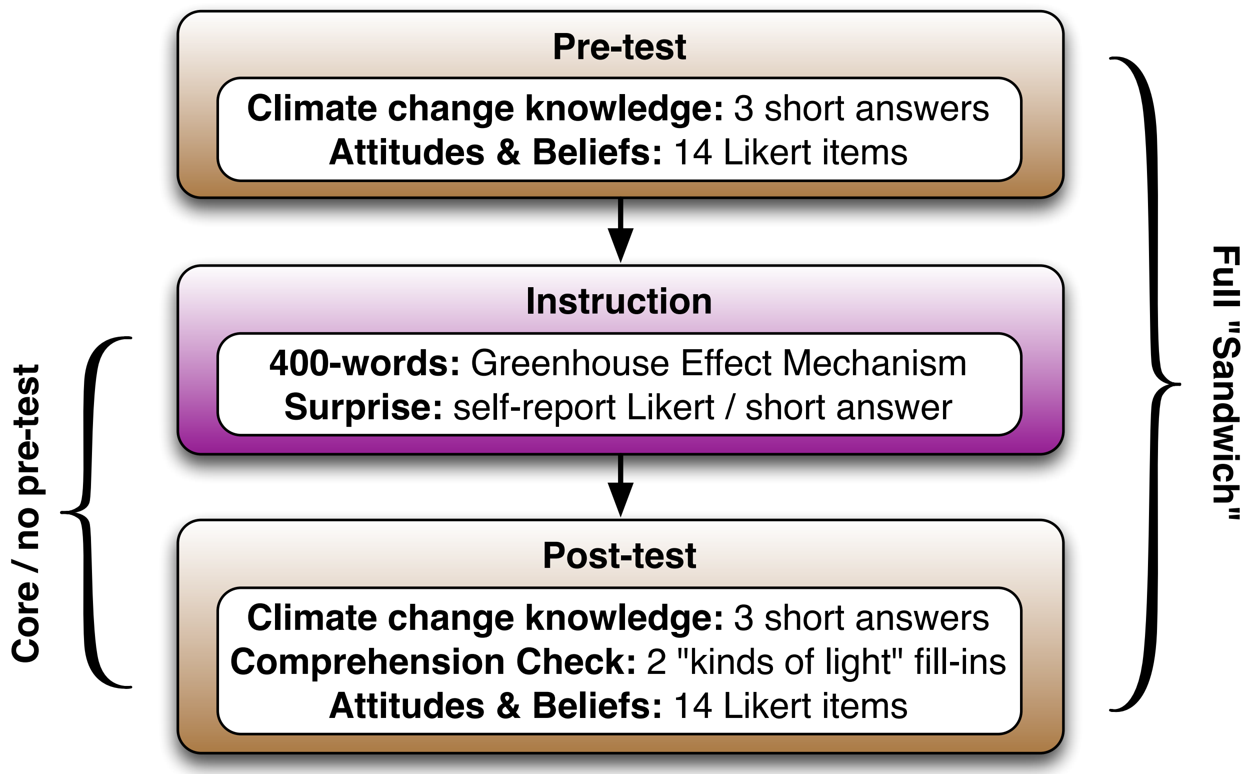


Figure 6.1: An overview of the experimental flow for Section 6.1. The flow for other experiments in Chapter 6 was similar. The analogy to a sandwich takes the knowledge and attitude tests to be slices of bread, and the educational intervention itself is the "jam."

concern was that for maximal power, it is preferable to sample naïve beliefs prior to the intervention. In such a design, we are able to use repeated measures statistics and consequently have much greater power. On the downside, however, problems can arise from an assessment prior to an intervention. For example, we were concerned that individuals might exhibit an increase in their stated belief in anthropogenic climate change merely by dint of experimental demand. This is evaluated by comparing our sandwich and no-pretest groups. As described in Section 3, pretest responses can be used to asses naïve knowledge and attitudes in the general public.

6.1.1 Methods

Participants

One-hundred three University of California, Berkeley, and 46 University of Texas, Brownsville, undergraduates were randomly assigned to one of our two groups: "sandwich" or "no-pretest." UT Brownsville is an "Hispanic-Serving Institution" with over 96%

Table 6.1: Number of students from both sub-populations stating membership in a given political party. Note that the demographic survey was given after the intervention (and thus may have influenced participants' willingness to state their political affiliation).

Party	Berkeley	Brownsville
decline to state	5	7
democrat	40	7
independent	4	2
libertarian	9	1
none	24	16
other	1	1
republican	2	3

of the student body reporting as Hispanic. UC Berkeley is somewhat more racially diverse, with no overarching mission in that respect. Students were recruited from large lecture courses. Those from Brownsville were recruited from courses in the physical sciences, while Berkeley participants were recruited from a course in cognitive science. Below, we report data from the 85 Berkeley and 41 Brownsville students who completed the survey as intended and had been U.S. residents for ten years or more (because we expressly consider U.S. exceptionalism/nationalism). Of the Berkeley data, we analyzed 43 no-pretest surveys and the pretest part of 42 sandwich surveys—but due to anticipated time constraints, only 30 sandwich posttests could be completed/obtained. Of the Brownsville data, we analyzed 22 complete no-pretest and 19 complete sandwich surveys.

Of the 85 Berkeley students analyzed, two did not complete the demographic test. Forty-three were female and 40 were male. Mean conservatism was 3.69 (i.e., slightly liberal on our 9-point scale; 1.65 standard deviation). Of the 41 Brownsville students 21 were female, 20 were male. Here, mean conservativism was 4.95 (i.e., moderate; 1.77 *sd*). Political affiliation is reported in Table 6.1.

Materials and procedure

The general flow of the intervention is given in Figure 6.1, and was collected anonymously. A sample of the format of the core no-pretest intervention is given in Appendix F. Participants were split into two groups, receiving either the "no-pretest" version of the intervention (sometimes called "open-faced"), or the full "sandwich" (filled with nourishing descriptions of climate change!).

The climate change knowledge portion of the pre and posttests consisted of the three questions described in the Appendix, Section I.1. For this experiment, the Likert items (all on a 1-9 scale) consisted only of knwgbl followed by the first 13 items in Table A.2 in the Appendix. Both groups read the educational text regarding the mechanism of greenhouse gases (reproduced in full in Appendix G), and indicated any surprise they may have experienced (again, on a 1-9 scale). The "kinds of light" check consisted of two fill-in-the-blank questions regarding both the kinds of light coming to earth from the sun and radiating away. Here, "sunlight" or "visible light" were considered correct for incoming, and "infrared" was considered correct for outgoing. Some participants wrote "ultraviolet" for incoming light, which one could charitably ascribe to a partially correct understanding.

After completing the intervention described above, participants also completed a demographic survey, detailed in Table A.1. The experiment began with a page of instructions, including the assertion that no tricks or deceptions were involved in this study. Lastly, given that their experimental intervention was shorter, individuals in the no-pretest group were asked to provide some feedback on the intervention, and also on Al Gore's "An Inconvenient Truth" (if they had seen it). These responses were used to refine our methods going forward, and Bem (2011) notwithstanding, should have had no effect on the results.

Participants were run simultaneously for each of the two classes. Instructions were administered by the course instructor, and students received one of two packets—placing them into one of the two groups described above. After completing the consent form on the front of the packet, individuals proceeded to read and answer questions. The entire experiment required approximately 25 minutes to complete.

Analysis

Handwritten responses were coded and placed into a spreadsheet (for details, see Appendix I). Given the rich nature of these data, many analyses were employed. As such, please see the Results section that follows for details of the analysis used for each question.

6.1.2 Results

Please note that all statistical tests are reported in full in the tables associated with this section. In the text, I primarily indicate only a basic value to give a sense for the strength of the result.

Scored knowledge: Learning the global warming mechanism

Even our rather sophisticated samples initially exhibited incorrect or non-normative understandings of the greenhouse effect's mechanism (e.g., on the roles of ultraviolet light, the ozone layer's depletion, non-greenhouse-gas pollution, and the reflection of incoming light). Most notably, not a single pretest explanation mentioned different light/radiation types or atmospheric retention time, despite an explicit prompt to explain any differences between the energy traveling toward and away from Earth. However, after reading the 400-word description, 61% of the Berkeley participants across both groups correctly answered that "infrared" light was emitted from Earth (in its fill-in-the-blank space), as did 55% of the Brownsville students who responded.

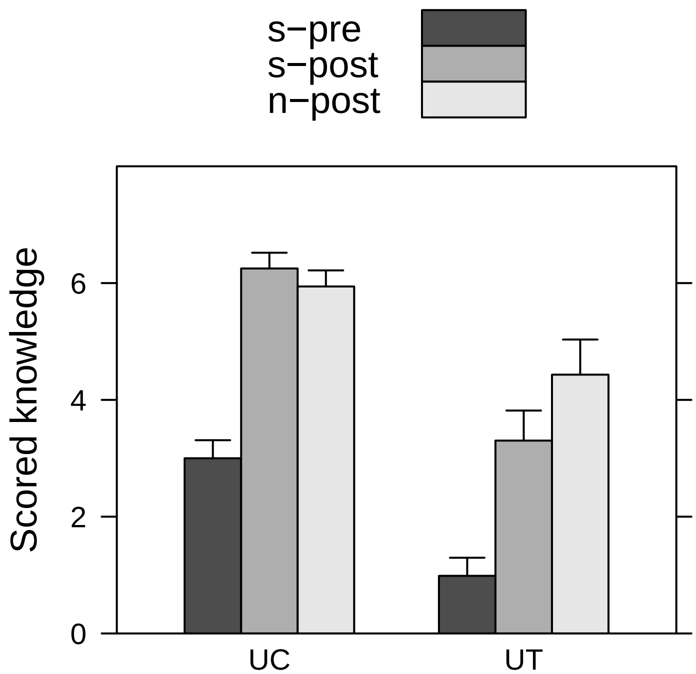


Figure 6.2: Combined scored knowledge for participants in our lecture room interventions. Posttest scores for both groups improved significantly relative to sandwich pretest scores ($p < .01$)

Beyond the blank-filling items, we statistically analyzed individuals' qualitative explanations—creating scoring rubrics for three central concepts:

Light Differentiating between the types of light entering and exiting the atmosphere

GHGs Atmospheric greenhouse gases' interactions with radiation

Energy The increased atmospheric retention time of energy

Inter-rater reliability was computed using a weighted modification of Cohen's κ, described in full in Appendix J. This reliability was high (weighted $\kappa = .71$ based on about one-third of the Berkeley data; $\kappa = .67$ across the full Brownsville dataset). Scores were generated based on three separate aspects of understanding captured in the coded texts: "Light," "GHGs," and "Energy." Significant improvements were observed across all three subscores ($p < .05$ for all six improvement possibilities across the two conditions and three subscore categories). We had no particular hypotheses, however, regarding specific effects of a given concept. Therefore, the data reported below and in Figure 6.2 use combined knowledge scores (and each sub-score contributes three of nine total possible points).

Improvements in participant knowledge were readily obtained with different approaches

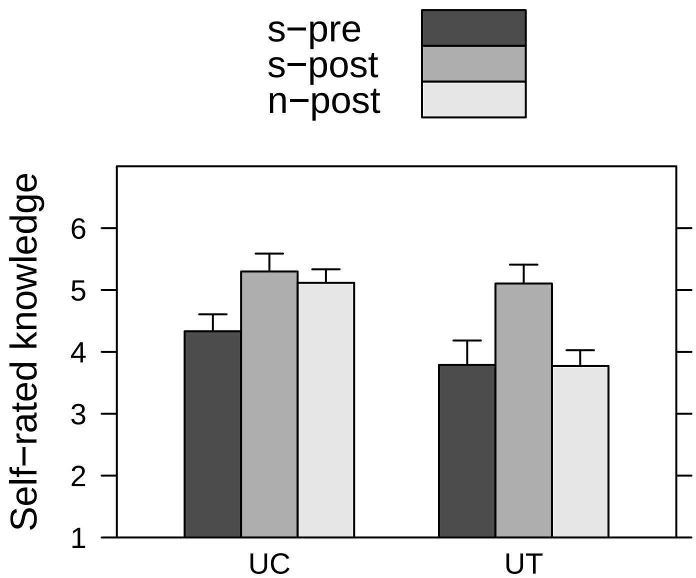


Figure 6.3: Gains in *self-rated* knowledge scores for participants in our lecture room interventions. Gains in our sandwich groups were significant ($p < .01$), while gains for our no-pretest groups were non-existent in the Brownsville group, and smaller (but still significant; $p < .05$) in the Berkeley group.

to analysis. For example, for our Berkeley students' responses, few items received a "mechanistic" code on the pretest (including incorrect codes; 11/42), but the majority of responses received such a code on the posttest (26/30 for the sandwich group, 39/43 for the no-pretest group).

Self-rated knowledge

Participants from both universities experienced significant gains in their self-rated knowledge after the intervention as well ($p < .01$). However, for our Brownsville students in the no-pretest group, they reported a posttest self-rated knowledge rating almost numerically identical to pretest ratings in the sandwich group. And while Berkeley students in the no-pretest group increased significantly ($p < .05$), that increase was numerically smaller than the sandwich group. These ratings are reported in Figure 6.3.

Global warming acceptance through mechanistic learning

To arrive at an easily comparable measure of global warming acceptance, we averaged together all of the items starting with "gw" used in this study (See Appendix A for a full list of these items). The lifsty item was omitted due to some concerns regarding multiple

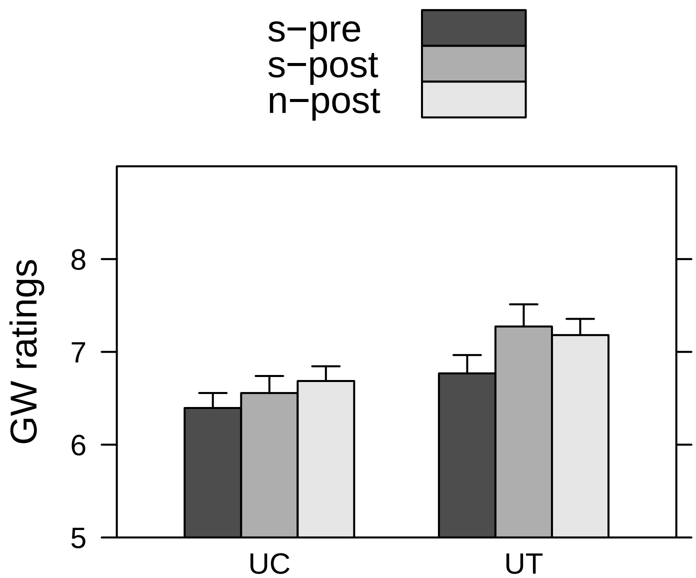


Figure 6.4: Changes in mean of Climate Change related beliefs and attitudes. Improvements from pre to posttest were significant for Berkeley students ($p < .05$ using a combined t-test with imputation). Improvements for Brownsville students were also significant using imputation ($p < .0001$), as well as looking only at the sandwich group ($p < .01$).

interpretation. This concern was in fact unfounded—this construct shifted similarly to the others—but we retain this set of items throughout our statistical testing to maintain consistency and genuine *a priori* hypothesis testing.

It may seem quite remarkable, but participants' global warming acceptance increased dramatically after our brief intervention, as predicted. Proportionally, participants shifted on average 14% closer to "extreme" agreement with climate change items. To assess this, we used all of the 73 Berkeley posttest ratings in a paired t-test, and used imputation for pretest scores for the no-pretest group (this method is detailed in Appendix K. We found a significant change in global warming acceptance on the posttests, as compared to pretest measures ($t(72) = 2.28$, $p = .01$). This result was replicated with the Brownsville surveys ($t(39) = 4.24$, $p < .0001$). These ratings are given in Figure 6.4.

Predicting naïve GW beliefs and attitudes

The relationship between knowledge and attitudes was also reflected in Berkeley students' naïve pretest data, in which participants' *self-perceived* ratings of their own global warming knowledge correlated significantly with their global warming attitudes ($r = .39$, $p = .01$). This was not the case with Brownsville students ($r = .15$, $p = .55$). This may be reflective of an overall lower self-perceived knowledge by Brownsville students. But

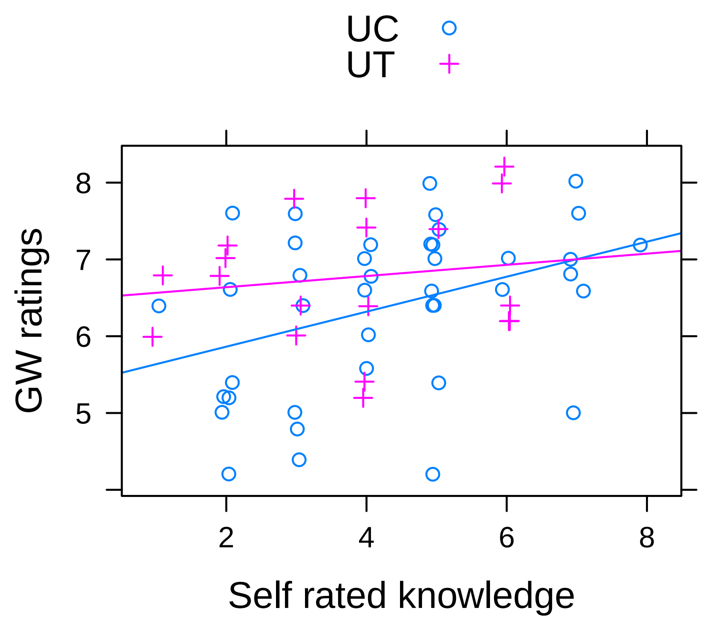


Figure 6.5: Relationship between naïve pretest self-rated knowledge and mean GW beliefs and attitudes. These data were only available for participants who took the pretest (i.e., the "sandwich" group). A significant relationship obtains for the Berkeley students ($p < .01$), while there is little relationship in the Brownsville sample (i.e., the slope of the regression line above is much flatter). Note that no (significant) prediction of attitudes was possible based on scored knowledge.

consider the findings above, in which we see an even more striking difference in terms of self-rated knowledge between our Berkeley and Brownsville populations. It seems that Brownsville students may simply have a much less grounded notion of their self-knowledge when they are not provided with any context on the matter. The relationship between self-rated knowledge and GW attitudes is depicted in Figure 6.5. Note that in this study, we conducted a fairly exhaustive examination of factors that might be predictive of naïve GW attitudes and beliefs. *Self-rated* knowledge was the only factor that was significantly predictive, and here only in one of our populations.

Surprise

Please recall that we had also predicted a between-conditions difference in surprise ratings due to reduced hindsight biases among the sandwich participants. The difference for Berkeley students was at the significance border-line ($t(42.08) = 1.65$, $p = .05$). These surprise ratings increased from a mean of 2.3 to 3.0 on a 9-point scale. It is a bit curious that their ratings are so low in general! The surprise ratings only reached "6" in the no-pretest condition (out of 9, with "5" being "somewhat surprising"), but were as high as "9" (i.e., "extremely surprising") in the sandwich condition. This difference in distribution is depicted in Figure 6.6. Among Brownsville students, surprise was uniformly higher, with a numerically similar difference between conditions, although this result was not significant ($t(38.1) = 0.92$, $p = .18$).

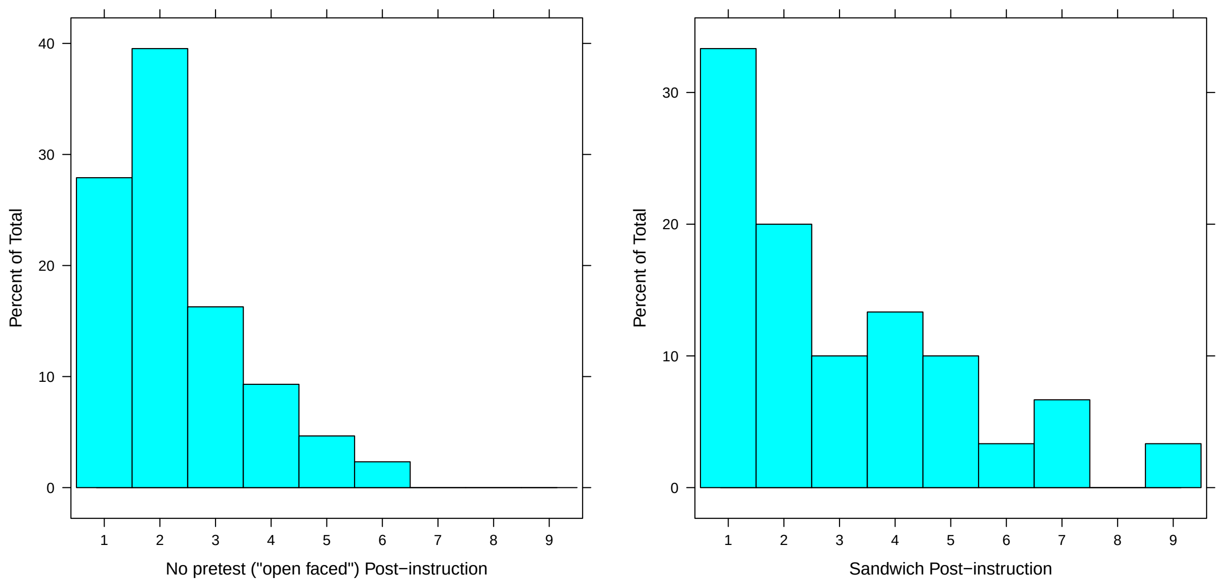


Figure 6.6: Distributions of surprise ratings for the sandwich and open-faced conditions in the Berkeley sample. Note that for the sandwich condition, the slight increase in "1" ratings (which may indicate resistance to the intervention) co-occurs with an increase (from none) in ratings 7–9.

I suspect that it is unlikely that individuals experienced the same kind of "visceral" surprise from the blurb that can be obtained by, for example, statistics we've used regarding issues like abortion and the death penalty. Further, while it may be due to a limitation of imagination, I have difficulty imagining an evolution item that would elicit that kind of surprise, either.

6.1.3 Full Tables of Results

As this was one of the very first experiments in this climate education program, we explored a large number of hypotheses. While it would be burdensome to treat all of these in the text, they are included here for posterity. Descriptions are provided in table captions.

