A method to evaluate the usability of interactive climate change impact decision aids

Gabrielle Wong-Parodi, Baruch Fischhoff & Ben Strauss

ONLINE FIRST

NO AUTHOR CHARGES FOR COLOR PRINTING

🖄 Springer

🖄 Springer

Springer

LETTER

A method to evaluate the usability of interactive climate change impact decision aids

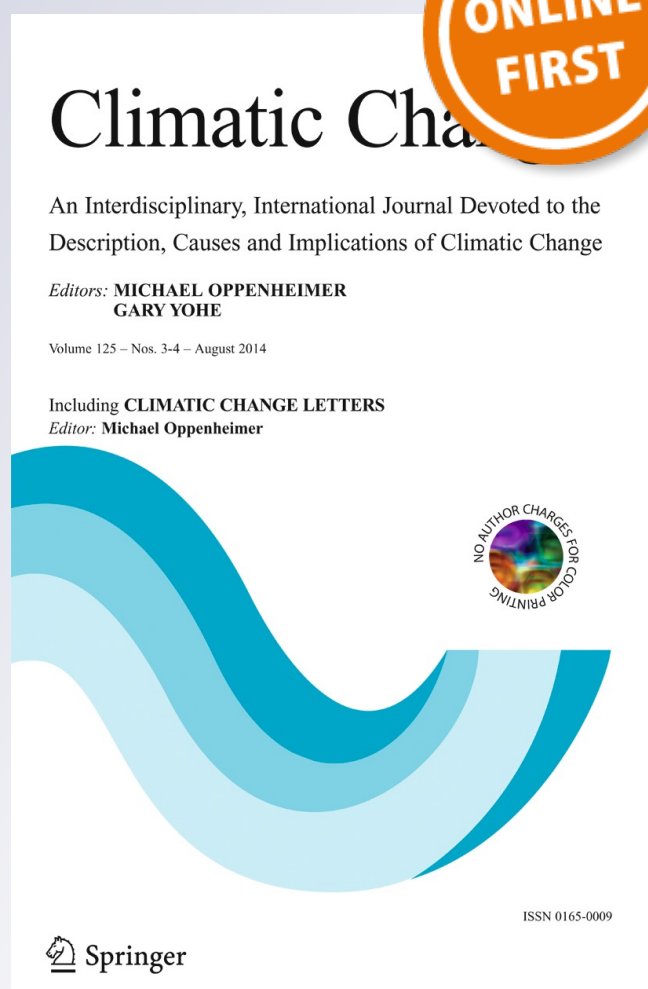**Gabrielle Wong-Parodi · Baruch Fischhoff · Ben Strauss**

Abstract Reducing the impacts from climate change requires people to make decisions that may prompt substantial changes in their lives. One possible way to help them is with personalized decision aids. Here we describe a method for evaluating such aids, in terms of how they affect users' understanding of their situation, defined in terms of their (a) knowledge, (b) consistency of preferences, and (c) active mastery of the material. Our method provides a simple way to evaluate the usability of climate-change decision aids, and to address concerns that the choice of display could bias users' attitudes. We demonstrate it with the Surging Seas Risk Finder, a decision aid focused on coastal flooding (http://sealevel.climatecentral.org/).

1 Introduction

The specter of climate change looms on the horizon. In policy and scientific communities, there is growing recognition that reducing its impacts will require people to make decisions leading to substantial changes in their lives (Dietz et al. 2009; Füssel 2007; Padgett et al. 2008). The hope of improving those decisions has prompted the creation of many decision aids, from 'carbon calculators,' focused on decisions about transportation or food consumption (Bottrill 2007; Chatterton et al. 2009), to seasonal forecasts for farmers making decisions about planting, water allocation, and fertilization (Meinke and Stone 2005). These aids often offer rich information, drawing on resources created for professionals (Matthies et al. 2007), such as Geographic Information Systems (GIS) (Simao et al. 2009), forecasts (Meinke and Stone

G. Wong-Parodi (✉)
Department of Engineering and Public Policy, Carnegie Mellon University, 129 Baker Hall, Pittsburgh, PA 15213, USA
e-mail: gwparodi@gmail.com

Gabrielle Wong-Parodi
e-mail: gwongpar@cmu.edu

B. Fischhoff
Departments of Engineering and Public Policy and Social and Decision Sciences, Carnegie Mellon University, 219E Baker Hall, Pittsburgh, PA 15213, USA
e-mail: baruch@cmu.edu

B. Strauss
Climate Central, One Palmer Square, Suite 330, Princeton, NJ 08542, USA
e-mail: bstrauss@climatecentral.org

🖉 Springer

2005), and simulations (Nicholson-Cole 2005). However, that very richness runs the risk of inundating lay users with more, and more technical, information than they can handle. That cognitive overload may also render them vulnerable to biases induced by how issues are framed, as they look for clues to orient them in the flood of information (Kahneman 2011).