Table 6.2: Summary of "improvement" results for Berkeley lecture room interventions. All results were *a priori* unless description starts with *"post hoc"*.

Result	p-value	Statistic
Individuals rarely provided mechanistic responses on the pretest (11/42 responses received a mechanistic code), but they often do on the posttest (26/30 sandwich and 39/43 no-pretest responses received a mechanistic code on the posttest). The statistical test is computed only for the sandwich group pre versus posttest, which has a lesser prevalence of mechanistic responses on the posttest of the two conditions	3.2×10^{-7}	Fisher's exact test (two-sided)
Misconceptions are common in the pretest but not the post test, total .38 pre to .10/.12 posttest (sandwich & no-pretest groups). Ozone .19 to .03/.02 (sandwich/no-pretest), wrong GHG .24 to .07/.09 (sandwich/no-pretest). (test on total misconceptions, comparing sandwich pretest to group-specific posttest)	0.01	Fisher's exact test (two-sided)
Participants don't mention energy leaving the earth until prompted. Specifically, of the four codes that deal with this topic, only 6 mention something about "trapped heat" in the pretest on the first (i.e., the only unscaffolded) question.	0.0002	Fisher's exact "two-sided"
Use of infrared is greater posttest than pretest. Goes from 0 to 16 / 22 in sandwich / no-pretest groups.	3.5×10^{-8}	Fisher's exact "two-sided"
Sandwich: GHG Objective knowledge scores improve after the blurb	5.08×10^{-5}	$t(29) = -4.75$ (paired)
No pretest: GHG Objective knowledge scores improve after the blurb	2.00×10^{-6}	$t(78.2) = -5.14$ (Welch)
Sandwich: Light Objective knowledge scores improve after the blurb	3.94×10^{-7}	$t(29) = -6.51$ (paired)
No pretest: Light Objective knowledge scores improve after the blurb	1.20×10^{-4}	$t(79.02) = -4.06$ (Welch)

Table 6.2: Improvements in Berkeley lecture room interventions, continued

Result	p-value	Statistic
Sandwich: Energy Objective knowledge scores improve after the blurb	0.04	$t(29) = -2.15$ (paired)
No pretest: Energy Objective knowledge scores improve after the blurb	4.60×10^{-4}	$t(80.82) = -3.6547$ (Welch)
Differences in mean GW attitudes are significant	0.013	$t(72) = -2.28$ (paired / imputed)
Sandwich pre to posttest: Increase in self rated knowledge is highly significant	1.40×10^{-5}	$t(29) = 4.96$ (paired)
No pretest: posttest (compared to sandwich pretest) increase in self rated knowledge is significant	0.014	$t(78.7) = 2.23$ (Welch)

Table 6.3: Summary of individual and group differences for Berkeley lecture room interventions. All results were *a priori* unless the description starts with *"post hoc."*

Result	p-value	Statistic
Surprise is significantly greater in sandwich group than no-pretest group	0.053	$t(42.08) = 1.65$ (Welch)
Posttest, slopes (i.e., correlations) between surprise and self-rated knowledge differ between no-pretest group (negative, significant) and sandwich group (which was numerically positive).	0.036	$t(69) = 2.137$ (interaction term in a significant linear model)
The no posttest group had a significantly higher word count than the sandwich group's posttest answers for the first (objective) knowledge question.	2.5×10^{-4}	$t(82.91) = -383$ (paired)
No pretest (posttest): Females are significantly more accepting of climate change than males	0.048	$t(40.19) = -1.71$ (Welch)
There is a significant positive correlation between number of times seeing An Inconvenient Truth and GW attitudes	0.022	$r(41) = .309$

Table 6.4: Summary of relationships between variables for Berkeley lecture room interventions. All results were *a priori* unless description starts with *"post hoc"*.

Result	p-value	Statistic
Surprise is significantly positively correlated with change in total objectively scored knowledge	0.047	$r(28) = .355$
Post hoc: There is a significant correlation between self-rated knowledge and GW attitudes on the pretest *only* (differences in self-rated knowledge are also insignificant). NB: we predicted the opposite result!	0.012	$r(40) = .386$
Post hoc: Sandwich: Negative correlation between posttest self-rated knowledge and *change* in objective score	0.011	$r(28) = -.458$
Post hoc: Sandwich: reversal in slope for the interaction term between scored and self-rated knowledge pre to posttest	0.047	$t(68) = -0.324$ (interaction term in a significant linear model)
Self-ratings on carefulness in reading are significantly correlated with posttest GW attitudes.	0.035	$r(41) = .279$
Rereading is significantly correlated with posttest GW attitudes.	0.055	$r(41) = .247$
Post hoc: Counter to our initial hypothesis, there is a negative correlation between rereading and posttest objective knowledge scores. NB: we predicted the opposite result!	0.019	$r(41) = -.356$

Table 6.5: Summary of "improvement" results for Brownsville lecture room interventions. All results were *a priori* unless the description starts with *"post hoc."*

Result	p-value	Statistic
Gains from pretest to posttest in mean GW attitudes are significant	1.30×10^{-4}	$t(38 = -4.02$ (paired / imputed)

Table 6.5: Improvements in Brownsville lecture room interventions, continued

Result	p-value	Statistic
Use of "infrared" is greater posttest than pretest. Goes from 0 to 7/6 in sandwich/no-pretest groups respectively (tested only for sandwich group)	8.00×10^{-3}	Fisher's exact "two-sided"
Sandwich: pre to posttest: Increase in self rated knowledge is highly significant	0.001	$t(18) = 18$ (paired)
Sandwich: GHG Objective knowledge scores improve after the blurb	0.0034	$t(18) = 3.38$ (paired)
No pretest: GHG Objective knowledge scores improve after the blurb	5.8×10^{-4}	$t(33.2) = 3.81$ (Welch)
Sandwich: Light Objective knowledge scores improve after the blurb	0.0095	$t(18) = 2.9$ (paired)
No pretest: Light Objective knowledge scores improve after the blurb	1.4×10^{-4}	$t(25.6) = 4.48$ (Welch)
Sandwich: Energy Objective knowledge scores improve after the blurb	0.02	$t(18) = 2.5$ (paired)
No pretest: Energy Objective knowledge scores improve after the blurb	2.9×10^{-4}	$t(36.8) = 4$ (Welch)

Table 6.6: Summary of failures to replicate and associated results with Brownsville lecture room interventions. All results were *a priori* unless the description starts with *"post hoc."*

Result	p-value	Statistic
No pretest: posttest increase in self-rated knowledge (compared to sandwich pretest) is not significant	0.51	$t(31.4) = -0.036$ (Welch)
Post hoc: Self-rated knowledge on posttest is significantly lower for no-pretest than sandwich group	0.0019	$t(36.6) = -3.36$ (Welch)
There is no correlation between self-rated knowledge and GW attitudes on the pretest	0.55	$r(17) = .15$

Table 6.6: Failures to replicate with Brownsville lecture room interventions, continued

Result	p-value	Statistic
Surprise is not significantly greater in sandwich group than no-pretest group	0.18	$t(38.1) = 0.92$ (Welch)

6.1.4 Discussion

This experiment replicates and extends findings from prior interviews and Cohen's (2012) survey, such that even rather well-educated people initially held mostly non-normative understandings of global warming's mechanism. Only 400 words later, though (roughly the duration of a TV commercial break), dramatic increases were observed in (1) mechanistic knowledge and (2) global warming acceptance. Further, the increases were found in divergent U.S. states and colleges. Certainly, this suggests that this educational intervention is a reasonable object of study! Differences in surprise ratings between the sandwich and no-pretest ("open-faced") conditions further support the notion that eliciting an explanation or theory prior to offering information increases surprise and reduces post hoc rationalization and hindsight bias. (On surprise, see Chapter 2; Munnich et al., 2007.)

In addition, we may note that there is scant difference between our sandwich and no-pretest conditions in terms of posttest attitudes (across our two populations, in one case the sandwich condition rates higher on posttest GW attitudes, while in the other, the no-pretest condition rates higher). Thus, it seems unlikely that a pretest incurs a greater burden of experimental demand over the core intervention (400 words followed by a posttest). Moreover, in some populations, the pretest may help to anchor self assessment (as with our UT Brownsville data). And finally, the sandwich intervention appears to increase reported feelings of surprise, and likely decreases post hoc bias (Rinne et al., 2006). Given these many benefits, in the work that followed, we standardized on using a sandwich style intervention.

6.2 Study 2: A Web-Based Intervention at UC Berkeley

Given the replicated demonstrations of significant attitude changes described above, we proceeded to assess whether the mechanism-explanation effects we had obtained were durable rather than transient. This study extended prior work by employing a delayed followup test several days after our posttest. We also wondered whether any "experimental demand" from the lecture room setting might have driven our prior results, so we provided the intervention on-line; that is, we assessed whether our materials would elicit significant attitude change even though students participated using their own computers, without experimenter observation. Thus we concurrently explored both the longevity (with a delayed followup) and format (online) aspects of our phenomenon. We also extended our prompts to incorporate more demographic and introspection queries.

6.2.1 Methods

Participants

Undergraduates (N = 80) were recruited through the Research Participation Program (RPP), administered by the University of California, Berkeley (UCB) psychology department. For this study, I specifically recruited conservative individuals and individuals with low GW acceptance based on data from the RPP prescreening phase (completed 324 times during the RPP prescreening period, though some students complete the form more than once—even though they get no additional credit!). One participant was excluded due to technical problems with his data. (this was one one of the participants who took the RPP pretest).

Of the analyzed participants, 46 were female, 33 male. In addition, this study reveals that (at least for Berkeley undergraduates), reported political party affiliation is somewhat unstable! Of the 36 students who *did* take the RPP prescreening survey, we obtained two reports of political party. 12 of the 36 students reported something different before and after, though they did not shift in a coherent way, as shown in Table 6.7. The complete breakdown of stated party affiliation is given in Table 6.8.

A final demographic measure of interest is conservativism. Given the large number of students who completed the RPP pretest survey, we were able to selectively contact those that we identified as the most conservative of the 300+ students who took the pretest. The RPP survey is standardized and shared with all experimenters. In this case, the RPP survey includes 2 questions for economic and social conservativism, while our survey only includes a question for "conservativism." A full treatment of the relationship between these variables is not justified here. Briefly, social conservativism is quite notably correlated with our later measure of non-specific conservativism ($r = .84$; economic conservativism was correlated at $r = .68$). The addition of economic conservativism as a second predictor or as a way to generate a mean conservativism score on the pretest *does* increase goodness of fit, but the reductions in error are small. Moreover, the Bayesian Information Criterion (BIC) weighs heavily in favor of a model using only social conservativism. However we slice things, though, participants reported higher values for non-specific conservativism after the intervention as compared to economic *or* social conservativism on the RPP pretest, as shown in Table 6.9. Thus, there is no unequivocal evidence here that folks have moved in a conservative direction, but certainly they weren't shifting in a more liberal direction! It does seem that recruitment of conservatives may have been somewhat successful, based on the 0.5-point higher score for participants who did the RPP pretest (and were thus the subject of targeted conservative recruitment).

Table 6.7: Stated party affiliation for the 12 of 36 individuals for whom we have two responses for political party. It is difficult to discern a clear pattern here, beyond a general instability in stated party affiliation—though descriptively, there were more switched *to* a party than *away* from a party to a choice like "none" or "independent." There are also more individuals claiming affiliation with conservative parties (Libertarian or Republican) than prior to the intervention. For example, an equal number of "none" responses shifted to "democrat" and "republican." Similarly, three "democrats" shifted away, while four individuals shifted to become "democrats." Notably, no individuals shifted to "green."

RPP pretest	After Intervention
democrat	libertarian
democrat	none
democrat	independent
independent	democrat
independent	libertarian
none	democrat
none	democrat
none	republican
none	republican
none	independent
decline to state	none
decline to state	democrat

Materials and procedure

The survey and instructional materials were largely analogous to those reported in Section 6.1. The primary difference was that administration was conducted entirely online, using the Qualtrics Inc. (Provo, UT) system, much as in Chapter 5. Eight items were added to pre and posttest attitude surveys to add reliability to the related RTMD metrics (specifically nationalism and religious affinities; these metrics will be reported elsewhere). The engage item was added to provide a clearer assessment of behavioral intentions (again, see Appendix A). Five further questions were introduced immediately following the instructional material to elicit introspection (about embarrassment, disagreement, etc.). These questions were distilled from the result of guided interviews with participants in a pilot study that was quite similar to this one, but carried out with participants using a computer in our laboratory testing space.

RPP recruitment allowed us to administer a pretest to about half of the undergraduates ($n = 36$) between 8 and 26 days ($\mu = 18.5$ days) before any of the 80 students participated in the study, which may have allayed test-retest effects (although we found little evidence for them in the experiments reported above). Thus, as with the above, some participants received the full survey testing "sandwich" while for others we lacked the demographics

Table 6.8: Total number of individuals reporting affiliation with a given party. Note that we aggressively recruited conservatives to participate from the RPP prescreening population, yet still ended up with more republicans coming from the remainder of our population!

	RPP pretest (fraction)	Study posttest (fraction)
decline to state	3.00 (.08)	1.00 (.01)
democrat	12.00 (.33)	26.00 (.33)
green	0 (0)	1.00 (.01)
independent	2.00 (.06)	3.00 (.04)
libertarian	1.00 (.03)	4.00 (.05)
none	17.00 (.47)	37.00 (.47)
republican	1.00 (.03)	7.00 (.09)
total	36.00 (1.00)	79.00 (1.00)

Table 6.9: Conservativism scores from the RPP pretest and intervention posttest. Note that scores are not directly comparable, as the differing levels of specificity entailed somewhat different wording on the questions. That said, individuals were certainly close to the middle of the scale, and if anything moved closer to the middle on average

	Reported conservativism		
	Social (pretest)	Economic (pretest)	Non-specific (posttest)
Full sandwich	3.86	3.25	4.31
No (RPP) pretest			3.80

and attitude portion of the pretest. Note that this differs from previous “no-pretest” groups, in that this group still provided their naïve description of the mechanism of global warming. Thus, we will refer to our conditions as having completed the “full sandwich” or a “partial pretest.” Note that demographics were *also* collected following the primary intervention, because we did not have them for all participants. Thus, for some participants, we collected some demographic information twice (some of which is reported on below). A delayed posttest was given to all participants between 1 and 8 days later ($\mu = 4$ days). The delay was not controlled, and as such, we should avoid making inferences about the timecourse of retention over that delay. This delayed test had the same format as the immediate posttest. This range of delays prior to this test was used to assess the timecourse of retention in planning subsequent studies. We lack the power to test forgetting over time here, but we did not observe any numerically.

Analysis

As we have now collected 3 observations of the same measure within subjects, for many of the tests in this study, we shifted from using t-tests to using robust mixed-effects regression. For this, we used the lmer function from the lme4 package in R, which handles a variety of unbalanced designs. Goodness of fit for these regression models was assessed utilizing Type-II sums of squares with the Anova function from the car package. Once a null model was rejected, Tukey-style contrasts were computed as simultaneous comparisons with glht from the multcomp package. Unless otherwise stated, all tests were *a priori*.

6.2.2 Results

As before, a full report of statistical tests for this study is given at the end of this section in Table 6.12. Statistics are provided in-text when appropriate.

Scored knowledge: Learning the global warming mechanism

In general and as anticipated, we replicated results from Study 1 (Section 6.1) and extended them by finding that shifts were retained over the mean (four-day) delay. Objectively scored knowledge was comparable to previously tested UC students, rising from 3.8 on pretest to 6.5 posttest and 6.3 on delayed test (gains from pretest were significant for both subsequent scores at $p < .0001$; drop from post to delayed was not significant). These are plotted in Figure 6.7.

Self-Rated knowledge

Self-rated knowledge means also increased markedly from pre to posttest (4.5 to 5.6 on a 9-point scale, $t(79) = 8.5$, $p < .001$). Retention of this increase, gratifyingly, was also noted on the delayed posttest (5.2, $t(79) = 6.2$, $p < .001$). The immediate increase in self-rated knowledge, replicates results from the "sandwich" interventions in Study 1 (Section 6.1). Scores are shown in Figure 6.8.

Global warming acceptance through mechanistic learning

GW belief ratings (with higher ratings being more in concert with science's consensus) increased numerically from a 6.20 pretest mean to a 6.54 posttest mean (a healthy improvement on our 1–9 Likert scales!). Some of this improvement diminished over the following days, but most was retained: the mean score on the delayed posttest was 6.44. significant using a naïve imputation approach. A more recent analysis, however, implies that these results may not be so clear. In particular, using the lmer package yielded insignificant improvements even from pre to posttest. Upon closer inspection, the breakdown for GW attitudes is quite different across full-sandwich and partial-pretest conditions, as seen in Table 6.10. Unlike in previous studies, the full-sandwich conditions (i.e., the condition with

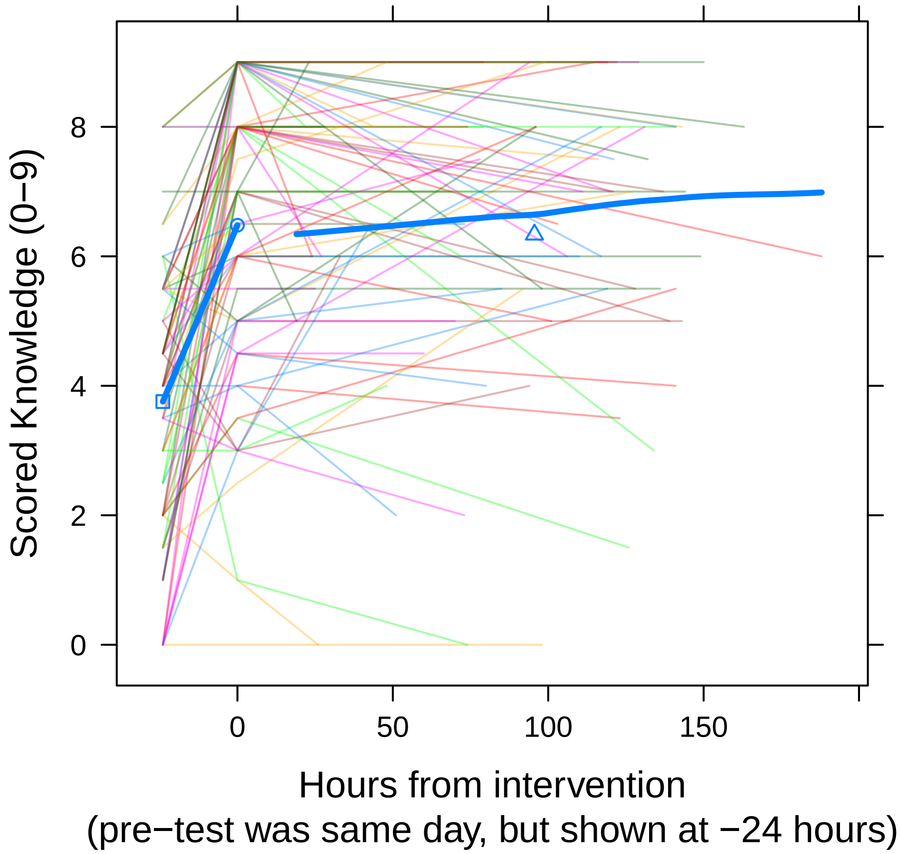


Figure 6.7: Objectively scored knowledge before and after our intervention. Faint lines represent individual performance, while the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). Mean delayed test score is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. Note that the pretest here was immediately prior to instruction (but shown at -24 hours). Participants overwhelmingly increased their scored knowledge following the intervention. While this study was not designed to assess forgetting in an individual over time (participants chose their own time to take the delayed test), it is interesting to note that earlier respondents tended to be scored lower than later respondents. Most importantly, it is easy to spot individuals who improved or worsened over the delay.

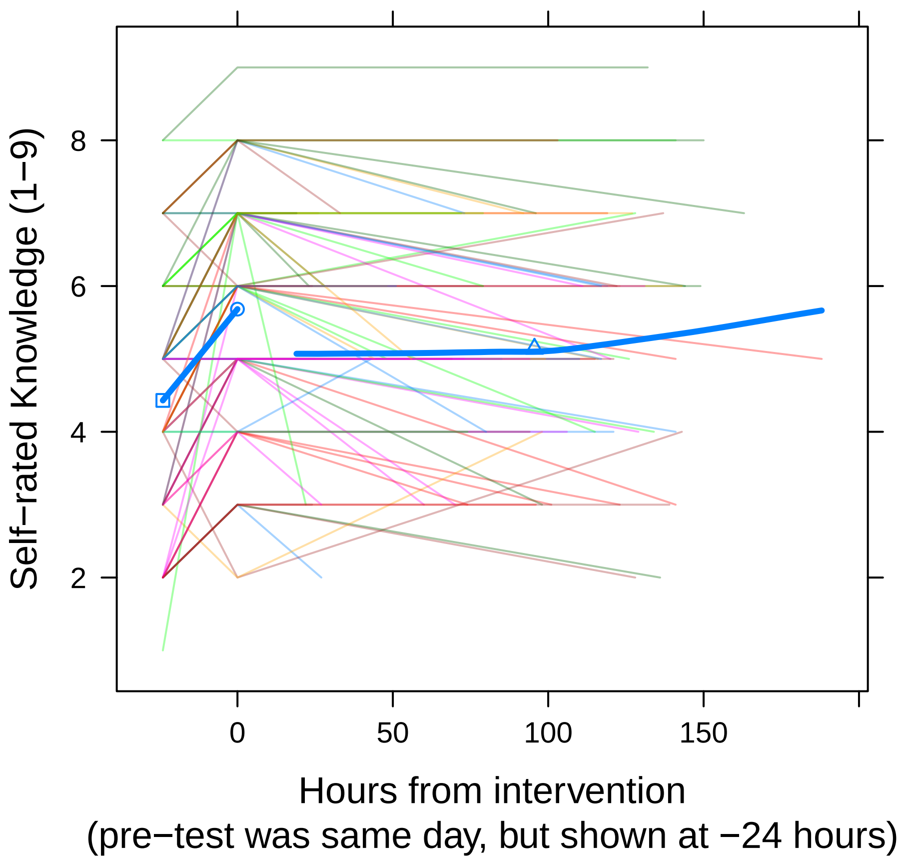


Figure 6.8: Self-rated knowledge before and after our intervention. Note that the pretest here was immediately prior to instruction (but shown at -24 hours). Again, the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). The mean delayed self-rating is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. As in Figure 6.7, there is a modest upward trend from early to late respondents, but recall that individuals chose their own time to respond to the delayed test.

GW attitudes available from the RPP pretest) contained a good number of participants who were recruited for their conservativism (and there is a clear condition difference on this measure). Thus, we are not as justified in assuming the pretest is representative of *all* participants. In the terminology described by Fox (2008), these data are not *missing at random* (MAR). Thus, we cannot apply the same sort of simple imputation to obtain a valid result. Individual scores are depicted in Figure 6.9. While we may be able to salvage this data using more sensitive techniques, such efforts are unwarranted given the robust successes reported in Sections 6.1 and 6.3, and the agreement here with the pattern obtained there.

Table 6.10: Mean GW ratings for full-sandwich and partial-pretest conditions. Note that the increase from pretest to posttest is only 0.16 for individuals in the sandwich group, and delayed test results are lower than where they started. The partial-pretest group, however, starts much higher and stays higher.

	Pretest	Posttest	Delayed test
Partial (no RPP) pretest	NA	6.68	6.65
Full sandwich	6.20	6.36	6.18

Surprise

As explained in the Methods, some participants received a full "sandwich" intervention, including beliefs and attitudes administered during RPP prescreening. All participants, however, completed a knowledge pretest. Thus, we were able to informally compare these two partially novel conditions with our previous profiles for reported surprise (i.e., with Figure 6.6). The surprise results from our current conditions are provided in Figure 6.10, and provide further support for the centrality of the initial *knowledge* pretest.

In addition to repeating the surprise question from Study 1, on the basis of informal interviews, we had included another related question asking if people were "surprised or embarrassed about their own lack of knowledge." This item does seem to have elicited somewhat higher scores, with a mean of 3.6 versus 2.9 for the original surprise question. Interestingly, this not only appears to (at least partially) alleviate a floor effect obtained with the straight surprise question, but it also demonstrates a reversal of the ranking of affective experiences as compared to the study with effective, representative numbers in Section 5.2. The distributions for these responses are shown in Figure 6.11.

Micro-analysis of GW ratings

Table 6.11 reports the mean rating across participants for agreement with individual items. (For the full text of these items, see Appendix A.) The largest gains were found in agreeing with item `gw1_2`: "Human activities are largely responsible for the climate changes…" (a 0.25 gain) and certainty that global warming is occurring (a 0.19 gain). In

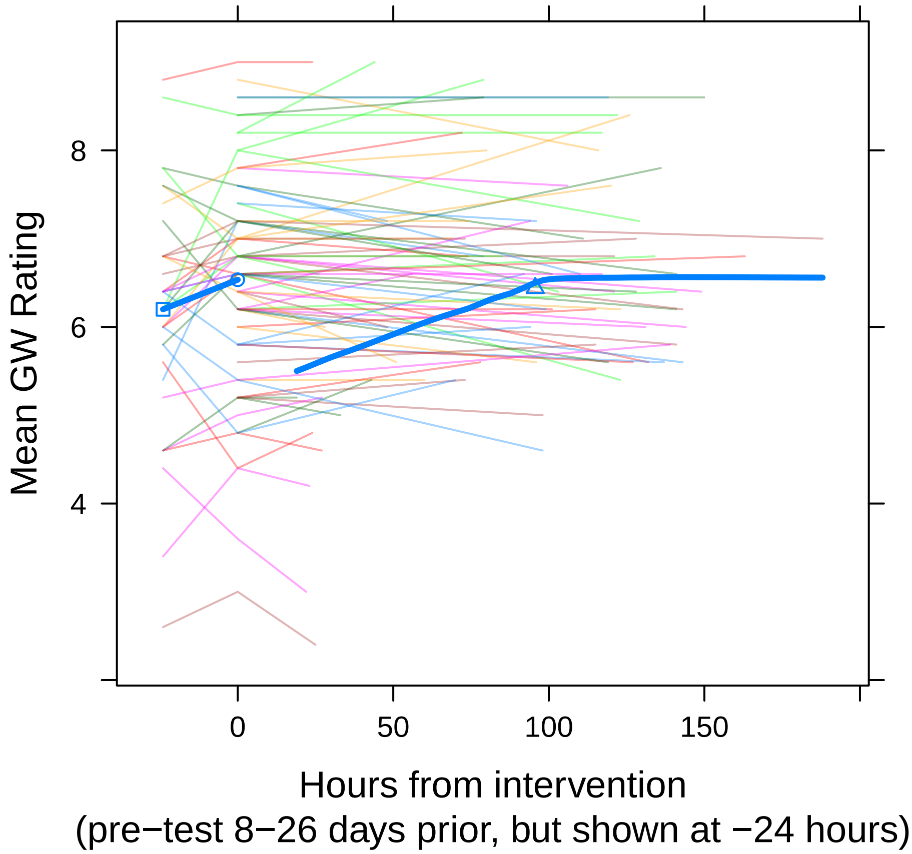


Figure 6.9: Reported global warming (GW) beliefs and attitudes before and after our intervention. Note that the pretest here was administered on average 18.5 days prior to instruction as a part of UC Berkeley's undergraduate participant pool (RPP) prescreening (but is shown at -24 hours). Once again, the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). The mean delayed self-rating is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. We can also notice here that all of the individuals who yielded the lowest ratings for normative beliefs and attitudes took the delayed test quite early. But given self-selection of the time for the delayed test, we should certainly be taking those initial dips with a grain of salt. We cannot assume that the figure represents an actual retention timecourse!

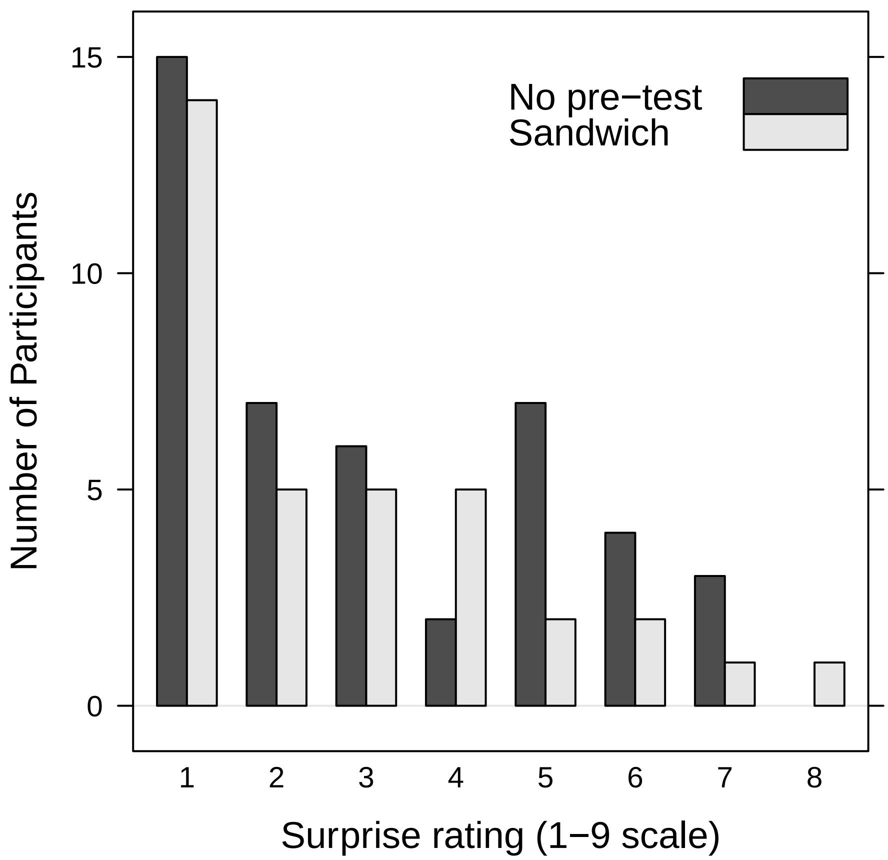


Figure 6.10: Surprise ratings for individuals in Study 2. The shape of both of these distributions much more closely approximate that of the sandwich participants in Study 1 (See, e.g., Figure 6.6). In particular, both conditions here yield surprise values in the 7+ range. Thus, it seems to matter less if or when participants receive the beliefs and attitudes surveys—the higher surprise values appear to be linked to asking participants to "put their cards on the table" with a knowledge pretest.

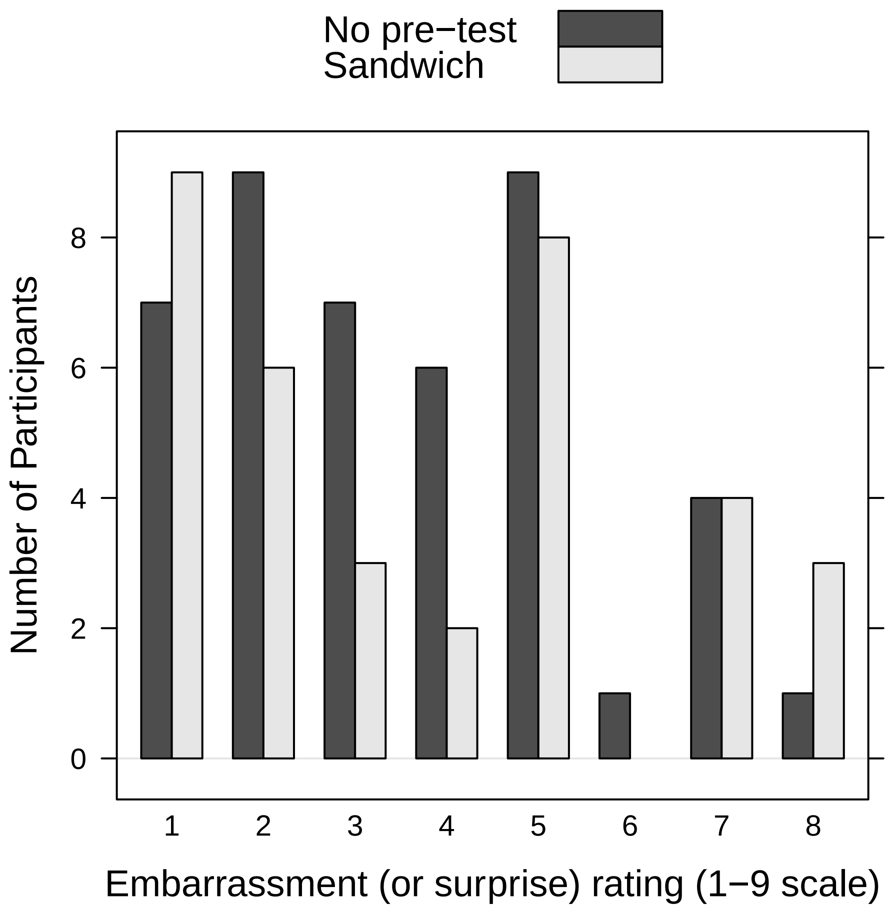


Figure 6.11: Ratings of "embarrassment or surprise at their own lack of knowledge" for individuals in Study 2. Distributions are again similar in notable ways between groups. In particular, there appears to be less of a floor effect here than we obtain with our original phrasing of surprise (consistent with a higher mean). Individuals are again using most of the range, including ratings in the 7+ range.

Table 6.11: Mean GW ratings, online with UC Berkeley undergrads. For means over these items, please refer to Table 6.10

	Pretest	Posttest	Delayed test
gw1_2	6.61	6.86	6.36
gw2_1	5.19	5.31	5.25
gw2_2	6.61	6.81	6.67
gw2_3	5.81	5.97	5.97
gw2_4	6.78	6.86	6.67
engage	5.91	5.86	6.11
lifesty	4.83	5.11	4.94

general, gains were fairly consistent across all GW measures, ranging only down to 0.08 at the lowest (the relatively broad, accusatory item gw2_4: "humans are severely abusing the environment"). Interestingly importance of lifestyle changed the most (0.27, though this was not included in the tested average GW variable). Expectation of engagement, dishearteningly, clocks in at a 0.05 *drop*!

Correlations

Scored knowledge and self-rated knowledge are significantly correlated pretest, so participants have reasonable meta-cognition here. However, unlike with our Berkeley students in Study 1, there was only a not quite marginally significant (and smaller) correlation between naïve pretest self-rated knowledge and GW attitudes ($r(34) = .27, p = .11$). Thus, while the validity of self-rated knowledge (as a predictor of actual knowledge) appears robust across Berkeley student populations, the relationship between self-rated knowledge and GW attitudes is weak, or perhaps even spurious.

6.2.3 Discussion

In sum, this study extends the finding that well-considered information, even received online, increases GW acceptance and behaviorally relevant attitudes; the conceptual changes that result from reading even 400 words have notable longevity. As we'll see below, these effects have been replicated with members of the general public as well. Computer-based interventions often scale well, enhance reliability, and prove cost-effective; given our results, we recommend the online distribution of mechanistic explanations, especially about climate change. A collection of videos to this effect is currently online at http://HowGlobalWarmingWorks.org.

Delayed followup tests occurred over a range of delays, and retention here was used in determining intervals for future studies. Given the almost total lack of mean forgetting over the observed interval, we start our subsequent study's delayed test after a longer delay (specifically, for a one week period starting 4 days after the initial intervention).

Table 6.12: Summary of results from Study 2.

Result	p-value	Statistic
Scored knowledge is different across the three testing times in a repeated-measures ANOVA.	2.2×10^{-16}	$\chi^2(2) = 128.39$
Scored knowledge is higher in the posttest than in the pretest.	5×10^{-3}	$z = 10.09$
Scored knowledge is higher in the delayed test than in the pretest.	5×10^{-3}	$z = 9.52$
Scored knowledge is not significantly lower in the delayed test than in the posttest.	0.813	$z = -0.612$
Self-rated knowledge is different across the three testing times in a repeated-measures ANOVA.	2.2×10^{-16}	$\chi^2(2) = 110.25$
Self-rated knowledge is higher in the posttest than in the pretest.	1×10^{-5}	$z = 10.47$
Self-rated knowledge is higher in the delayed test than in the pretest.	1×10^{-5}	$z = 5.922$
Self-rated knowledge is lower in the delayed test than in the posttest.	1.5×10^{-5}	$z = -4.55$
We are unable to reject the null hypothesis that GW attitudes are the same across the three testing times given this dataset (but note that the pattern is still consistent with Studies 1 and 3).	0.21	$\chi^2(2) = 3.09$
Self-rated knowledge is significantly correlated with scored knowledge on the pretest.	4.7×10^{-6}	$r(76) = .49$
Self-rated knowledge is *not* quite marginally correlated with GW attitudes on the pretest.	0.11	$r(34) = .27$

6.3 Study 3: An Intervention with Amazon Mechanical Turk

6.3.1 Methods

Experimental methods in this study were nearly identical to Study 2, above (detailed in Section 6.2.1). The primary difference here was that participants were recruited through Amazon's Mechanical Turk (discussed in more detail in Section 5.2.1. In general, the remainder of the methods focuses only on the differences with those reported in Section 6.2.1.

Participants

"Workers" on the Amazon Mechanical Turk platform (N = 41) were recruited to complete a two-part survey. Language used in recruitment made no mention of climate change, and was titled "Politics, Science, and Your Attitudes." Approximately 75% of these individuals (n = 30) completed the delayed test. After removal of problematic participants (as described in the "data quality" section below), we were left with 38 participants in the primary intervention, with 28 in the delayed test. 17 of the 38 retained participants were female. Two of our participants reported as being born outside the United States, but residing here for at least 22 years. Stated party affiliations are listed in Table 6.13. Mean conservativism was 3.9 out of 9, which is comparable to our college students above. Participants, however, were far more likely to declare being a Democrat or Republican than our college students above. While the sample is still clearly biased towards the Democratic/liberal end of the political spectrum, the ratio between Republicans and Democrats is far less extreme than in previous samples. Note also that this is in line with the results reported by Richey and Taylor (2012), in which approximately 73% of polled workers on Mechanical Turk reported voting for Obama (versus approximately 51% in the election results).

Table 6.13: Stated party affiliations for participants in Study 3 (through Mechanical Turk).

Party	Number (percentage)
democrat	22 (58%)
republican	8 (21%)
independent	6 (16%)
none	1 (3%)
other	1 (3%)
total	38.00 (100%)

Materials and procedure

On the basis of our previous two studies, the most convenient design that still elicits the highest surprise was pretest immediately before the mechanism explanation. Moreover, this pretest can include all measures—eliciting descriptions prior to instruction maximized surprise, and the timing of other attitudinal measures seemed to have little effect (or may have actually reduced our power to observe shifts in attitudes). Thus, here, all participants completed a pretest immediately prior to the main intervention and posttest, most closely approximating the sandwich intervention from Study 1 with the addition of a delayed test some days later.

The most interesting difference in procedure between this and Study 2 is the utilization of Mechanical Turk. A minor change is that the retention interval was extended to begin 4 days later (with participants choosing when to take the delayed posttest), with the longest interval being 10.8 days ($\mu = 5.5$ days). A reminder was sent through the Mechanical Turk bonus system, by which participants were paid 5 cents and given a message that the delayed test was open for completion. The majority of payment was provided at the time of completion of the delayed test. Perhaps unsurprisingly, most participants responded shortly after the first reminder, and the majority of the remainder responded shortly after the final reminder.

Materials were identical, apart from the addition of a question requesting the individual's "worker ID"—a unique identifier used to track and pay individuals, akin to a social security number assigned by Amazon.

Data quality on Amazon Mechanical Turk

As discussed in Section 5.2.1, the use of an anonymous, on-line labor pool raises a number of additional concerns about data quality. Here, we apply the same approach to IP-based testing for participant honesty about their location, and checking against re-takes. An additional concern arises in this study—somewhat bizarrely, as this does not reduce time required, or increase payment—participants may copy and paste from online sources.

Again, Amazon attempted to prevent re-takes as requested, and no duplicate IP addresses were detected. Two participants, however, lacked an IP address. Manual inspection revealed nothing anomalous about these participants, however, so they were retained.

The Mechanical Turk system was again set to allow only individuals in the U.S. as participants. Verification of location was again obtained using GeoLite data created by MaxMind (available from http://www.maxmind.com). An additional check was available given IP addresses captured during both the primary intervention as well as the delayed test. On the primary survey, participants indicated the state they were in. Participant IP addresses from both intervention and delayed followup surveys were subsequently checked against this reported location. Here, two individuals IP's appeared to be located in India and Turkey, and thus these participants were excluded. Most other participants had IP addresses that resolved to the state they claimed to be from. One participant's IP address was not listed in the MaxMind database, and was traced to Hughes Net, a U.S.-only satellite internet provider.

Checking for plagiarism is a well-trod topic in academia today, though there is no one clear approach. For the purposes of this survey, we relied on a combination of coder judgement, and automated checking using Google's Custom Search API. A complete description of the process I used is provided in Appendix H. In summary, I used a combination of my own and others' judgement to identify clear cases of plagiarism. I then developed an approach using Google search to identify texts that already existed on the internet, being sure that this method caught the clear cases we'd identified using manual inspection. Ultimately, two individuals clearly copied and pasted materials from the internet. These could have both been identified by the unusual presence of extended unicode characters in the text of their answers, as well as the presence of newlines.

Note that one individual was *both* foreign *and* copied text directly form the web. Thus, we exclude only 3 subjects total.

While survey consistency seemed reasonable for all remaining participants, it did highlight two individuals who answered almost exclusively 1 or 9 to all items. These individuals were retained in the analyses below, as they still provided information on their beliefs. In particular, their responses were consistent on reverse-coded catch-trials, and their textual descriptions were consistent with taking the task seriously.

6.3.2 Results

Overall, we replicate our central results from Studies 1 and 2 above. As before, full statistics are reported in Table 6.15 at the end of this section. Only information directly supporting the narrative is included in the textual exposition.

Problems and data quality

While the text of the delayed test was like the pretest (i.e., the first question made no affordance for "if you would add anything") three individuals still put "nothing to add" for all three questions. These individuals were retained as, if anything, they should weaken retention effects after a delay. Moreover, they all still answered the Likert survey questions. While not definitive, this is certainly one clue that individuals recruited through Mechanical Turk are perhaps rushing a bit more than our previous college samples. This may also be the reason we observed two individuals who only provided 1's and 9's above.

Scored knowledge: Learning and global warming mechanism

Again, we replicated results from Studies 1 and 2 by finding that shifts were retained over the mean 5.5-day delay. Scored knowledge was comparable to previously tested students, rising from 1.9 on pretest to 4.8 posttest and 3.9 on delayed test (on a 0–9 scale; gains from pretest were significant at $p < .002$ for both subsequent scores). These are plotted in Figure 6.12.