How well any communication achieves its goals is an empirical question. Recognizing that the design teams for many decision aids lack the resources for full-scale evaluations, we offer a simple, general approach that any team could use, so that no aid relies solely on the intuitions of its designers – whose perspectives may be very different from that of the decision makers whom they hope to serve. Our approach draws on principles from risk communication and human-computer interaction research (Fischhoff et al. 2011; Fischhoff 2013; Noar et al. 2007; Szwajcer et al. 2009; Salvendy 2012; Olson and Olson 2003). We illustrate it with an aid for decisions regarding coastal flooding.

Informed decision-making requires people to understand the benefits, risks, and uncertainties of their options well enough to identify choices consistent with their personal preferences (Braddock III et al. 1999). The value of providing people with good information has been observed in domains as diverse as carbon mitigation (de Best-Waldhober et al. 2009), financial investments (Lusardi 2008), domestic violence (Ver Steegh 2002), and energy policy (von Winterfeldt 2013). Studies related to decision aids have focused on health decisions (Fischhoff et al. 2011). For example, women using an aid regarding prenatal screening for Down Syndrome had a better understanding of information (including test results) and felt less conflict about their decisions compared to women receiving routine care (Bekker et al. 2004). Similar results have been found with aids for prostate cancer screening (Volk et al. 2003), hormone therapy (O'Connor et al. 1998), and other decisions.

Here, we extend those methods to evaluating decision aids related to climate change, focusing on three aspects of understanding essential to informed decision-making: (a) knowledge, (b) consistent preferences, and (c) active mastery. Thus, a decision aid should impart *knowledge* of decision-relevant facts, allow people to integrate those facts with their values well enough to form *consistent preferences,* and achieve the *active mastery* needed to make sound inferences related to practical decision problems. We begin by describing an aid related to coastal flooding, the Surging Seas Risk Finder. We then evaluate it for these three aspects of understanding. We also contrast responses to three time horizons (2020, 2050, 2100), showing how to assess whether different ways of framing a decision (here, through the choice of time frame) biases users' responses.

2 The risk finder decision aid

Climate Central, an independent organization studying and reporting climate change, has developed Surging Seas, a decision aid with authoritative forecasts of coastal sea level rise. Figure 1 shows its website. This aid is intended to inform a range of decision makers, from homeowners deciding whether to buy flood insurance (or move) to city planners developing recovery and resilience programs.

The website's Risk Finder tool (Fig. 2) provides information regarding potential flood area, likelihood and exposure (Strauss et al. 2012; Tebaldi et al. 2012) for specified coastal locations (e.g., Kings County (Brooklyn), NY) and time horizons (e.g., between today and 2020, between today and 2050, between today and 2100). We created a simplified version for this study (Wong-Parodi and Strauss 2014), one that lets users select a flood height, and then learn the chances of it being reached, along with the percentages of land, people, schools, homes, road miles, power plants, and sewage plants that would be inundated. We chose to focus on countywide rather than neighborhood impacts, as they are more likely to be considered when
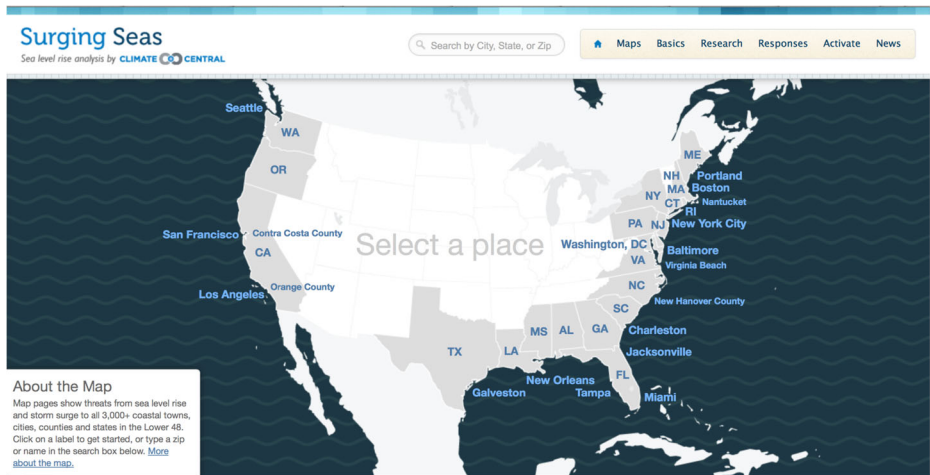
Fig 1 Overall Surging Seas website (http://sealevel.climatecentral.org, retrieved January 2014). It contains information beyond that found on Risk Finder (Fig. 2) and was not seen during the study