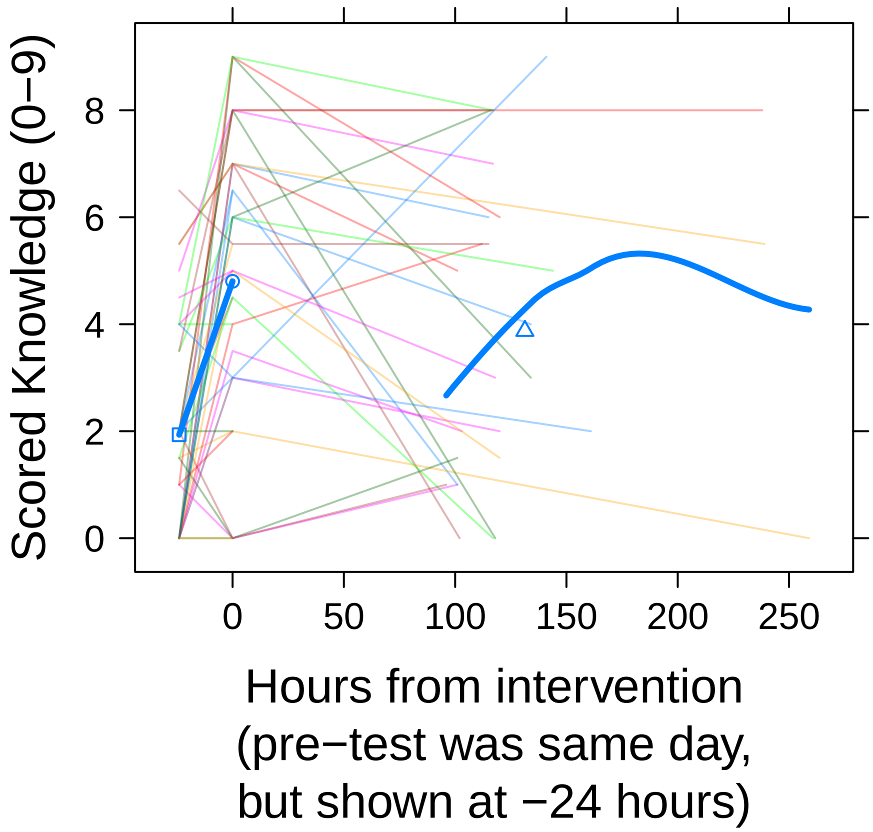


Figure 6.12: Objectively scored knowledge before and after our intervention. Faint lines represent individual performance, while the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). Mean delayed test score is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. Note that the pretest here was immediately prior to instruction (but shown at -24 hours). Participants overwhelmingly increased their scored knowledge following the intervention. As before, this study was not designed to assess forgetting in an individual over time, but it is interesting to note that, cross-sectionally, we again obtain relatively high scores later in our delayed testing. In contrast to the study in Section 6.2, most participants took the study the day they received their reminder, as is reflected by the mean delay being closer to the left edge of the LOESS regression line.

Table 6.14: Mean GW ratings for the Mechanical Turk mechanism intervention. Note that means are only computed over the items with codes starting with "gw." Item codes are explained in Table A.2 in the Appendix. Note that these gains, taken individually, are not significant (though gains in their mean is significant).

	Pretest	Posttest	Delayed test
gw1_2	6.87	7.34	7.07
gw2_1	6.03	6.55	6.25
gw2_2	6.82	7.05	7.04
gw2_3	6.05	6.42	6.36
gw2_4	6.89	7.05	6.86
engage	6.37	6.42	6.46
lifsty	5.37	5.63	6.04
mean	6.34	6.64	6.58

Self-rated knowledge

As with our UC Berkeley students, participants demonstrated a significant correlation between self-rated knowledge and actual knowledge ($r(36) = .49$; $p = .0017$). Self-rated knowledge means also increased markedly from pre to posttest (4.2 to 4.7 on a 1–9 scale, $p < .01$). Retention of this increase, however, failed to exceed our significance threshold (4.6, $p = .11$). These scores are reported in Figure 6.13. It seems not unlikely, given our previous results, that more subjects would yield a significant difference here. Even so, it should be noted that studies with college students tended to yield increases of over a point on self-rated knowledge (at least with "sandwich" interventions such as this one). Thus, even our significant pre to posttest gains are small compared to the UCB population.

Global warming acceptance through mechanistic learning

Happily, while introspection regarding knowledge seemed to yield somewhat smaller gains than in Studies 1 and 2, GW belief ratings increased significantly from a 6.34 pretest mean to a 6.64 posttest mean ($p = .001$). Some of this improvement diminished over the following days, but most was retained: the mean score on the delayed posttest was 6.58 ($p = .006$). Note that these values are very much in line with those obtained in the previous studies. Individual attitude breakdowns are reported in Table 6.14. Individual data are depicted in Figure 6.14.

Surprise

As in Study 2 above, individuals ranked their "embarrassment or surprise at their own lack of knowledge" higher than straight surprise. Mean values were 2.9 and 4.1 respectively for straight surprise and surprise/embarrassment, quite comparable to Study 2 in this

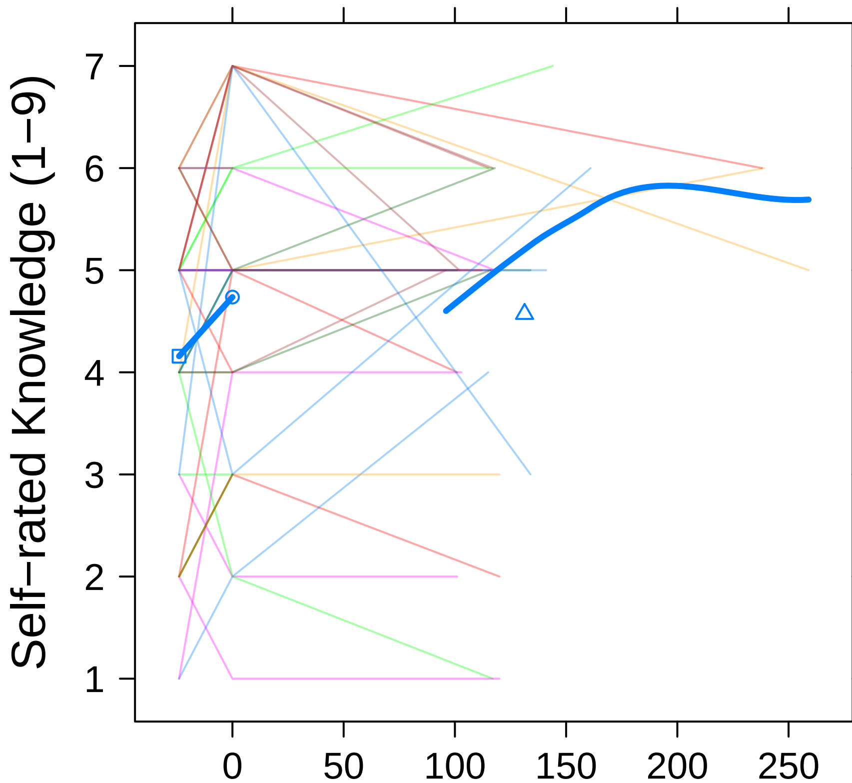


Figure 6.13: Self-rated knowledge before and after our intervention. Again, faint lines represent individual performance, while the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). Mean delayed test score is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. Note that the pretest here was immediately prior to instruction (but shown at -24 hours). Participants overwhelmingly increased their scored knowledge following the intervention. As before, this study was not designed to assess forgetting in an individual over time, but it is interesting to note that we again obtain relatively high scores later in our delayed testing. In contrast to Study 2 in Section 6.2, most participants took the study the day they received their reminder, as is reflected by the mean delay being closer to the left edge of the LOESS regression line.

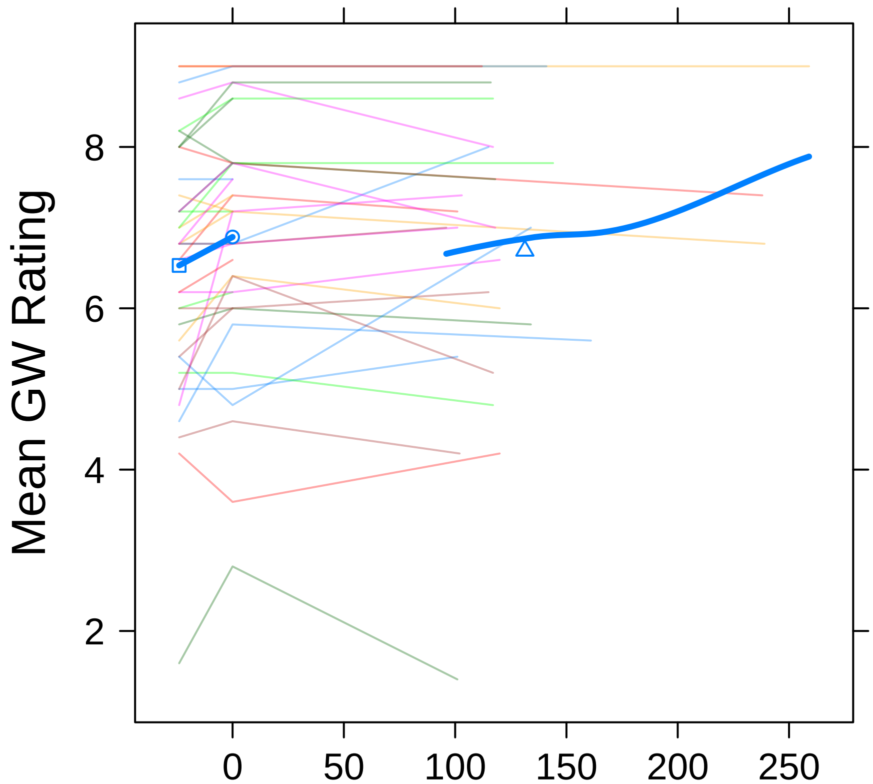


Figure 6.14: Reported global warming (GW) beliefs and attitudes before and after our intervention on Mechanical Turk. A significant increase in participant ratings is observed after the intervention, and this gain is significantly retained (compared to pretest) in the delayed test. Once again, faint lines represent individual ratings, while the bold line connects mean pre and posttest scores (indicated by a square and circle respectively). Mean delayed test score is indicated by a triangle at the mean time for taking the delayed test. A LOESS robust, smooth regression line is fit over the time period in which participants completed the delayed test. The pretest here was immediately prior to instruction (but shown at -24 hours).

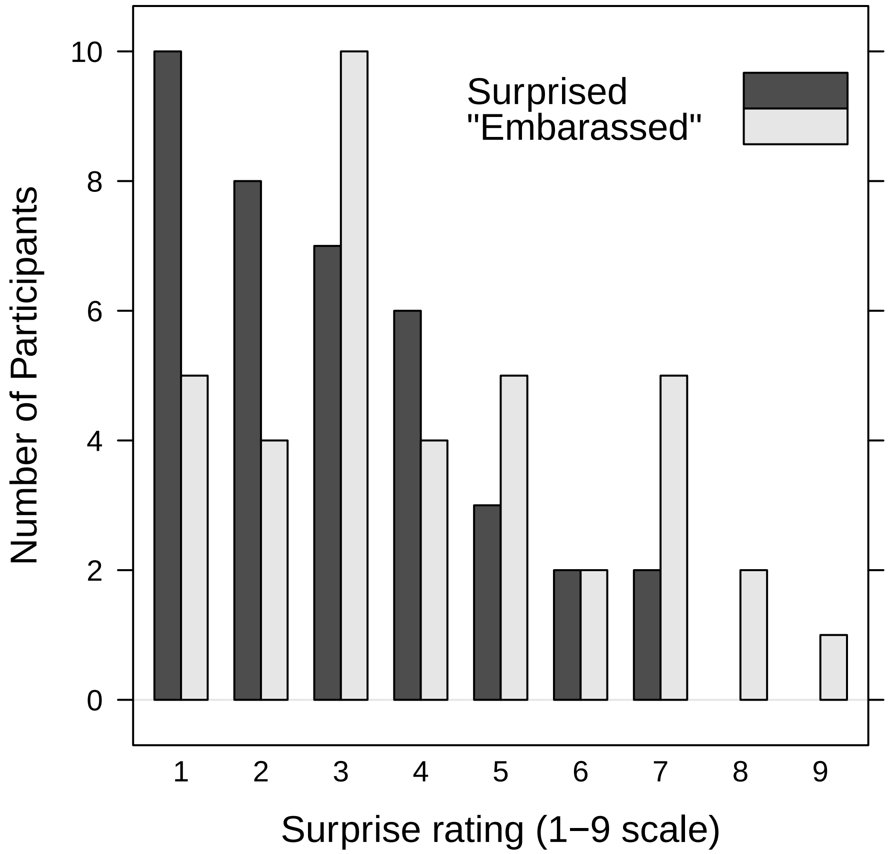


Figure 6.15: Surprise and "embarrassment or surprise at their own lack of knowledge" ratings for individuals in Study 3. Embarrassment appears to suffer less of a floor effect than surprise.

chapter, and again a reversal from Study 2 in Section 5.2. Distributions for both questions are depicted in Figure 6.15).

Factors with no observed effect

And, as in Study 2, we again failed to replicate a significant relationship between self-rated knowledge and GW attitudes on the pretest. Given these repeated failures to replicate, it seems prudent to abandon the naïve relationship between self-rated knowledge and GW attitudes (at least in this general sense).

6.3.3 Discussion

Getting interventions into classrooms is a difficult process that might take years (and even then, might not be successful). This study provides an evaluation of a more efficient online approach to delivering our educational intervention to the general population. Often in psychological studies, one attempts to control for self-selection effects. Here, we recruited individuals from Mechanical Turk who were willing to participate in a study regarding "Science, Politics, and Your Attitudes." Thus, our sample was likely representative of individuals that might engage with our materials were they to be made generally available on line (as is now being done on HowGlobalWarmingWorks.org). This is probably more useful than attempting to, for example, balance particular demographics in the study population.

In addition, as noted above, workers on Mechanical Turk appear to be more liberal (and perhaps, thus more GW-accepting) on average than the general population. If one were interested in reaching a broader segment of the population, it seems incumbent upon the experimenter to target populations with better conservative representation. In particular, it would seem useful to develop recruitment methods for attracting the population of interest (that is, one should experiment with something like one's study description to discover a way of piquing the interest of, say, more Republicans). More generally, however, we have now seen three populations, all of which have been productively engaged by the materials in our global warming mechanism intervention.

Table 6.15: Summary of results from Study 3.

Result	p-value	Statistic
Scored knowledge is different across tests in a repeated-measures ANOVA.	2.7×10^{-8}	$\chi^2(2) = 34.87$
Scored knowledge is higher in posttest than in pretest.	5×10^{-5}	$z = -5.86$
Scored knowledge is higher in delayed test than in pretest.	0.0014	$z = 3.30$
Self-rated knowledge is different across tests in a repeated-measures ANOVA.	0.021	$\chi^2(2) = 7.69$
Self-rated knowledge is higher in posttest than in pretest.	0.0088	$z = 2.72$
Self-rated knowledge is *not* quite marginally higher in the delayed test than in pretest.	0.11	$z = 1.69$

Table 6.15: Results from Study 3, continued.

Result	p-value	Statistic
Mean GW attitude is different across tests in a repeated-measures Anova.	0.000 91	$\chi^2(2) = 14.01$
Mean GW attitude is higher in posttest than in pretest.	0.000 85	$z = 3.445$
Mean GW attitude is higher in delayed test than in pretest.	0.0062	$z = 2.84$
Self-rated knowledge is significantly correlated with scored knowledge at pretest.	0.0017	$r(36) = .49$
Self-rated knowledge is *not* significantly correlated with GW attitudes at pretest.	0.32	$r(36) = .16$

6.4 Summary and Conclusions

We've shown across a number of populations that ignorance of the basic physical/chemical mechanism of the greenhouse effect is nearly universal (along with the notion that global warming is an extra, anthropogenic effect). In addition to this research, Felipe (2012) describes successes with a curriculum involving the mechanism with 11th graders (and see Chapter 5 for more on the numerical estimation aspects of that study). Across a variety of intervention styles, we have shown that individuals are able to markedly increase their ability to describe the greenhouse effect, and that such an intervention additionally shifts climate-related beliefs and attitudes.

As can be seen from the work described above, the act of educating the American public about the basic mechanism of climate change is a daunting, multi-faceted challenge. Furthermore, even this is not sufficient to know if such an endeavor will truly effect some more direct positive action towards the larger problem of climate change. For that, we will need to examine some connection to behavior. This is the clearest lack in the current research. Such an endeavor will be even more challenging than the above, but it is of critical importance!

Overall, however, we have seen evidence that materials such as those exhibited in Appendix G are likely to be effective both in college lecture rooms as well as online. Evidence from related studies has provided additional support for the effective application of such materials in high school classrooms.

Acknowledgements

The work reported in this chapter has been previously published, in part, in Ranney et al. (2012a, 2012b), and Clark et al. (2013). All such material is re-used here with the permission of my co-authors, the publishers, and the Graduate Division at the University of California, Berkeley.

Chapter 7

Conclusion

We have provided an evidential medley that effectively disconfirms the notion that climate change-relevant knowledge and attitudes are locked in cognitive stasis. Moreover, contrary to those who over-problematize a "knowledge deficit" (or "information deficit") approach to climate change communication, we see a "wisdom deficit." We have demonstrably and considerably un-problematized it with the "cognitive levers" of the interventions described in the previous chapters. In contrast, it is unlikely that offering either an ill-structured list of uncompelling facts to an unprepared mind or thinly veiled rhetoric (cf. Lord et al., 1979) will notably alter beliefs or behaviors—especially about the difficult topic of climate change. Rather, one must be sensitive to specific (mis)understandings that may be relevant to a learner grappling with a domain. Ultimately, we will likely need to engage virtually all people, assisting them in connecting their long-term values to the long-term effects of their behaviors.

In Chapter 4, we also showed, disturbingly, that one can readily erode climate change acceptance with misleading, cherry-picked numbers. We can think of no better protection against such "evil" interventions than to provide the context necessary to recognize them for the clever misinformation that they are. Such prophylactic interventions may represent promising targets for further research and educational initiatives (cf. Lewandowsky et al., 2012).

Our research group is currently studying ways of disseminating the information that we have found to elicit worthwhile cognitive and belief changes. For instance, we are producing on-line instructional materials (e.g., videos available on http://HowGlobalWarmingWorks.org) that can widely convey both global warming's mechanism and the statistics that reflect the scientific consensus of climate change—so the public can join that consensus.

We have shown above that on-line survey interventions, brief curricula, and classroom lessons can have marked and persistent effects on knowledge, understanding, beliefs, and attitudes about climate change. In spite of arguments to the contrary, some simple cognitively-informed interventions might be fundamental in building the resolve to tackle global climate change.

7.1 On the Structure of Successful Interventions

One of the clearer answers in the educational research literature is the superiority (for retention) of "test-enhanced learning" (Roediger & Karpicke, 2006). The basic idea is that one should always try to answer a question rather than simply studying the answer. It should come as no surprise then, that in our own work we've found enhancement in our interventions from eliciting an answer from the participant prior to revealing the correct answer (for example, the enhanced surprise seen in our sandwich group in Section 6.1). Thus, it should not seem controversial that in the experiments above, we have often defaulted to a pretest without a thorough exploration of the effects of leaving one off. If one is interested in developing effective interventions, including a pretest is likely a good starting point!

Regarding the timing of a posttest, while some of our studies in Chapter 6 included a delayed test, they *all* included an immediate posttest. This was more for convenience, and a hedge against rapid forgetting (at least we might observe some immediate effects, even if they were lost later—though fortunately, they were not!). Looking at the work presented in Chapter 2, as well as other literature on optimal timing of practice (e.g., Cepeda et al., 2009), it is likely that an immediate posttest is probably *less* effective than a delayed posttest for long-term retention. Any posttest, however, will likely improve retention. Clearly, determining the optimal timing for a posttest (or several) remains an open question for further research. In practice, however, one is likely to be subject to practical constraints. As such, I would recommend an immediate pretest, as well as a posttest that occurs after as much of a reasonable lag as one can build into a curriculum. Critically, as one moves to different media (such as the video-based materials at http://HowGlobalWarmingWorks.org), the best timecourse for practice and retention may differ markedly.

7.2 A Note on Demographics

While a random sample of workers on Mechanical Turk yielded a relatively Democratic and GW-accepting population, we were still able to capture some conservatives within this sample. As we've noted in the preceding chapters, liberal skew on Mechanical Turk seems to be the norm (Richey & Taylor, 2012). Efforts should be made to evaluate similar science education efforts directed at communities where conservatives are better-represented. There may also have been some self-selection at play in our online studies. But, given the applied nature of this work, self-selection for this experiment likely reflects self-selection for engagement with this material online. This would do little, then, to erode the claim that online science and numeracy interventions can have a net positive impact on individuals. Note also that the "polarization" claim seems highly untenable in the face of results from Chapter 4 (given how UC Berkeley undergraduates are from a liberal pool of Americans).

7.3 On Retrospective Reports of Surprise

It is also interesting to note the different amounts of surprise reported for two numbers in the interventions in Chapter 5 versus the surprise reported for the 400 words in Chapter 6. Specifically, while the 400 words *contains* two of those numbers, individuals report *less* surprise to the textual mechanistic description. An interesting analogy is provided by the account of retrospection on colonoscopies in Kahneman (2003). Here, participants report retrospective pain that is a weighted combination of their peak experience of pain, and the pain at the end of the procedure. Thus, while participants may experience a blip of surprise in the 400 words, the fact that it ends with a summary of the preceding information almost certainly means it ends on an unsurprising note. Thus, participants *may* in fact experience comparable surprise in both styles of intervention, but when surprise is reported only after the unsurprising conclusion of the 400 words, this may drive retrospective surprise down.

Less theoretically interesting, perhaps, but still practically useful, is that we've seen above that different formulations of surprise may be more relevant to different forms of information. In particular, participants appear more willing to say that they were "surprised or embarrassed at their own lack of knowledge" for our mechanistic explanation (as seen in the study in Section 6.3), while they were more willing to report straight surprise at numerical information (as seen in the study in Section 5.2). Use of the appropriate question may allow for better assessment of individual variability in a given intervention. Some care should be exercised, as certain forms of question appear to be rated uniformly high (or low) for a given participant (as seen in Figure 5.1). Such differences might be hashed out in the development of more refined assessments for future studies.

7.4 Summary and Recommendations

Is there value in educating individuals about climate change? In part, it depends on who you are. If you are a science educator, I hardly needed to complete this research to tell you that the answer is an emphatic *yes*. But importantly, by using surprising information that stays close to verifiable facts, you may likely avoid problems with polarization. If you are seeking to influence behavior or policy, however, it is a complex task. But in contrast to the view that climate education is dangerously polarizing, it seems that science education might at least push us in the right direction (even if it, alone, is not the most effective or efficient route to conservation). But in the end, if it turns out that we are indeed educable in a meaningful way, there is perhaps more value in conserving the environment that sustains us than if we were mere automata to be shoved around with propaganda!

At the highest level, the question becomes, what would we have govern us as a society at the highest level: Superstition, the preservation of national power, market forces, or our best shot at objective truth? As we've seen in some of the discussion above, these approaches are intertwined. But I hope I've made it clear that, apart from philosophical arguments in it's favor, science education seems to be an effective approach to tackling the behavioral problem of climate change.

References

Aggleton, J. P. & Brown, M. W. (2006). Interleaving brain systems for episodic and recognition memory. *Trends in Cognitive Sciences*, *10*(10), 455–463. doi:10.1016/j.tics.2006.08.003

Archer, D. & Brovkin, V. (2008). The millennial atmospheric lifetime of anthropogenic CO2. *Climatic Change*, *90*(3), 283–297. doi:10.1007/s10584-008-9413-1

Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 5th ser., *41*(251), 237–276.

Barnosky, A. D. (2009). *Heatstroke: nature in an age of global warming* (2nd Edition). Washington, DC: Island Press.

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425. doi:10.1037/a0021524

Boroditsky, L. (2001). Does language shape thought?: mandarin and english speakers' conceptions of time. *Cognitive psychology*, *43*(1), 1–22.

Carey, S. (2011). *The origin of concepts*. Oxford University Press, USA.

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right-and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice. *Experimental Psychology*, *56*(4), 236–246.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: why some misconceptions are robust. *The Journal of the Learning Sciences*, *14*(2), 161–199.

Clark, D. & Ivry, R. B. (2010). Multiple systems for motor skill learning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 461–467. doi:10.1002/wcs.56

Clark, D. & Ranney, M. A. (2010). Known knowns and unknown knowns: multiple memory routes to improved numerical estimation. In K. Gomez, L. Lyons, & J. Randinsky (Eds.), *Learning in the disciplines: proceedings of the ninth international conference of the learning sciences (ICLS 2010)* (Vol. 1, Full Papers, pp. 460–467). Chicago, IL: International Society of the Learning Sciences, Inc.

Clark, D., Ranney, M. A., & Felipe, J. (2013). Knowledge helps: mechanistic information and numeric evidence as cognitive levers to overcome stasis and build public consensus on climate change. In *Proceedings of the 35th annual meeting of the cognitive science society*. Austin, TX: Cognitive Science Society.

Clark, D. & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia*, *41*(3), 304–317.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi:10.1177/001316446002000104

Cohen, S. (2012). *San Diego's IPCCC: the implications and politics of climate change confusion* (Undergraduate thesis, University of California, Berkeley).

Compayré, G. (1889). *The history of pedagogy* (W. H. Payne, Trans.). Available at: http://archive.org/details/historyofpedagog00compuoft. Boston: D. C. Heath & company.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, *100*(4), 2157–2162. doi:10.1073/pnas.0337195100

Dickson, D. (2005). The case for a 'deficit model' of science communication - SciDev.Net. Retrieved January 19, 2013, from http://www.scidev.net/en/editorials/the-case-for-a-deficit-model-of-science-communic.html

diSessa, A. A. & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191.

Felipe, J. (2012). *Numerical reasoning, knowledge, and environmental behavior regarding climate change* (Masters project paper, University of California, Berkeley).

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd edition). SAGE Publications, Inc.

Garcia de Osuna, J., Ranney, M. A., & Nelson, J. (2004). Qualitative and quantitative effects of surprise: (mis)estimates, rationales, and feedback-induced preference changes while considering abortion. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the twenty-sixth annual conference of the cognitive science society* (pp. 422–427). Mahwah, NJ: Erlbaum.

Harte, J. & Harte, M. E. (2008). *Cool the earth, save the economy: solving the climate crisis is EASY*. Retrieved from http://www.cooltheearth.us/

Hoffman, A. J. (2010, October). Climate change as a cultural and behavioral issue: addressing barriers and implementing solutions. *Organizational Dynamics*, *39*(4), 295–305. doi:10.1016/j.orgdyn.2010.07.005

Hoffman, A. J. (2011). The growing climate divide. *Nature Climate Change*, *1*, 195–196.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346–363.

Jastrow, R., Nierenberg, W., & Seitz, F. (1991). Global warming: what does the science tell us? *Energy*, *16*(11-12), 1331–1345. doi:10.1016/0360-5442(91)90006-8

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. doi:10.1038/nclimate1547

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963–973. doi:10.1111/j.1467-9280.2009.02402.x

Knowlton, B. J. (1998). The relationship between remembering and knowing: a cognitive neuroscience perspective. *Acta psychologica*, *98*(2-3), 253–265.

Krosnick, J. A. & MacInnis, B. (2013). Does the american public support legislation to reduce greenhouse gas emissions? *Daedalus*, *142*(1), 26–39.

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

Leiserowitz, A., Maibach, E., & Roser-Renouf, C. (2010). *Climate change in the american mind: americans' global warming beliefs and attitudes in january 2010*. Yale Project on Climate Change. Yale University. Retrieved from http://environment.yale.edu/uploads/AmericansGlobalWarmingBeliefs2010.pdf

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012, December). Misinformation and its correction continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. doi:10.1177/1529100612451018

Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change*, *3*(4), 399–404. doi:10.1038/nclimate1720

Lomborg, B. (2007). *Cool it: the skeptical environmentalist's guide to global warming*. Random House Digital, Inc.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109.

Lundmark, C. (2007). The new ecological paradigm revisited: anchoring the NEP scale in environmental ethics. *Environmental Education Research*, *13*(3), 329–347. doi:10.1080/13504620701430448

Martinez, C. (2009). *Factors possibly related to evolution acceptance in the U.S.* (Master's thesis, University of California, Berkeley).

Mayhew, P. J., Jenkins, G. B., & Benton, T. G. (2008). A long-term association between global temperature and biodiversity, origination and extinction in the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1630), 47–53. doi:10.1098/rspb.2007.1302

McCright, A. M. & Dunlap, R. E. (2011). The politicization of climate change and polarization in the american public's views of global warming, 2001–2010. *The Sociological Quarterly*, *52*(2), 155–194. Retrieved January 25, 2013, from http://onlinelibrary.wiley.com/doi/10.1111/j.1533-8525.2011.01198.x/full

Montaño, D. E. & Kasprzyk, D. (2008). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education*. Jossey Bass.

Mooney, C. (2011, March). The science of why we don't believe in science. *Mother Jones*. Retrieved from http://www.motherjones.com/politics/2011/03/denial-science-chris-mooney

Munnich, E. L., Ranney, M. A., & Appel, D. M. (2004). Numerically-driven inferencing in instruction: the relatively broad transfer of estimation skills. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the twenty-sixth annual conference of the cognitive science society* (987–992). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., & Bachman, M. L. N. (2005). The longevities of policy-shifts and memories due to single feedback numbers. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual conference of the cognitive science society* (1553–1558). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., Nelson, J., Garcia de Osuna, J., & Brazil, N. (2003). Policy shift through numerically-driven inferencing: an EPIC experiment about when base rates matter. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the twenty-fifth annual conference of the cognitive science society* (834–839). Mahwah, NJ: Erlbaum.

Munnich, E. L., Ranney, M. A., & Song, M. (2007). Surprise, surprise: the role of surprising numerical feedback in belief change. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the twenty-ninth annual conference of the cognitive science society* (503–508). Mahwah, NJ: Erlbaum.

Nadel, L. & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.

Nelson, J. (2007). *Criminal justice beliefs: cognition, emotion and polarization in punishment and rehabilitation decisions* (Doctoral dissertation, University of California, Berkeley).

Oreskes, N. & Conway, E. M. (2010). *Merchants of doubt: how a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing.

Potsdam Institute for Climate Impact Research and Climate Analytics. (2013). *Turn down the heat: climate extremes, regional impacts, and the case for resilience* (tech. rep. No. 78424). World Bank. Washington, DC. Retrieved from https://openknowledge.worldbank.org/handle/10986/14000

R Development Core Team. (2009). *R: a language and environment for statistical computing*. ISBN 3-900051-07-0. Vienna, Austria. Retrieved from http://www.R-project.org

Ranney, M. A. (2012). Why don't americans accept evolution as much as people in peer nations do? a theory (reinforced theistic manifest destiny) and some pertinent evidence. In K. Rosengren, M. Evans, G. Sinatra, & S. Brem (Eds.), *Evolution challenges: integrating research and practice in teaching and learning about evolution* (pp. 233–269). Oxford: Oxford University Press.

Ranney, M. A., Cheng, F., Garcia de Osuna, J., & Nelson, J. (2001). Numerically driven inferencing: a new paradigm for examining judgments, decisions, and policies involving base rates. Paper presented at the annual meeting of the Society for Judgment and Decision Making. Orlando, FL.

Ranney, M. A., Clark, D., Reinholz, D. L., & Cohen, S. (2012a). Changing global warming beliefs with scientific information: knowledge, attitudes, and RTMD (reinforced theistic manifest destiny theory). In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual meeting of the cognitive science society* (pp. 2228–2233). Austin, TX: Cognitive Science Society.

Ranney, M. A., Clark, D., Reinholz, D. L., & Cohen, S. (2012b). Improving americans' modest global warming knowledge in the light of RTMD (reinforced theistic manifest destiny) theory. In K. van Aalst, K. Thompson, M. M. Jacobson, & P. Reimann (Eds.), *The future of learning: proceedings of the tenth international conference of the learning sciences* (Vol. 2, pp. 2-481 to 2-482). International Society of the Learning Sciences, Inc.

Ranney, M. A., Rinne, L. F., Yarnall, L., Munnich, E. L., Miratrix, L., & Schank, P. (2008). Designing and assessing numeracy training for journalists: toward improving quantitative reasoning among media consumers. In P. A. Kirschner, F. Prins, V. Jonker, & G. Kanselaar (Eds.), *International perspectives in the learning sciences: proceedings of the eighth international conference for the learning sciences, vol. 2* (pp. 2246–2253). International Society of the Learning Sciences, Inc.

Ray, J. & Pugliese, A. (2010). *Worldwide, blame for climate change falls on humans*. Gallup. Retrieved June 30, 2013, from http://www.gallup.com/poll/147242/worldwide-blame-climate-change-falls-humans.aspx

Richey, S. & Taylor, B. (2012, December). How representative are amazon mechanical turk workers? Retrieved July 23, 2013, from http://themonkeycage.org/2012/12/19/how-representative-are-amazon-mechanical-turk-workers/

Rinne, L. F., Ranney, M. A., & Lurie, N. H. (2006). Estimation as a catalyst for numeracy: micro-interventions that increase the use of numerical information in decision-making. In S. Barab, K. Hay, & D. Hickey (Eds.), *Proceedings of the seventh international conference on learning sciences* (pp. 571–577). Mahwah, NJ: Erlbaum.

Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., … Wickham, C. (2013). A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, *1*(1). doi:10.4172/gigs.1000101

Sachs, O., Tiefenbeck, V., Duvier, C., Qin, A., Cheney, K., Akers, C., & Roth, K. (2012). *Field evaluation of programmable thermostats*. Fraunhofer Center for Sustainable Energy Systems CSE. Cambridge, MA. Retrieved July 1, 2013, from http://cse.fraunhofer.org/publications/field-evaluation-of-programmable-thermostats/

Schrag, D. P. (2011). Hope and the climate scientist: response to McKibben's human flourishing depends on what we do now. In D. Swearer & S. McGarry (Eds.), *Ecologies of human flourishing* (pp. 169–174). Cambridge, MA: Harvard University Press.

Shepherd, S. & Kay, A. C. (2012). On the perpetuation of ignorance: system dependence, system justification, and the motivated avoidance of sociopolitical information. *Journal of personality and social psychology*, *102*(2), 264.

Shtulman, A. & Calabi, P. (2008). Learning, understanding, and acceptance: the case of evolution. In *Proceedings of the 30th annual conference of the cognitive science society* (235–240). Austin, TX: Cognitive Science Society.

Siegler, R. S. (2000). Unconscious insights. *Current Directions in Psychological Science*, *9*, 79–83.

Sinatra, G., Kardash, C., Taasoobshirazi, G., & Lombardi, D. (2012). Promoting attitude change and expressed willingness to take action toward climate change in college students. *Instructional Science*, *40*(1), 1–17. doi:10.1007/s11251-011-9166-5

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological bulletin*, *119*(1), 3. Retrieved July 23, 2013, from http://psycnet.apa.org/journals/bul/119/1/3/

Straw, A. D. (2008). Vision egg: an open-source library for realtime visual stimulus generation. *Frontiers in neuroinformatics*, *2*. doi:10.3389/neuro.11.004.2008

Thagard, P. (2006). *Hot thought: mechanisms and applications of emotional cognition*. The MIT Press.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007, April). Schemas and memory consolidation. *Science*, *316*(5821), 76–82. doi:10.1126/science.1135935

World Bank. (2012). *Turn down the heat: why a 4°C warmer world must be avoided*. World Bank. Washington, DC. Retrieved from https://openknowledge.worldbank.org/handle/10986/11860

Appendix A

Survey Items Used in Chapters 4, 5, & 6

Table A.1: Demographic questions. Numeric codes assigned by survey software are given in parentheses and not shown to participants. If no list of choices is given for a question, a blank space was provided.

Question ID	Wording / Answers
state	In what U.S. state or territory to you reside? Please use the 2-letter abbreviation (e.g., MD, DC, PR)
gender	What is your gender?
	Male (1), Female (2)
citizen	Are you a U.S. citizen or permanent resident?
	Yes (1), No (2)
us_born	Were you born in the U.S.?
	Yes (1), No (2)
english	Is English your first language?
	Yes (1), No (2)
party	What is your strongest political party affiliation?
	None (1), Democrat (2), Green (3), Independent (4), Libertarian (5), Republican (6), Other (7), Decline to State (8)

Table A.1: Demographic questions, continued

Question ID	Wording / Answers
conserv	On the following scale, indicate the extent to which you consider yourself to be liberal or conservative on most political and social issues:
	1 Extremely Liberal, 2, 3 Somewhat Liberal, 4, 5 Moderate, 6, 7 Somewhat Conservative, 8, 9 Extremely Conservative (numerical scale given to participants)
faith	What is your main religious faith?
	Atheist (1), Agnostic (2), Buddhist (3), Christian (4), Hindu (5), Jewish (6), Muslim (7), Spiritual but not religious (8), Other (9), Decline to state (10)[1]
ed_level	What is the highest level of education you have completed?
	No higher than 8th grade (1), Some high school (9-12th grade) (2), High school diploma / GED (3), Some college (4), Bachelor's degree (5), Master's degree (6), Professional degree (7), Doctoral degree (8)
cc_inst	Have you received any instruction regarding Global Warming (Climate Change) in the last 2 years? If so, when was the most recent?
	No (1), Fall 2012 (2), Summer 2012 (3), Spring 2012 (4), Winter 2011-2012 (5), Fall 2011 (6), Summer 2011 (7), Spring 2011 (8), Winter 2010-2011 (9)
cci_desc	Please describe the instruction you've received regarding Global Warming (Climate Change)
us_years	You indicated that you were born outside the U.S. How many years have you been living in the U.S.? (Please round up to the nearest whole number.)
where_born	Where were you born?
eng_years	How many years have you been speaking English?
first_lang	What is your first language?

[1]Note that the RPP demographics survey further splits Christian into Protestant and Catholic

Table A.2: Belief / attitude ("RTMD") questions used in the various studies. Most items were on a 1–9 scale (noted below the table). Scales for items that deviate are reported below the question.

Question ID	Wording / Answers
evo1_1	Evolution accurately explains how plants, animals, and humans came to be as they are.
gw1_2	Human activities are largely responsible for the climate change (global warming) that is going on now.
nat1_3	The United States is one of the very best countries on our planet (e.g., "in the top three").
dty1_4	There exists a supernatural being/deity (e.g., God) or set of beings/deities (gods).
aft1_5	After a person dies, that person experiences an afterlife of some sort (for instance, heaven/hell, reincarnation, enlightenment, nirvana, etc.).
cre1_6	Biblical creation accurately explains how plants, animals, and humans came to be as they are.
gw2_1	Global warming or climate changes, when they happen at all, are just parts of a natural cycle.
gw2_2	I am certain that global warming is actually occurring.
gw2_3	I am worried about global warming.
gw2_4	Humans are severely abusing the environment.
evo2_5	Evolution is unable to explain much of the physical evidence regarding the origins and development of life on Earth.
evo2_6	Other living things may have evolved, but humans have not.
lifsty	Overall, how important is it to change your current lifestyle to reduce your carbon footprint (i.e., to decrease the amount of greenhouse gases you emit both directly and indirectly)? Not Important 1, 2, Slightly Important 3, 4, Somewhat Important 5, 6, Very Important 7, 8, Extremely Important 9

1–9 scale usually consisted of: Extremely Disagree 1, Strongly Disagree 2, Disagree 3, Mildly Disagree 4, Neither Agree Nor Disagree 5, Mildly Agree 6, Agree 7, Strongly Agree 8, Extremely Agree 9. Where different, the scale is noted in the table next to the question above.

Table A.2: Belief / attitude ("RTMD") questions.

Question ID	Wording / Answers
engage	I intend to personally engage in more environmentally-friendly (e.g., sustainable, recycling, and/or resource-minimizing) activities in the future, compared to what I do now.
aft2	After a person dies, that person lives on in some way.
aft3	I don't believe that heaven exists.
dty2	God is created by human imagination.
dty3	The only true God is that of my religion.
cre2	A supreme being has never played any role in the origin or development of life on earth.
cre3	To me, creation gives a more satisfying explanation of life on Earth than does evolution.
nat2	The United States can fix just about any problem it might unintentionally create.
natmil	How many countries could defeat the U.S. militarily without assistance from other countries? ______________ Countries
knwgbl	Please indicate how knowledgeable you think you are about climate change—by choosing a number on the 1 (not knowledgeable at all) to 9 (extremely knowledgeable) scale below. Not knowledgeable at all about Climate Change 1, 2, 3, 4, Moderately knowledgeable about Climate Change 5, 6, 7, 8, Extremely knowledgeable about Climate Change 9

1–9 scale usually consisted of: Extremely Disagree 1, Strongly Disagree 2, Disagree 3, Mildly Disagree 4, Neither Agree Nor Disagree 5, Mildly Agree 6, Agree 7, Strongly Agree 8, Extremely Agree 9. Where different, the scale is noted in the table next to the question above.

Appendix B

UNDP Millenium Goals and Climate-Related Funding Choices

Chapter 4 describes a series of fund allocation policy decisions made by participants. Below are the instructions given to participants in our 2-item intervention, followed by the text used to describe the two alternatives for each item.

B.1 Funding Policy Instructions

As a result of the UN Millennium Summit, in the year 2000, the United Nations adopted eight goals for increasing the economic and social conditions of the world's poorest countries, called the Millennium Development Goals. These goals are to: (1) end poverty and hunger, (2) achieve universal primary education, (3) promote gender equity, (4) reduce child mortality rates, (5) improve maternal health, (6) combat HIV/AIDS and other diseases, (7) ensure environmental sustainability, and (8) develop a global partnership for development.

Imagine that you have been hired as a consultant to the United Nations. Your task is to allocate funds between projects oriented toward global climate change and projects focused on achieving other Millennium Development Goals. You will provide from two to four policy allocations in total.

For each policy, first you will be asked to estimate the value of a policy-relevant statistic. Then you will make an initial policy recommendation. You will be asked to describe your estimation process—in particular, what knowledge and reasoning you used to make your estimate. (*You will write all of this information inside of this packet.*)

After making an initial recommendation for each of the Millennium Development Goals, you will put this packet away, and begin on Packet 2 (the yellow packet). At that time, you will be given the true values of the statistics, as well as an opportunity to revise each recommendation you made.

B.2 Funding Alternatives

All 4 variants of the 2-item intervention used the same policy choices. The first (policy one) was:

1. Create initiatives to reduce extreme poverty and hunger; or
2. Invest in new technologies to reduce the levels of greenhouse gases in the atmosphere.

The second (policy two) was:

1. Invest in providing sustainable access to safe drinking water and basic sanitation; or
2. Invest in renewable energy technologies, such as solar and wind power.

Appendix C

Format of "Evil" NDI Intervention

Below are the full instructions and examples only for our "evil" NDI intervention. Interventions consisted of two interventions of the sort illustrated in the following pages, or eight drastically simplified estimations. Note that the following pages only include the "example" pages. The pages immediately following each example followed the same format, but included numbers that demonstrably eroded acceptance and concern for global warming (as described in Chapter 4).

Packet 1 (Initial Allocations)

Please read now: General Instructions

Closely related to today's lecture, you are asked to take part in an informative 22-minute study. Thank you for your participation!

Associated with this survey is a consent form. If you will, please read it and sign it now. We will collect it soon, and you will be offered a copy of it later.

Once we begin, you may also ask a question at any time. (Pilot-testing suggests that the survey is rather clear, but one never knows!)

The statistics that you will eventually be given are true and accurate. These statistics really do come from the sources that are cited; for instance, you could truly share the information with your family members tonight.

Please *don't* look at your neighbors' surveys or speak to them during this study. Also, please don't skip ahead and don't go back to an earlier page.

For items that use a 1-9 scale, please respond to them by indicating the degree appropriate—for instance, by circling a number on the 1 to 9 scales below (1 for the least/lowest and 9 for the most/highest).

Please answer honestly regarding your true thoughts and beliefs. We underlined words that might be easy to misread like "not" and "don't," but please be sure to read each item carefully.

We have a limited time to administer this survey, so please answer the short-answer items with some brevity.

Again, your participation is sincerely appreciated—and it is for a good cause. You will receive feedback regarding what this research is for during the lecture, and you can ask anything you wish at that time.

Do you have any questions?

Thanks again!

Please indicate the degree to which you are knowledgeable about global climate change—by circling a number on the 1 (not knowledgeable at all) to 9 (extremely knowledgeable) scale below.

1	2	3	4	5	6	7	8	9
Not knowledge-able at all about Climate Change				Moderately knowledge-able about Climate Change				Extremely knowledge-able about Climate Change

Please go on to the next page. At that point, please do not return to this page.