people are deciding to move to a new area. People likely consider more granular information when deciding exactly where to move, assuming the overall risk is tolerable. Users were randomly assigned to one of three time frames: near-term ("between today and 2020"), mid-term ("between today and 2050"), or long-term ("between today and 2100"). (As our goal was to assess what users could absorb from an aid, we did not include a no-aid control group. Given the specialized information that Risk Finder offers, without it, people are likely to have no more than general impressions.) Fig. 3 illustrates the simplified Risk Finder for Kings County (Brooklyn), NY, with a 5 ft flood height and mid-term time frame.

Designers face many choices in creating an aid, each of which might frame users' thinking in ways that biased their choices (Kahneman 2011; Lichtenstein & Slovic 2006). In some cases, basic psychological research can help designers to choose the most neutral frame. For coastal flooding, one such design choice is the time frame to present. Unfortunately, the
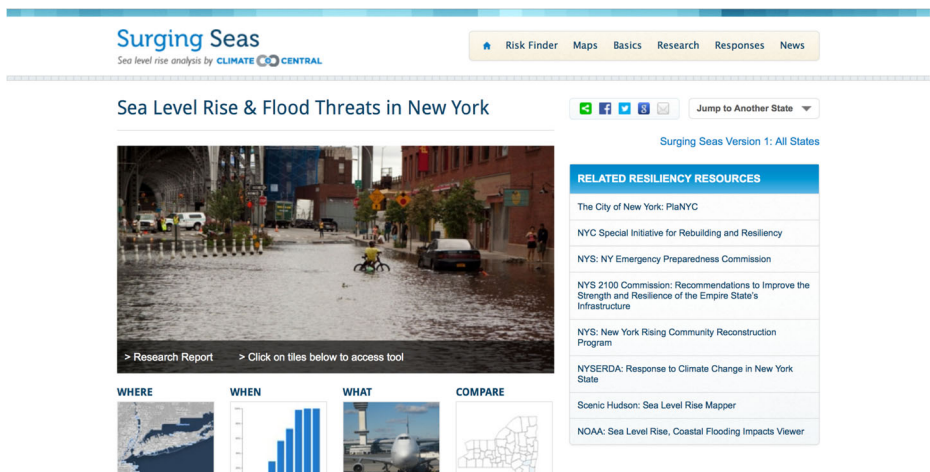

Fig 2 Risk Finder tool for New York (http://sealevel.climatecentral.org/ssrf/new-york, retrieved January 2014)
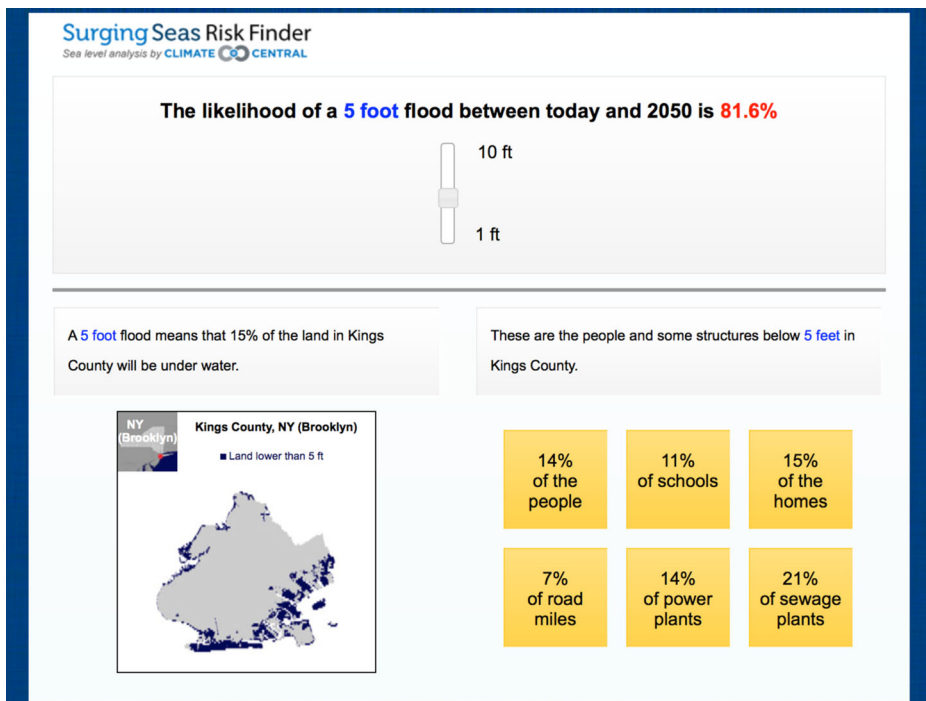
Fig 3 Simplified experimental Risk Finder tool for Kings County (Brooklyn), NY showing mid-term flood risk