Please respond to the following items, if you will, by indicating the degree to which you agree with each statement—by circling a number on the 1 (extremely disagree) to 9 (extremely agree) scale below.

Evolution accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Human activities are largely responsible for the climate change (global warming) that is going on now.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

The United States is one of the very best countries on our planet (e.g., "in the top three").

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

There exists a supernatural being/deity (e.g., God) or set of beings/deities (gods).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

After a person dies, that person experiences an afterlife of some sort (for instance, heaven/hell, reincarnation, enlightenment, nirvana, etc.).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Biblical creation accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please go on to the next page. At that point, please do not return to this page.

Global warming or climate changes, when they happen at all, are just parts of a natural cycle.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am certain that global warming is actually occurring.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am worried about global warming.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Humans are severely abusing the environment.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Evolution is <u>un</u>able to explain much of the physical evidence regarding the origins and development of life on Earth.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Other living things may have evolved, but humans have <u>not</u>.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please note the change in wording of the following scale

Overall, how important is it to change your current lifestyle to reduce your carbon footprint (i.e., to decrease the amount of greenhouse gases you emit both directly and indirectly)?

1	2	3	4	5	6	7	8	9
Not Important		Slightly Important		Somewhat Important		Very Important		Extremely Important

I expect to personally engage in more environmentally-friendly (e.g., sustainable, recycling, and/or resource-minimizing) activities in the future, compared to what I do now.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please go on to the next page. At that point, <u>please do not return to this page</u>.

As a result of the UN Millennium Summit, in the year 2000, the United Nations adopted eight goals for increasing the economic and social conditions of the world's poorest countries, called the Millennium Development Goals. These goals are to: (1) end poverty and hunger, (2) achieve universal primary education, (3) promote gender equity, (4) reduce child mortality rates, (5) improve maternal health, (6) combat HIV/AIDS and other diseases, (7) ensure environmental sustainability, and (8) develop a global partnership for development.

Imagine that you have been hired as a consultant to the United Nations. Your task is to allocate funds between projects oriented toward global climate change and projects focused on achieving other Millennium Development Goals. You will provide from two to four policy allocations in total.

For each policy, first you will be asked to estimate the value of a policy-relevant statistic. Then you will make an initial policy recommendation. You will be asked to describe your estimation process—in particular, what knowledge and reasoning you used to make your estimate. (*You will write all of this information inside of this packet.*)

After making an initial recommendation for each of the Millennium Development Goals, you will put this packet away, and begin on Packet 2 (the yellow packet). At that time, you will be given the true values of the statistics, as well as an opportunity to revise each recommendation you made.

Please *do not* go back to review or revise any previous page (or packet) once you have moved on.

On the following pages you will be asked to estimate some statistics. (After making all of your initial estimates, you will later be given the true values of the statistics, so try to make a mental note of your estimates.) An example of how you should fill out these pages is given below.

EXAMPLE:

The United Nations has 1 billion dollars to allocate between two programs. Please indicate what percentage of the 1 billion dollars you would allocate to each program (for a total of 100% between the two programs). Before providing any policy recommendations, please provide an estimate for the following statistic:

According to a 2005 survey by the Center for Survey and Research Analysis at the University of Connecticut, (estimate) __X__% of Americans believe in extraterrestrial life.

What % of Americans do you think believe in extraterrestrial life? That is, what is X?
Write your estimate here: ___40_____%

In a sentence or two: *Describe how you came up with that estimate.*

"About 60% of my friends believe in extraterrestrial life, but I think we know a little bit more than the average American does, so the actual number is probably lower."

How *low* would the actual number have to be to surprise you? (low) ________20_________ %

How *high* would the actual number have to be to surprise you? (high) _______60_________ %

How confident are you that the actual value falls in the range between (low) and (high) above? (circle one)

95%	85%	75%	65%	55%	_______%
almost totally sure	quite sure	moderately sure	mildly sure	almost unsure	other (write-in)

Now, what percent of the 1 billion dollars would you allocate between the following two policies?:

(#1) Invest in providing sustainable access to safe drinking water and basic sanitation; or

(#2) Invest in defense technologies to prevent an extraterrestrial invasion.

__45__% toward policy **(#1)** ___55___% toward policy **(#2)**

In a sentence or two: Why would you allocate funds that way?

"Having safe drinking water is really important, but if aliens take over the world it doesn't really matter anymore."

Now that you have read this example, please fill out the following pages with your estimates and policy recommendations.

Thank you for providing policy recommendations for how the United Nations should allocate its funds toward achieving the UN Millennium Development Goals. ***At this time, it is perfectly okay to go back to look at (but <u>not</u> change) your prior answers, so please take a moment to review your answers, just to make sure that you wrote all of your estimates in this packet.*** *Also, please check if you remembered to indicate "increase" or "decrease" if/where appropriate. If you forgot to write any answers, please do so now.*

Now, please put this packet inside of your envelope and take out Packet 2 (the yellow packet).

Packet 2 (Final Allocations)

Note: This packet is for the second part of today's survey. If you have not completed Packet 1, please return this yellow packet to your envelope and first complete Packet 1. *Please do not turn to the next page of this packet unless you have already completed Packet 1 (the white packet).* If you have already completed Packet 1, please make sure it is put away in your envelope before continuing.

A Reminder: General Instructions

Once we begin, you may also ask a question at any time. (Pilot-testing suggests that the survey is rather clear, but one never knows!)

The statistics that you will eventually be given are true and accurate. These statistics really do come from the sources that are cited; for instance, you could truly share the information with your family members tonight.

Please *don't* look at your neighbors' surveys or speak to them during this study. Also, please don't skip ahead and don't go back to an earlier page.

For items that use a 1-9 scale, please respond to them by indicating the degree appropriate—for instance, by circling a number on the 1 to 9 scales below (1 for the least/lowest and 9 for the most/highest).

Please answer honestly regarding your true thoughts and beliefs. We underlined words that might be easy to misread like "not" and "don't," but please be sure to read each item carefully.

We have a limited time to administer this survey, so please answer the short-answer items with some brevity.

Again, your participation is sincerely appreciated—and it is for a good cause. You will receive feedback regarding what this research is for during the lecture, and you can ask anything you wish at that time.

Further Instructions:

If you have Packet 1 (the white packet) in front of you now, please put it inside of your envelope before you begin. Right now you should have Packet 2 (the yellow packet) in front of you.

Do you have any questions? Thanks again!

On the following pages you will be given the true values of the statistics you estimated, and asked to revise your policy recommendations. An example of how you should fill out these pages is given below.

EXAMPLE Allocation revisited:

Previously, you provided an estimate before you completed EXAMPLE Allocation. Please think about the estimate you gave before reading on.

Your advisor now informs you of the true value:

According to a 2005 survey by the Center for Survey and Research Analysis at the University of Connecticut, ***60%*** *of Americans believe in extraterrestrial life.*

Note how this number compares to (your memory of) the estimate you gave.

Did you find this number surprising? Please rate according to the following scale:

1	2	3	4	5	6	7	8	9
Not Surprising At all				Somewhat Surprising				Extremely Surprising

Briefly, what specifically did you find surprising (if anything)?

"I didn't realize so many Americans believed in extraterrestrial life!"

Given this information, what percent of the 1 billion dollars would you allocate between these two policies?:

(#1) Invest in providing sustainable access to safe drinking water and basic sanitation; or

(#2) Invest in defense technologies to prevent an extraterrestrial invasion.

__*25*__% toward policy **(#1)**　　　　__*75*__% toward policy **(#2)**

In a sentence or two: Why would you allocate funds that way?

"It's very likely that there are aliens out there, and we need to be ready for them!"

Now that you have read this example, please fill out the following pages with your revised policy recommendations.

Please indicate the degree to which you are knowledgeable about climate change—by circling a number on the 1 (not knowledgeable at all) to 9 (extremely knowledgeable) scale below.

1	2	3	4	5	6	7	8	9
Not knowledge-able at all about Climate Change				Moderately knowledge-able about Climate Change				Extremely knowledge-able about Climate Change

Please go on to the next page. At that point, please do not return to this page.

Evolution accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Human activities are largely responsible for the climate change (global warming) that is going on now.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

The United States is one of the very best countries on our planet (e.g., "in the top three").

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

There exists a supernatural being/deity (e.g., God) or set of beings/deities (gods).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

After a person dies, that person experiences an afterlife of some sort (for instance, heaven/hell, reincarnation, enlightenment, nirvana, etc.).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Biblical creation accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please go on to the next page. At that point, please do not return to this page.

Global warming or climate changes, when they happen at all, are just parts of a natural cycle.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am certain that global warming is actually occurring.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am worried about global warming.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Humans are severely abusing the environment.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Evolution is unable to explain much of the physical evidence regarding the origins and development of life on Earth.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Other living things may have evolved, but humans have not.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please note the change in wording of the following scale

Overall, how important is it to change your current lifestyle to reduce your carbon footprint (i.e., to decrease the amount of greenhouse gases you emit both directly and indirectly)?

1	2	3	4	5	6	7	8	9
Not Important		Slightly Important		Somewhat Important		Very Important		Extremely Important

I expect to personally engage in more environmentally-friendly (e.g., sustainable, recycling, and/or resource-minimizing) activities in the future, compared to what I do now.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please go on to the next page. At that point, please do not return to this page.

Please circle, as appropriate, regarding your background

What is your gender? M/F

Are you a U.S. citizen or permanent resident? Y/N

Were you born in the US? Y/N

If not, how many years have you been living in the U.S? __________

Is English your first language? Y/N

What is your strongest political party affiliation?
1. None
2. Democrat
3. Green
4. Independent
5. Libertarian
6. Republican
7. Other
8. Decline to state

On the following scale, indicate the extent to which you consider yourself to be liberal or conservative on most political and social issues:

1	2	3	4	5	6	7	8	9
Extremely Liberal		Somewhat Liberal		Moderate		Somewhat Conserv-ative		Extremely Conserv-ative

What is your main religious faith?
1. Atheist
2. Agnostic
3. Buddhist
4. Christian
5. Hindu
6. Jewish
7. Muslim
8. Spiritual but not religious
9. Other
10. Decline to state

In general, did you accept the validity and relevance of the numerical facts? Explain if necessary.

Finally, did you think that you were supposed to change your mind during the survey? Please explain.

Thank you. When finished, please put both packets inside your envelope.

Appendix D

Format of Representative-Number NDI Intervention

Below is an example format for an NDI-style estimation intervention. Unlike in the “evil” NDI intervention, all participants encountered seven or eight such estimations (described in Appendix E). Formatting of the survey items for the web was a straightforward: the statement for each survey item (provided in Apendix A) was provided with a set of radio buttons directly underneath to allow selection of a response ranging from “1” (extreme disagreement) to “9” (extreme agreement). Note that I was unable to capture the actual formatting of the web site used via the Qualtrics system. The general layout is correct, however the color scheme was UC Berkeley themed. One detail that is lost occurs where participants are asked to provide percent confidence: on the web version, there is a slider for that response.

Please estimate the following quantity. First, you must select the choice that corresponds to the format for your estimate - even if there is only one choice! Then, please write the number *by itself*.

According to observation data collected at Mauna Loa Observatory in Hawaii, what is the percent change in atmospheric CO2 levels from 1959 (when observation began) to 2009?

My estimate is a...

○ % increase (in CO2) ○ % decrease (in CO2)

The number I estimate is (whole numbers only):

How low and how high would that number have to be to surprise you? If you wrote decrease above, and one of your answers is an increase (or vice versa) use the minus sign (-) to indicate the opposite direction, and put that number in "Below." Your "Above" answer should *always* be the same direction as your estimate (and probably larger)!

Below:

Above:

Approximately how confident are you that the actual value falls in the range between the *low* and *high* numbers above? (Note that a confidence below 50 would imply you think it is *more likely* that the number is *outside* your range!)

	Unsure				Moderately Sure				Totally Sure		
	50	55	60	65	70	75	80	85	90	95	100
Percent Confidence											

Appendix E

Numerical Information Used in Chapter 5

Tables E.1 and E.2 provide the wording of the numerical information used in Studies 1 and 2 in Chapter 5. These were not the only changes made, and for a more complete description of the changes, please see section 5.2.1.

Note that we do not report the specific information used in Chapter 4, as we have little desire to contribute to the precision of the already formidable propaganda campaign arrayed against climate change acceptance. However, below are the numerical items used in Chapter 5.

Acknowledgements

Thanks to Daniel Reinholz and his team of undergraduates from the Reasoning Group for their fine work in developing this list of numbers.

Table E.1: Representative numerical information used in the study in Section 5.2.

Code	Textual Description	Format / Value
hy	According to the IPCC Fourth Assessment Report (released in 2007), what is the number of years between 1995-2006 (a 12 year period) that rank among the hottest 12 years for average global temperature?	"# of years" / **11 years**
am	What is the change in the atmospheric levels of methane (a greenhouse gas) since 1750?	"% increase" or "% decrease" / **151% increase**
oi	What is the change in percentage of the world's ocean ice cover since the 1960s?	"% increase" or "% decrease" / **40% decrease**
ac	According to observation data collected at Mauna Loa Observatory in Hawaii, what is the percent change in atmospheric CO2 levels from 1959 (when observation began) to 2009?	"% increase" or "% decrease" / **22.6% increase**
rc	Based on a set of 1372 climate researchers, what is the percentage of researchers most actively publishing in the field of climate research who support the tenets of anthropogenic climate change, as outlined by the IPCC?	"% of researchers" / **97.5%**
pg	In 1850 there were approximately 150 glaciers present in Glacier National Park. How many are present today?	"# of glaciers" / **25 glaciers**
ag	From 1850 to 2004, what is the percent change of volume of glaciers in the European Alps?	"% increase" or "% decrease" / **50% decrease**
sl	According to a study published in Geophysical Research Letters, by how much has the average sea level changed from 1870 to 2004? *Note: actually reported 6.4 feet on all but final Mechanical Turk studies, where this item was not used because it is actualy not very impressive.*	"feet increase" or "feet decrease" / **0.64 feet**

Table E.2: Final version of representative numerical information used in the study in Section 5.2 (Study 2). Note that not all items are changed from Table E.1, and only seven items are used here (as opposed to eight in previous interventions).

Code	Textual Description	Format / Value
hy	Global surface temperatures have been recorded since 1850. According to the 2007 report from the Intergovernmental Panel on Climate Change, how many of the years between 1995-2006 (a 12 year period) are one of the hottest 12 years recorded?	"# of years" / **11 years**
am	What is the change in the atmospheric levels of methane (a greenhouse gas) since 1750?	"% increase" or "% decrease" / **151% increase**
oi	What is the change in percentage of the world's ocean ice cover since the 1960s?	"% increase" or "% decrease" / **40% decrease**
ac	According to observation data collected at Mauna Loa Observatory in Hawaii, what is the percent change in atmospheric CO2 levels from 1959 (when observation began) to 2009?	"% increase" or "% decrease" / **22.6% increase**
rc	A 2010 article examines the 908 active researchers with at least 20 climate publications on Google Scholar. What percentage of them have stated that it is "very likely" that human-caused emissions are responsible for "most" of the "unequivocal" warming of the Earth in the second half of the 20th century?	"% of researchers" / **97.5%**
pg	In 1850 there were approximately 150 glaciers present in Glacier National Park. How many are present today?	"# of glaciers" / **25 glaciers**
ag	From 1850 to 2004, what is the percent change of volume of glaciers in the European Alps?	"% increase" or "% decrease" / **50% decrease**

Appendix F

Pencil-and-Paper Version of the Mechanism Intervention

On the pages that follow is a faithful reproduction of the core intervention given to individuals in the no-pretest condition. The sandwich intervention included an exact copy of the posttest as the pretest, except that it did not include the “kinds of light” comprehension check. Both versions included page numbers (omitted here to avoid confusion with dissertation page numbers). Finally, the no-pretest survey included a brief set of open-ended questions that are not reported on in this dissertation.

The online version was quite similar, with largely identical instructions (the primary difference being the addition of some provisions for quitting the experiment by closing the browser). In addition, online the survey items were randomized, and a few more were added.

<u>Please read now</u>: General Instructions

Intimately related to today's lecture, you are asked to take part in an informative 15-minute study. Thank you for your participation! We believe that you will find this interesting, and we hope that it will also result in some good for society.

<u>The survey looks longer than it is</u>. Some pages have only one item on them.

Associated with this survey is a consent form. If you will, please read it and sign it now. We will collect it soon, and you will be offered a copy of it later.

Once we begin, you may also ask a question at any time. (Pilot-testing suggests that the survey is rather clear, but one never knows!)

This study involves <u>NO deceptions</u>. There is NO "trick" involved, and what we are asking about is what we are actually interested in. Further, any information that we provide you is accurate; for instance, you can share the information with your family tonight, if you wish.

Please *don't* look at your neighbors' surveys. We are using multiple versions, and it will confuse you/us if you have straying eyes. Also, please don't skip ahead and <u>don't go back to an earlier page</u>.

For items that use a 1-9 scale, please respond to them by indicating the degree appropriate—for instance, by circling a number on the 1 to 9 scales below (1 for the least/lowest and <u>9 for the most/highest</u>).

Please answer honestly regarding your true thoughts and beliefs. We underlined words that might be easy to misread like "not" and "don't," but please be sure to read each item carefully.

We have a limited time to administer this survey, so please <u>answer the short-answer items with some brevity</u>. Note that <u>some items only ask you if you would "add anything"</u> to what you wrote on a page that is only 1-2 pages back. On these items, there is no need to repeat what you wrote those 1-2 pages back. Add what you will, and if you have nothing to add, simply indicate that and move onto the next item.

Again, your participation is sincerely appreciated—and for a good cause. You will receive feedback regarding what this research is for during the lecture, and you can ask anything you wish at that time.

Do you have any questions?

Thanks again!

Important! Please read and understand this page.

How does climate change ("global warming") work? The mechanism of the greenhouse effect

[Or: "Why do some gases concern scientists—like carbon dioxide (CO_2)—but not others, like oxygen?"]

Scientists tell us that human activities are changing Earth's atmosphere and increasing Earth's average temperature. What causes these climate changes?

First, let's understand Earth's "normal" temperature: When Earth absorbs sunlight, which is mostly visible light, it heats up. Like the sun, Earth emits energy—but because it is cooler than the sun, Earth emits lower-energy infrared wavelengths. Greenhouse gases in the atmosphere (methane, carbon dioxide, etc.) let visible light pass through, but absorb infrared light—causing the atmosphere to heat up. The warmer atmosphere emits more infrared light, which tends to be re-absorbed—perhaps many times—before the energy eventually returns to space. The extra time this energy hangs around has helped keep Earth warm enough to support life as we know it. (In contrast, the moon has no atmosphere, and it is colder than Earth, on average.)

Since the industrial age began around the year 1750, atmospheric carbon dioxide has increased by 40% and methane has increased by 150%. Such increases cause *extra* infrared light absorption, further heating Earth above its typical temperature range (even as energy from the sun stays basically the same). In other words, energy that gets to Earth has an even *harder* time leaving it, causing Earth's average temperature to increase—producing global climate change.

[In molecular detail, greenhouse gases absorb infrared light because their molecules can vibrate to produce asymmetric distributions of electric charge, which match the energy levels of various infrared wavelengths. In contrast, non-greenhouse gases (such as oxygen and nitrogen—that is, O_2 and N_2) don't absorb infrared light, because they have symmetric charge distributions even when vibrating.]

Summary: (a) Earth absorbs most of the sunlight it receives; (b) Earth then emits the absorbed light's energy as infrared light; (c) greenhouse gases absorb a lot of the infrared light before it can leave our atmosphere; (d) being absorbed slows the rate at which energy escapes to space; and (e) the slower passage of energy heats up the atmosphere, water, and ground. By increasing the amount of greenhouse gases in the atmosphere, humans are increasing the atmosphere's absorption of infrared light, thereby warming Earth and disrupting global climate patterns.

Shorter summary: Earth transforms sunlight's visible light energy into infrared light energy, which leaves Earth slowly because it is absorbed by greenhouse gases. When people produce greenhouse gases, energy leaves Earth even more slowly—raising Earth's temperature.

Please go on to the next page. At that point, please do not return to this page.

Did you find anything in this explanation surprising? Please rate according to the following scale:

1	2	3	4	5	6	7	8	9
Not Surprising At all				Somewhat Surprising				Extremely Surprising

Briefly, what specifically did you find surprising (if anything)?

Please go on to the next page. At that point, please do not return to this page.

Please respond to the following items, if you will, with a brief textual answer. Questions are on separate pages to prevent backtracking, and it is expected that you will leave a large amount of empty space on these pages.

Please write 1-3 sentences (about **30** words or less) that you could use to explain how climate change occurs to a senior in high school:

Please go on to the next page. At that point, please do not return to this page.

On the previous page, you responded to the following request:

“Please write 1-3 sentences (about 30 words or less) that you could use to explain how climate change occurs to a senior in high school.”

*Briefly (**25** words or less), what would you add, if anything, in response to the following?*

Please explain any differences regarding how energy (i.e., heat, light) travels to the Earth from the sun compared to how energy travels away from the Earth:

Please go on to the next page. At that point, please do not return to this page.

On the previous pages, you responded to the following requests:

1) "Please write 1-3 sentences (about 30 words or less) that you could use to explain how climate change occurs to a senior in high school."

2) "Please explain any differences regarding how energy (i.e., heat, light) travels to the Earth from the sun compared to how energy travels away from the Earth."

*Briefly (**25** words or less), what would you add, if anything, in response to the following questions?:*

Are all gases "greenhouse gases?" If not, what makes something a greenhouse gas?

Please go on to the next page. At that point, please do not return to this page.

The sun mostly emits ____________ light towards the Earth.

The Earth mostly emits ____________ light out into space.

Please go on to the next page. At that point, please do not return to this page.

Please indicate the degree to which you are knowledgeable about climate change—by circling a number on the 1 (not knowledgeable at all) to 9 (extremely knowledgeable) scale below.

1	2	3	4	5	6	7	8	9
Not knowledge-able at all about Climate Change				Moderately knowledge-able about Climate Change				Extremely knowledge-able about Climate Change

Please go on to the next page. At that point, <u>please do not return to this page</u>.

Evolution accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Human activities are largely responsible for the climate change (global warming) that is going on now.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

The United States is one of the very best countries on our planet (e.g., “in the top three”).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

There exists a supernatural being/deity (e.g., God) or set of beings/deities (gods).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

After a person dies, that person experiences an afterlife of some sort (for instance, heaven/hell, reincarnation, enlightenment, nirvana, etc.).

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Biblical creation accurately explains how plants, animals, and humans came to be as they are.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please go on to the next page. At that point, <u>please do not return to this page</u>.

Global warming or climate changes, when they happen at all, are just parts of a natural cycle.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am certain that global warming is actually occurring.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

I am worried about global warming.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Humans are severely abusing the environment.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Evolution is unable to explain much of the physical evidence regarding the origins and development of life on Earth.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Other living things may have evolved, but humans have not.

1	2	3	4	5	6	7	8	9
Extremely Disagree	Strongly Disagree	Disagree	Mildly Disagree	Neither Agree Nor Disagree	Mildly Agree	Agree	Strongly Agree	Extremely Agree

Please note the change in wording of the following scale

Overall, how important is it to change your current lifestyle to reduce your carbon footprint (i.e., to decrease the amount of greenhouse gases you emit both directly and indirectly)?

1	2	3	4	5	6	7	8	9
Not Important		Slightly Important		Somewhat Important		Very Important		Extremely Important

Please go on to the next page. At that point, please do not return to this page.

Please circle, as appropriate, regarding your background

What is your gender? M/F

Are you a U.S. citizen or permanent resident? Y/N

Were you born in the US? Y/N

If not, how many years have you been living in the U.S? __________

Is English your first language? Y/N

What is your strongest political party affiliation?

1. None
2. Democrat
3. Green
4. Independent
5. Libertarian
6. Republican
7. Other
8. Decline to state

On the following scale, indicate the extent to which you consider yourself to be liberal or conservative on most political and social issues:

1	2	3	4	5	6	7	8	9
Extremely Liberal		Somewhat Liberal		Moderate		Somewhat Conserv-ative		Extremely Conserv-ative

What is your main religious faith?

1. Atheist
2. Agnostic
3. Buddhist
4. Christian
5. Hindu
6. Jewish
7. Muslim
8. Spiritual but not religious
9. Other
10. Decline to state

Please go on to the next page. At that point, please do not return to this page.

Appendix G

The 400 Words

How does climate change ("global warming") work? The mechanism of the greenhouse effect

[Or: "Why do some gases concern scientists—like carbon dioxide (CO_2)—but not others, like oxygen?"]

Scientists tell us that human activities are changing Earth's atmosphere and increasing Earth's average temperature. What causes these climate changes?

First, let's understand Earth's "normal" temperature: When Earth absorbs sunlight, which is mostly visible light, it heats up. Like the sun, Earth emits energy—but because it is cooler than the sun, Earth emits lower-energy infrared wavelengths. Greenhouse gases in the atmosphere (methane, carbon dioxide, etc.) let visible light pass through, but absorb infrared light—causing the atmosphere to heat up. The warmer atmosphere emits more infrared light, which tends to be re-absorbed—perhaps many times—before the energy eventually returns to space. The extra time this energy hangs around has helped keep Earth warm enough to support life as we know it. (In contrast, the moon has no atmosphere, and it is colder than Earth, on average.)

Since the industrial age began around the year 1750, atmospheric carbon dioxide has increased by 40% and methane has increased by 150%. Such increases cause *extra* infrared light absorption, further heating Earth above its typical temperature range (even as energy from the sun stays basically the same). In other words, energy that gets to Earth has an even harder time leaving it, causing Earth's average temperature to increase—producing global climate change.

[In molecular detail, greenhouse gases absorb infrared light because their molecules can vibrate to produce asymmetric distributions of electric charge, which match the energy levels of various infrared wavelengths. In contrast, non-greenhouse gases (such as oxygen and nitrogen—that is, O_2 and N_2) don't absorb infrared light, because they have symmetric charge distributions even when vibrating.]

Summary: (a) Earth absorbs most of the sunlight it receives; (b) Earth then emits the absorbed light's energy as infrared light; (c) greenhouse gases absorb a lot of the infrared light before it can leave our atmosphere; (d) being absorbed slows the rate at which energy escapes to space; and (e) the slower passage of energy heats up the atmosphere, water, and ground. By increasing the amount of greenhouse gases in the atmosphere, humans are increasing the atmosphere's absorption of infrared light, thereby warming Earth and disrupting global climate patterns.

Shorter summary: Earth transforms sunlight's visible light energy into infrared light energy, which leaves Earth slowly because it is absorbed by greenhouse gases. When people produce greenhouse gases, energy leaves Earth even more slowly—raising Earth's temperature.

Acknowledgements

Thanks to Michael Ranney, Daniel Reinholz, and Lloyd Goldwasser—the principal authors of the above explanation.

Appendix H

Detecting Plagiarism using Google Custom Search Engine

First, one needs to set up a custom search engine at http://www.google.com/cse. A small number of queries are provided for free each day, beyond which the service requires a small payment per 1,000 queries. It's not obvious from the documentation, but to use the service one needs to create a custom engine using some (any) URL. The logic is that this service is for a "custom" search focused on your particular web properties. We, however, simply wish to get generic google results for our texts. Thus, after creating our custom-tailored search engine, in the settings, one can enable searching the entire web, and then remove the initial URL. Now, one has the ability to search using Google's Custom Search API using one's developer key and custom search engine ID. The results are probably identical with what a regular google search would provide, but there is no guarantee from Google that this is the case.

Google's API will accept a maximum of 32 search terms, and as such, I only used the first 32 terms from each text for search. Each API call will return a number of search hits, including "snippets" that match the text. In the case of plagiarism, this will be an almost exact match. Thus, there is a drastic decrease in "edit distance" in the case of copy-paste plagiarism. This can be seen easily in Table H.1.

The difference appears to be coarse enough that a variety of metrics are sufficient to separate out cases of plagiarism. Metrics evaluated were "FuzzyWuzzy" (provided by SeatGeek at https://github.com/seatgeek/fuzzywuzzy) and the classic Levenshtein, or "edit" distance from the NLTK package (available at http://nltk.org/). Both of these are measures of partial string similarity. The code used to perform these analyses is available upon request.

Table H.1: Partial list of (sorted) distance measures from the top google hit for a number of texts. As you can see, the first three rows are markedly different from those that followed. The First column is a true "distance" and as such, larger numbers indicate less similarity. The "FuzzyWuzzy" measures are calculated as a percentage.

Levenshtein	FuzzyWuzzy	FuzzyWuzzy (partial)
17	94	95
17	94	95
54	88	97
134	32	31
122	32	34
146	32	37
102	31	34
127	30	35
126	30	29
139	30	30

Appendix I

Mechanism Items and Coding Scheme for Responses

I.1 Materials Used by All Coders

Development of the coding scheme was a multi-step process. Initially, two members of our group, Sarah Cohen and Roxana Farjadi, saught to identify conceptions that occurred across multiple surveys. These conceptions were assigned numerical codes, and these codes were arranged into general categories. Following this, I developed a more complete progression, describing relationships between the various categories, as well as grouping them into "misconceptions," "ignorance," and "mechanistic description." This allowed the beginnings of a scoring rubric to be developed. We then iterated the process with a larger group of coders to arrive at the final product reproduced below. What follows is the full text of the coding packet developed by Ms. Cohen and Ms. Farjadi, which also contains the text for the mechanism questions we asked in our interventions. Given the centrality of these questions, we produce them here as well:

1. Please write 1-3 sentences (about 30 words or less) that you could use to explain how climate change occurs to a senior in high school.
2. Please explain any differences regarding how energy (i.e., heat, light) travels to the Earth from the sun compared to how energy travels away from the Earth.
3. Are all gases "greenhouse gases?" If not, what makes something a greenhouse gas?

Note that S. Cohen (2012) also reports on a coding schema, though that scheme exhibits differences with the one described here. Following this are a diagram representing relationships between the codes. A section containing a set of notes provided by Myles Crain are included in the next section. They provides a set of criteria for chosing between notes, and was used by the final set of coders. See chapter 6 for details.

Instructions

1. Responses can be classified in three categories at most. Give them as many codes as possible.
2. If the respondent talks about the differentiation of energy, refer to the "definition of differentiation of energy" table for additional help in categorizing.
3. If the respondent talks about *how greenhouse gases work*, refer to the "definition of greenhouse gases" table for additional help in categorizing.
4. If the respondent mentions greenhouse gases, refer to the "says/mentions greenhouse gases" table for additional help in categorizing.
5. If the respondent talks about *any type of mechanism for climate change*, refer to the "mechanism of climate change table." This table is broken into the sub-categories of energy, source, general chemical reactions, and respondent confusion. Please note that sometimes *a response can fit into more than one subcategory* under the overarching mechanism category.
6. If the respondent leaves a question blank, writes "do not know," or "same as above," refer to the last table, "Don't Know."
7. If the response prompts categorization ambiguities, first look at the response as a whole to look for phrases that might provide a clearer indication of what they mean. If the ambiguity can be clarified without coder inferences or assumptions, categorize the response into the code that provides the most possible credit (i.e., "be charitable within reason"). If the coder cannot clear up the ambiguity or must make assumptions, code the response into the category which best describes what the respondent actually says and not what the coder might think they are trying to say (i.e., "don't infer extra credit"). Also, note whether the respondent is defining something, explaining how climate change works, or *both*. To be doing both, the ideas must be clearly a definition and a mechanism. For instance, to say "greenhouse gases do X and thus trap heat on earth" would be both a definition and a mechanism. Even if a definition is embedded in a phrase that describes the mechanism, give them credit for both the mechanism and the definition.
8. Unless otherwise noted, all the categories listed can be applied to Know_1, Know_2, or Know_3.
9. See example column for examples of each code. Please note that for each example, the response may have been coded into more categories than just the category in which the example is placed (e.g., the example for MCCS2 was coded into SGHG1 as well as MCCS2).

<u>Definition of Terms</u>

Know_1: Please write 1-3 sentences (about 30 words or less) that you could use to explain <u>how climate change occurs</u> to a senior in high school.

Know_2: Please explain any differences regarding how energy (i.e., heat, light) travels to the Earth from the sun compared to how energy travels away from the Earth.

Know_3: Are all gases "greenhouse gases?" If not, what makes something a greenhouse gas?

Categories (Listed) – Please see tables for cutoffs, discussions, and comparisons between categories.

DD: Definition of the Differentiation of Light/Energy

DD1: Respondent differentiates between visible sunlight entering the atmosphere and infrared radiation/heat being emitted by the Earth.

DD2: Partial credit for differentiation: Respondent attempts to explain how energy differs when it enters the atmosphere and when it leaves, but does so in such a way that is either too incomplete or incorrect to fit into category DD1. Category DD2 is therefore "partial credit" for DD1. As long as the participant references some kind of asymmetry in how light is reflected, bounced, changed, etc. (even if mostly wrong), they fall in category DD2 and not DD3.

DD3: Completely incorrect attempt to differentiate kinds of light/energy – This only applies to when there is absolutely NO asymmetry referenced.

DGHG: Definition of Greenhouse Gases

DGHG1: Greenhouse Gas "right *definition*" – Respondent may or may not mention the exact phrase "greenhouse gas", but at least defines them in the right context. Respondent defines greenhouse gases as molecules that *absorb* energy, not as molecules that trap, stop, block, or reflect energy. Respondent may use the terms light, heat, radiation, or infrared radiation instead of energy in their definition.

DGHG2: Greenhouse Gas "partial credit *definition*" – Respondent may have demonstrated an understanding of some of the elements outlined in category DGHG1 but their answer is either too grammatically vague to pass judgment on correctness or contains elements of *incorrect* content ("partial credit"). To get a definition code, the respondent has to mention or allude to energy. Remember that responses in this category do *not* describe greenhouse gases as molecules that "*absorb energy."*

DGHG3: Not all gases are greenhouse gases: Respondent directly answers the question in Know_3 by stating in some way that not all gases are greenhouse gases.

DGHG4: Wrong concept of greenhouse gas: The participant holds obvious misconceptions about what a greenhouse gas is or how it works.

SGHG: Says/mentions greenhouse gases - If they give at least some definition or statement as to *what* greenhouse gases do or *how* they work, refer to DGHG categories.

SGHG1: *In know_1*: Simple mention of greenhouse gases (no explanation) –Participant uses the term "greenhouse gas," or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* explanation of climate change.

In know_2: Simple mention of greenhouse gases - participant uses the term "greenhouse gas," or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* explanation or strongly implied understanding of the concept of how energy functions in the atmosphere

In know_3: Simple mention of greenhouse gases - participant uses the term "greenhouse gas," or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* explanation or strongly implied understanding of the concept of a greenhouse gas.

SGHG2 : *In know_1*: Simple mention of greenhouse gases –Respondent uses the term "greenhouse gas," or provides a specific example of one, like carbon dioxide, in the context of a mostly *incorrect* explanation of climate change.

In know_2: Simple mention of greenhouse gases - participant uses the term "greenhouse gas," or provides a specific example, like carbon dioxide, in the context of a mostly *incorrect* explanation or strongly implied understanding of the concept of how energy functions in the atmosphere

In know_3: Simple mention of greenhouse gases - participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a mostly *incorrect* explanation or strongly implied understanding of the concept of a greenhouse gas.

SGHG3: Mentions greenhouse effect – Respondent explicitly uses the phrase “greenhouse effect,” or some variation thereof. The respondent may or may not offer an explanation of what the greenhouse effect is or how it works.

MCC: Mechanism of Climate Change, broken up by concept

MCCE: Mechanism of climate change, energy

MCCE1: Atmosphere Retention time: Respondent describes how long it takes for heat to leave the atmosphere in depth. They reference that there are “more” greenhouse gases now than there were before, which causes heat to stay in the atmosphere longer OR causes *more* heat to stay in the atmosphere (either time or amount are permissible in this category). The explanation must be in the context of comparing a previous instance when greenhouse gases existed to the presence of greenhouse gases in the atmosphere today.

MCCE2: Trapped heat as a *mechanism* for climate change: Respondent describes heat/energy/radiation as being trapped. They may describe energy changes but lack a comparison from our time to a previous time with greenhouse gases. For inclusion in this category, the respondent must use the idea of “trapping” or “stopping” heat from leaving and must NOT attempt to use the concept of energy being “trapped” as a definition of greenhouse gases– that would fall into category DGHG2. However, there are responses that may be coded as both categories MCCE2 and DGHG2 if the respondent separately defines greenhouse gases, as guided by the definition of category DGHG2, and describes the mechanism of climate change as trapping heat.

MCCE3: Input rate/amount of energy does not equal output rate/amount of energy – Respondent demonstrated some knowledge that rate/amount of energy input is different from the rate/amount of energy output, and so energy is “stuck” somewhere OR energy is “slowed down.” If the person does NOT reference a previous time with less GHGs, but does talk about heat being slowed or hindered from leaving the atmosphere, this category applies. Also, this category classifies responses that are vaguer than those in category MCCE2 or MCCE1.

MCCE4: Radiation from the sun directly heats the atmosphere – Respondent explicitly states or strongly implies that the atmosphere is heated by radiation from the sun. Respondent does not mention that Earth absorbs/reemits energy (i.e., the respondent skips differentiating energy).

MCCS: Mechanism of climate change, source

MCCS1: Human element: Respondent states or heavily implies that human emissions of greenhouse gases cause or contribute to global warming. This category includes references to fossil fuels and technology as causes of climate change.

MCCS2: Natural variation/weather patterns as an explanation for climate change: Respondent references natural variation in weather patterns as a cause of climate change thereby implying that anthropogenic emissions (“the human element”) are not the only causes of climate change.

MCCS3: Pollution: Respondent explicitly states or strongly implies that pollution causes global warming, with no explicit reference to energy’s function in the warming of the earth. This category also includes responses where the respondent seems to think that pollution physically “thickens the atmosphere” and thus causes warming. If the person references pollution (as opposed to greenhouse gases) as causing global warming, the response fits in this category.

MCCS4: Ozone: Respondent talked about the *depletion* of the ozone layer causing global warming.

MCCR: Mechanism of Climate Change, General Chemical Reactions

MCCR: Chemical Reactions and/or molecular properties explanations: participant attempts to explain the difference between energy entering Earth’s atmosphere and energy exiting Earth’s atmosphere from a strictly chemical perspective. Response does not include explicit differentiation between energies but rather uses chemical reactions in themselves as the cause of warming. A molecular perspective involving vibrations or other molecular properties may be used instead of chemical reactions or in addition to them. Response is too general to be given credit for categories DD1 or DGHG1.

MCCQ: Mechanism of Climate Change, Confused Respondent

MCCQ1: General Weather Confusion: Respondent thought we were asking about the seasons. The respondent may describe weather patterns, Earth's rotations, or the tilt of the Earth's axis.

MCCQ2: Did not understand: Respondent supplies a completely irrelevant answer (i.e. talks about high school perspectives).

DNK: Don't know or blank

DNK1: Don't know or N/A

DNK2: Code here if the participant uses a phrase similar to "I wouldn't add anything" or same as above.

Categories (organized by keyword)

Name of Category	Definition of Differentiation of Energy: DD In descending order from most thorough to least thorough	Distinctions:	Examples:
DD1	Respondent differentiates between visible sun light entering the atmosphere and infrared radiation/heat being emitted by the earth.	This category is fairly easy to find; if respondent say "reflected" IR (instead of absorbed and reemitted) that still fits here, provided that they made **some distinction** between light coming in and light going out.	"higher frequency radiation from the sun enters easily, but the lower frequency radiation reemitted by the cooler earth" (1Post) "the sun emits energy ans the earth absorbs that energy and then infared light comes back" (25Post)
DD2	Partial credit for differentiation: Respondent attempts to explain how energy differs when it enters the atmosphere and when it leaves, but does so in such a way that is either too incomplete or incorrect to fit into category DD1. Category DD2 is therefore "partial credit" for DD1. As long as the participant references some kind of asymmetry in how light is reflected, bounced, changed, etc. (even if mostly wrong), they fall in category DD2 and not DD3.	For example participant responses **may** include: -Failure to say how visible light becomes infrared -Failure to mention visible light AND infrared light (or heat) -Other **partially incorrect** attempts at differentiation	"Energy traveling to earth is converted to infared, [this energy can be absorbed by greenhouse gases]" (4Post). "The earth emits shorter wavelengths of energy whereas the sun emits longer ones." (6 Post)
DD3	Completely incorrect attempt to	Fails to understand that there is a difference in	"No difference on how energy travels."

	differentiate kinds of light/energy; this only applies to when there is absolutely NO asymmetry referenced	incoming and outgoing energy.	(27 Pre)

Name of category	Definition of Greenhouse Gas : DGHG In descending order from most thorough to least thorough	Distinctions:	Examples:
DGHG1	Greenhouse Gas “right *definition*” – Respondent may or may not mention the exact phrase “greenhouse gas”, but at least defines them in the right context. Respondent defines greenhouse gases as molecules that *absorb* energy, not as molecules that trap, stop, block, or reflect energy. Respondent may use the terms light, heat, radiation, or infrared radiation instead of energy in their definition.	If you are having trouble deciding between DGHG1 and DGHG2, look at the context in which the definition of a greenhouse gas is given. Furthermore, if you really cannot tell what they are saying (because of grammar or vagueness) pick DGHG2. To be qualified in DGHG1, the respondent has to give some indication that they know *how* greenhouse work, not just that they cause something to happen, resulting in warming. (If respondent uses the concepts of trapping, stopping, blocking, or reflecting energy the response belongs in category DGHG2.) It doesn’t matter for this category where the respondent thinks the energy comes from.	“Greenhouse gases absorb the reflected light…” (2Post) “Only the ones that can absorb infared light, like CO2 are considered greenhouse gases...”(3 Post)
DGHG2	Greenhouse Gas “partial credit *definition*” – Respondent may have demonstrated an understanding of some of the elements outlined in category DGHG1 but their answer is either too grammatically vague to pass judgment on correctness or contains elements of *incorrect* content (“partial credit”). To get a definition code, the respondent has to mention or allude to energy. Remember that responses in this category do *not* describe greenhouse gases as molecules that “*absorb energy.”*	Remember, this is the “Partial Credit” category. Cut-off: When respondent tries to explain the function of a greenhouse gas the response fits in DGHG2 when they do **not** say absorb.	“Climate change occurs due to the abundance of greenhouse gases in the atmosphere. Greenhouse gases, like co2, are slowly emitted into the atmosphere as energy, but as the abundance of this gas increases, it slowly warms up the earth, b/c greenhouse gases are created at a faster rate than they absorb infared light” (14 Post) “Carbon gases are released into the air that trap extra light” (16 Post)

DGHG3	Not all gases are greenhouse gases: Respondent directly answers the question in Know_3 by stating in some way that not all gases are greenhouse gases.	Just have to say “no” in some way, but do **not** have to understand why. Can also give counterexample to count in this category (e.g. saying, “N2 is not a greenhouse gas”).	“No, a greenhouse gas is referring to…” (21Pre) “not all gases are greenhouse gases. No clue what makes a greenhouse gas a greenhouse gas” (24Pre)
DGHG4	Wrong concept of greenhouse gas: The participant holds obvious misconceptions about what a greenhouse gas is or how it works.	If there is some modicum of correctness do **not** put the response here. Give them the credit for what they know.	“Greenhouse gases are the gases that remain in the earth's atmosphere. They are unable to leave” (30Pre)

Name of category	Says/Mentions Greenhouse Gases: SGHG In descending order from most thorough to least thorough	Distinctions:	Examples:
SGHG1	-*In know_1*: Simple mention of greenhouse gases (no explanation) – Participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* mostly *correct* explanation of climate change. -*In know_2:* Simple mention of greenhouse gases - participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* mostly *correct* explanation or strongly implied understanding of the concept of how energy functions in the atmosphere -*In know_3*: Simple mention of greenhouse gases - participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a moderately or mostly *correct* mostly *correct* explanation or	If they do **not** describe the behavior of greenhouse gases, examine the context. If they mention it in a moderately or mostly correct context, then the response fits in SGHG1. Parts of the response can be wrong or irrelevant, but if they use the term greenhouse gases in a mostly correct context, SGHG1 is appropriate. This response does **not** fit into category DGHG1 because it does **not** say that greenhouse gases trap heat. Saying that GHGs cause warming does **not** give enough indication of understanding of *how* GHGs interact with energy. This response also does **not** fit into category MCCS3 because it does **not** specify that GHGs intrinsically cause warming.	“Climate change … can also be induced unnaturally by greenhouse gas buildup from carbon emissions” (13 Pre)

	strongly implied understanding of the concept of a greenhouse gas.		
SGHG2	-*In know_1*: Simple mention of greenhouse gases –Respondent uses the term “greenhouse gas,” or provides a specific example of one, like carbon dioxide, in the context of a mostly *incorrect* explanation of climate change. -*In know_2:* Simple mention of greenhouse gases - participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a mostly *incorrect* explanation or strongly implied understanding of the concept of how energy functions in the atmosphere -*In know_3*: Simple mention of greenhouse gases - participant uses the term “greenhouse gas,” or provides a specific example, like carbon dioxide, in the context of a mostly *incorrect* explanation or strongly implied understanding of the concept of a greenhouse gas.	Responses fit into category SGHG2 when they mention GHGs (or a type of GHGs) but do so in a mostly **incorrect** explanation. When participants refer to ozone depletion as the main cause of global warming, for example, it is **incorrect**. Because this response does **not** explain how GHGs work, and the context is incorrect, it fits into SGHG2.	“Climate change occurs when the weather patterns abruptuly change and are abnormal. It occurs because of greenhouse gases such as carbon dioxide released into the atmosphere.” (35 Pre)
SGHG3	Mentions greenhouse effect – Respondent explicitly uses the phrase “greenhouse effect,” or some variation thereof. The respondent may or may not offer an explanation of what the greenhouse effect is or how it works.	If respondent defines GHGs correctly and then mentions the greenhouse effect separately, SGHG3 and DGHG1 can be used to categorize the same response. However, usually SGHG3 is used in place of DGHG1.	“climate change occurs due to an increase of trapped infared light in our atmosphere which is caused by the greenhouse effect.” (21 Post)

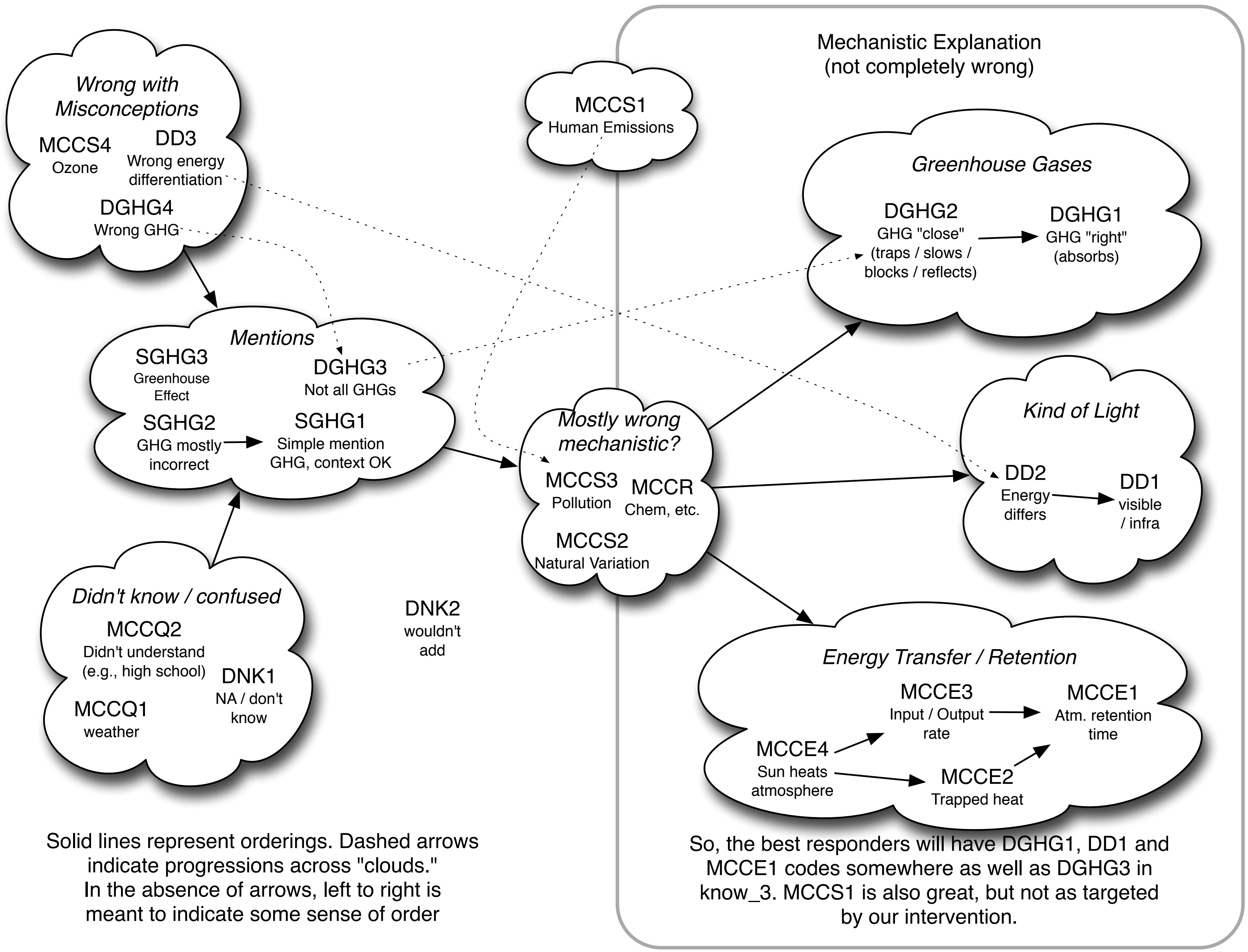
Name of category	Mechanism of Climate Change: MCC	Distinctions:	Examples:
ENERGY, Mechanism of Climate Change: MCCE			
MCCE1	Atmosphere Retention time: Respondent describes how long it takes for heat to leave the atmosphere in depth. They reference that there are “more” greenhouse gases now than there were before, which causes heat to stay in the atmosphere longer OR causes *more* heat to stay in the atmosphere (either time or amount are permissible in this category). The explanation must be in the context of comparing a previous instance when greenhouse gases existed to the presence of greenhouse gases in the atmosphere today	MCCE1 needs to have some sort of comparison to another time when there were not as many GHGs in the atmosphere. If they do not, then the response likely fits into MCCE2 or MCCE3. MCCE1 is the most specific category. Often there will be reference to “slowing” or “preventing” the escape of heat from the atmosphere	“Greenhouse gases absorb the reflected light and cause the earth to heat up (when more gases, slower rate of expulsion + therefore more heat” (2post) “but currently too much carbon gases are released into the air that trap extra light (heating earth up more than usual” (16 Post)
MCCE2	Trapped heat as a *mechanism* for climate change: Respondent describes heat/energy/radiation as being trapped. They may describe energy changes but lack a comparison from our time to a previous time with greenhouse gases. For inclusion in this category, the respondent must use the idea of “trapping” or “stopping” heat from leaving and must NOT attempt to use the concept of energy being “trapped” as a definition of greenhouse gases– that would fall into category DGHG2. However, there are responses that may be coded as both categories MCCE2 and DGHG2 if the respondent separately defines greenhouse gases, as guided by the definition of category DGHG2, and describes the mechanism of climate	This response fits into MCCE2 and **not** MCCE1 because it does **not** say that the more greenhouse gases there are in the atmosphere, the longer the energy stays in the atmosphere. Rather, it implies that there is a threshold beyond which energy “lingers” in the atmosphere. MCCE2 is almost MCCE1, but there is either a slight misunderstanding or miscommunication in the wording of the response (i.e., this category is partial credit). If energy being “trapped” is used to *define* a GHG, the response is coded in DGHG2 so as to avoid giving credit twice.	“Climate change is a gradual heating of the Earth's atmosphere due to trapped heat” (30 post) “co2. that creates a layer in our planet's atmosphere which traps sunlight and warms up the earth.” (12 Pre)

	change as trapping heat.		
MCCE3	Input rate/amount of energy does not equal output rate/amount of energy – Respondent demonstrated some knowledge that rate/amount of energy input is different from the rate/amount of energy output, and so energy is “stuck” somewhere OR energy is “slowed down.” If the person does NOT reference a previous time with less GHGs, but does talk about heat being slowed or hindered from leaving the atmosphere, this category applies. Also, this category classifies responses that are vaguer than those in category MCCE2 or MCCE1.	If trying to decide between MCCE1, MCCE2, and MCCE3, first ascertain if there is a comparison to a different time with a different level of greenhouse gases. If yes, then MCCE1. Otherwise, look at the clarity: if they say heat is being STOPPED or TRAPPED, the response goes in MCCE2; if the response talks about how energy is slowed or hindered, then MCCE3.	“Cimate change is the heating up of the earth - above its normal temperature. It is caused by waves of heat leaving the earth's atmosphere, but certain greenhouse gases has caused the waves to leave even more slowly, causing the earth to be at a higher temperature.” (6 post) “it releases infared light which gets absorbed by the greenhouse gases in our atmosphere causing the earth to heat up” (15 Post) – This is a good example of both a definition and a mechanism.
MCCE4	Radiation from the sun directly heats the atmosphere – Respondent explicitly states or strongly implies that the atmosphere is heated by radiation from the sun. Respondent does not mention that Earth absorbs/reemits energy (i.e., the respondent skips differentiating energy).	If the respondent only refers to radiation **from the sun** heating greenhouse gases, then it fits in MCCE4. In other words, it will **not** fit into category DD1 because it fails to explain differentiation. Additionally, if the mechanism by which energy from the sun reaches the Earth is ambiguous and there are no clear indications in the rest of the response to suggest that the energy reaches the Earth’s surface, then the response should be classified in MCCE3.	“The atmosphere traps energy traveling from the sun.” (49 Pre)
SOURCE, Mechanism of Climate Change: MCCS			
MCCS1	Human element: Respondent states or heavily implies that human emissions of greenhouse gases cause or contribute to global warming. This category includes references to fossil fuels and technology as causes of	This category will include any reference to how humans cause climate change, e.g. the Industrial Revolution, cars, oil combustion, etc.	“Greenhouse gases emited by our cars, and industrial process and other human activity involving the burning of fossil fues or other combustables” (18 Pre)

	climate change.		
MCCS2	Natural variation/weather patterns as an explanation for climate change: Respondent references natural variation in weather patterns as a cause of climate change thereby implying that anthropogenic emissions (“the human element”) are not the only causes of climate change.		“Climate change is a natural process (ice age - el nino) an can also be induced unnaturally by greenhouse gas buildup from carbon emissions” (13 Pre)
MCCS3	*MCCS3*: Pollution: Respondent explicitly states or strongly implies that pollution causes global warming, with no explicit reference to energy’s function in the warming of the earth. This category also includes responses where the respondent seems to think that pollution physically “thickens the atmosphere” and thus causes warming. If the person references pollution (as opposed to greenhouse gases) as causing global warming, the response fits in this category.	This category needs some sort of implication that humans or “waste” emissions warm up the atmosphere by themselves, with no regard for energy’s role.	“We produce too much carbon as waste. It ends up in the atmosphere. Heats up.” (31 Pre)
MCCS4	*MCCS4*: Ozone: Respondent talked about the *depletion* of the ozone layer causing global warming.	If the respondent claims that ozone depletion causes climate change, it goes into MCCS4.	“ozone depletion also affect how the sun's heat and light is absorbed in our atmosphere and cause climate change.” (28 Pre)
GENERAL CHEMICAL REACTIONS, Mechanism of Climate Change: MCCR			
MCCR	Chemical Reactions and/or molecular properties explanations: participant attempts to explain the difference between energy entering Earth’s atmosphere and energy exiting Earth’s atmosphere from a strictly chemical	Responses fit into this category if they provide a very general attempt to describe heat in the atmosphere. Often the respondent has misconceptions about the role of chemicals in the atmosphere and therefore their response **cannot** fit into categories DD1 or DGHG1 as	“The sun directly enters the earth causing many chemical reactions. The earths byproducts of these chemical reactions let out either heat or molecules. Some molecules reabsorb the heat and create global warming” (5

	perspective. Response does not include explicit differentiation between energies but rather uses chemical reactions in themselves as the cause of warming. A molecular perspective involving vibrations or other molecular properties may be used instead of chemical reactions or in addition to them. Response is too general to be given credit for categories DD1 or DGHG1.	well as this one.	post) "Energy travels to the earth from the sun in the rays of heat of the sun in the form on molecules in constant motion. Energy travels away from earth by the same force of interacting and fast moving molecules" (6 pre)
RESPONDENT CONFUSION, Mechanism of Climate Change: MCCQ			
MCCQ1	General Weather Confusion: Respondent thought we were asking about the seasons. The respondent may describe weather patterns, Earth's rotations, or the tilt of the Earth's axis.	Respondent could talk about seasons in conjugation with actual explanation of global warming. Read the whole response before coding.	"Climate change occurs when the sun is hitting the earth from a different angle. When it is winter, the sun's rays are less direct. In the summer, there are longer days w/ more direct sunlight" (21 Pre)
MCCQ2	Did not understand: Respondent supplies a completely irrelevant answer (i.e. talks about high school perspectives).		"It is senior year that students begin to get tired of the hgh school environment and are anxious to open a new chapter of their lives: colege. This is called senioritis. Therefore a climate change occurs to a senior in highschool when he/she is ready to leave high school and move on" (6 Pre)

Number of Category	Don't know: DNK	Distinctions:	Examples:
DNK1	N/A: maybe ran out of time.		"I do not know how climate change occurs I was never taught." (24Pre)
DNK2	Code here if the participant uses a phrase similar to "I wouldn't add anything" or "same as above."		"I wouldn't add anything." (3 Post)


Mechanistic Explanation
(not completely wrong)
Wrong with Misconceptions
MCCS4
Ozone
DD3
Wrong energy differentiation
DGHG4
Wrong GHG
MCCS1
Human Emissions
Greenhouse Gases
DGHG2
GHG "close" (traps / slows / blocks / reflects)
DGHG1
GHG "right" (absorbs)
Mentions
SGHG3
Greenhouse Effect
DGHG3
Not all GHGs
SGHG2
GHG mostly incorrect
SGHG1
Simple mention GHG, context OK
Mostly wrong mechanistic?
MCCS3
Pollution
MCCR
Chem, etc.
MCCS2
Natural Variation
Kind of Light
DD2
Energy differs
DD1
visible / infra
Didn't know / confused
MCCQ2
Didn't understand (e.g., high school)
DNK1
NA / don't know
MCCQ1
weather
DNK2
wouldn't add
Energy Transfer / Retention
MCCE3
Input / Output rate
MCCE1
Atm. retention time
MCCE4
Sun heats atmosphere
MCCE2
Trapped heat
Solid lines represent orderings. Dashed arrows indicate progressions across "clouds." In the absence of arrows, left to right is meant to indicate some sense of order
So, the best responders will have DGHG1, DD1 and MCCE1 codes somewhere as well as DGHG3 in know_3. MCCS1 is also great, but not as targeted by our intervention.

I.2 Notes on Choosing Codes

This “crib sheet” was generated by Myles Crain to identify a single defining characteristic and/or unique distinction within each code. Here are a few notes on how it was used:

- The crib sheet is NOT self-contained. Its meant to jog memory without having to constantly flip through the coding packet. The sheet is only useful if you are generally familiar with the coding scheme already.
- Assigning a code should be defensible with explicit references to the definitions and explanations of that code as provided in the packet.
- I've separated DGHG3 from the other DGHG codes intentionally (that is, DGHG3 coming after DGHG4 is NOT a typo).
- SGHG codes are only supposed to be used in the complete absence of a definition of GHGs. The SGHG category is primarily useful in coding for whether an explanation of climate change includes explicit reference to GHGs.
- Use MCCE codes to identify how a participant refers to energy within an explanation of climate change.
- Enquoted things are things that must appear in a response in order to apply the code (except when there are other options–for example, SGHG1 requires using the phrase ”GHG” *or* citing specific examples of GHGs).

Following is the “crib sheet” itself:

DD1 visible incoming & infrared outgoing

DD2 asymmetry/difference reference

DD3 wrong, no asymmetry/difference

DGHG1 GHGs ”absorb” energy

DGHG2 part correct, no ”absorb”

DGHG4 wrong

DGHG3 ”not all”, cite >1

SGHG1 ”GHG”/e.g., mostly accurate

SGHG2 ”GHG”/e.g., mostly wrong

SGHG3 ”greenhouse effect”

MCCE1 more gas/heat than before

MCCE2 heat/energy ”trapped”

MCCE3 different input/output rates/amounts

MCCE4 sun's radiation heats atmosphere

MCCS1 humans/tech/fossil fuels

MCCS2 natural variation

MCCS3 pollution

MCCS4 O3 layer

MCCR chemical/molecular exclusively

MCCQ1 weather, confusion

MCCQ2 irrelevant

DNK1 "don't know", n/a

DNK2 nothing added

I.3 Assigning Scores to Coded Responses

Knowledge scores were computed based on the ordering of mechanistic codes in the "cloud" diagram in Section I.1. For each category, points were assigned on the basis of the highest-scoring code present (e.g., if a one-point and two-point code were present, the category would be scored for two points). Each of three subscores was computed as follows:

GHG One point was assigned for any mention or partially incorrect description: SGHG1, SGHG2, SGHG3, MCCS3, or MCCR. Two points were assigned for DGHG2. Three were assigned for DGHG1.

Light Here, there were only two point-earning codes. Thus, we assigned a score of 1.5 for DD2, and three for DD1.

Engery Responses coded with MCCE4 earned one point, MCCE3 or MCCE2 earned two points, and MCCE1 earned the maximum of three points.

A combined score was computed by adding all subscores, resulting in a total score of 0–9.

Appendix J

Extension to Cohen's κ for Multiple Scores

J. Cohen (1960) describes a commonly used measure of inter-rater agreement, κ, in which different raters each apply one of a number of scores to a set of responses (usually written text, such as collected in the studies in Chapter 6). Specificaly, κ is defined as:

$$\kappa = \frac{p_o - p_c}{1 - p_c}$$

Here, p_o is observed probability (or fraction) of agreement, and p_c is our expectation of chance agreement. Thus when $\kappa = 1$, we have obtained perfect observed agreement ($p_o = 1$), and $\kappa \to 0$ as $p_o \to p_c$. Note that κ can also become a large negative number when structural disagreement is present.

There are weighted and nonweighted variants of this measure. However, given the greater complexity of the weighted version and equivalence of results across weighted and nonweigted versions of κ, for simplicity's sake we will focus here on nonweighted κ. In this formulation, agreement occurs when both raters provide the same code. Thus, if there are two possible codes ("a" and "b"), there are four possibilities for two coders rating the same item. Two such possibilities are agreement, thus $p_c = .5$. We would obtain p_o by counting the fraction of occurrences of agreement out of the total number of rated items.

Our texts recieve a number of codes, so we needed an extension to κ that could capture such a scheme. In such a scheme, it is natural that if both coders provide the same set of codes, they are given a 1, or credit for full agreement. If they provide no common codes, they recieve a 0. Partial credit when both coders provide the same number of codes is equally simple. Consider the case where coders provide N codes, m of which are common. This item then recieves a score of m/N. In the case of an unequal number of codes, N is set to the (potentially non-integer) average of the two numbers. Given these scores for each item, κ is computed normally as the ratio between observed and chance agreement.

Computationally, we construct a matrix of codes C_i for each coder. Each row represents a particular participant response, with a "1" in the column corresponding to each of the assigned codes (and "0" elsewhere). We can readily compute agreement (m) as the dot-product between the relevant rows. Given our usage of an unproven statistic, we will compute agreement between any random pairing of rows as well, deriving a "raw" measure of agreement:

$$R = C_1 C_2^T$$

We then compute the mean number of codes for each pair of rows, and norm each cell by that number. We can freely sample from this normed matrix of agreements to obtain an empirical distribution for chance agreement given the structure of two random sets of codes. Actual observed agreement will be available along the diagonal (or "trace") of this agreement matrix. This code, as with all code used in this dissertation, is available upon request.

A note on using κ in the text above: It seems completely admissable to continue to use κ for the measure described above, given it's similarity in spirit to Cohen's original measure. This seems doubly admissable considering that formal distributional considerations appear to be considered rarely, if ever, in the psychological literature. Thus, any deviations in the specifics of formal properties are unlikely to be a concern for the casual reader.

Appendix K

Using Imputation to Combine Participants with and without a Pretest

Imputation is a well-established approach to dealing with missing data (for an overview see Fox, 2008, p. Chapter 20). In a number of the experiments in this dissertation, multiple groups received a similar intervention, but one group may have been missing a pre-test in which we obtained their naïve baseline score (e.g., for a climate-relevant attitude). The approach we used in these cases was to use the participants for which we *did* have a pre-test score (i.e., our sandwich group), and use the average of those as an approximation to our other group's pre-test score. To be explicit, following is the exact R code used to compute this test for Study 1 in Chapter 6:

```
# Here, dfs is pre-populated with the measured values. We assign the mean of
# the sandwich (s) group scores to the pre-test scores for the no-pretest
# group (n). We then append the sandwich group scores unmodified.

imputed.df <- data.frame(pre.gw=mean(dfs$s.pre$total.gw),
                         total.gw=dfs$n.post$total.gw)

imputed.df <- rbind(imputed.df,
                    dfs$s.post[,c('pre.gw', 'total.gw')])

# Note - this gives the same result as a simple t-test on the difference
# scores, so we're not cheating on our degrees of freedom, or obtaining
# artificially lower variance on the pre-test scores.

with(imputed.df,
     t.test(pre.gw, total.gw, alternative='less', paired=TRUE) )
```