research provides no clear guidance on how time horizon will affect users (Maglio et al. 2013). Some research has found greater concern about coastal flooding with shorter time horizons (Spence et al. 2012), which could prompt deeper information processing and better understanding. On the other hand, near-term risks are necessarily lower, which could reduce concern, information processing, and understanding. We apply our approach to assessing the net effects of these conflicting processes.

3 Methods

3.1 Measures

Criterion 1: **Knowledge**.

We assessed users' understanding of the facts presented by the aid with three tests:

Test 1 – pattern recognition assessed whether users can see that higher flood heights are less likely than lower ones. Participants were asked to use the slider, and then answer two open-ended questions: (a) "What happens to the flood height (ft) as you move the slider up and down?" (b) "What happens to the flood likelihood (%) as you move the slider up and down?" Participants then manipulated the slider to see the likelihood of floods of 2, 3, 4, and 9 ft, after which they indicated their answer on a 0.0 to 100.0 % slider bar.

Test 2 – land impacts asked participants to look at the map and describe, "What happens as you move the slider from 1 to 10 ft?"

Test 3 – people and structure impacts asked participants to move the slider as they looked at the Risk Finder, and then describe, "What could happen to people and structures as you move the slider from 1 to 10 ft?"

Criterion 2:   **Consistency**.

We assessed the consistency of users' preferences with two tests:

Test 4 – government policy and willingness to move compared users' agreement with three statements: (a) "Keeping what [I] learned about coastal floods in [County] in mind …I would still move with my family to [County] (if I was planning on doing that already)." (b) "Imagine that the federal, state, and local government do nothing to prepare for the risks of coastal flooding in [County]…I would still move to [County] with my family." (c) "Imagine that the federal, state, and local government do nothing to slow down global warming…I would still move to [County] with my family." The scale anchors were 1 = completely disagree and 7 = completely agree.

Test 5 – flood tolerance and willingness to move asked users, "Now, please set the highest expected flood height that you and your family would be willing to live with, at some point between today and [2020/2050/2100], before deciding to move to [County]. What height did you pick?" Using the aid, users selected a flood height between 0 feet and 10+ feet. In addition to this measure of *absolute tolerance*, we created a value of *relative tolerance* by subtracting this value from their answer to Test 7 below. (Note: Assessing users' actual willingness to move would require providing information about flood risks at specific locations and about many other topics.)

Criterion 3:   **Active Mastery**.

We assessed users' ability to make inferences based on the facts in the decision aid with two tests:

Test 6 – extrapolation above 10 ft "Now, please move the sliders below to show your estimate of how likely between today and [2020/2050/2100] it is that there will be at least one 11 ft flood in [County]." Users responded with a 0.0 to 100.0 % slider bar. Correct responses were <0.7 % for 2020, <3.8 % for 2050, and <16.8 % for 2100.

Test 7 – estimation at different time horizons: "Now, please set the flood height to the highest level that you would expect to see between today and [2020/2050/2100], if you lived in [County]. What height did you pick?" Participants selected a value between 0 feet and 10+ feet. If users can make such inferences, these estimates should increase over time.

3.2 Experimental procedures

After a brief introduction, participants were screened for age (≥18). They then selected one city/county from 14 possibilities. In order to hold information constant across conditions, all participants received values for Kings County (Brooklyn), NY.

Participants were then asked to "Imagine that you and your family are planning on moving to [their selected county and state]. You and your family want to settle down in [their selected county], with no plans to move ever again. Your family is concerned about coastal flooding. This type of flooding can happen in an area that is directly on or near the coast. They want you to find out about coastal flooding in [their selected county]." To deepen participants' engagement with this role-playing exercise, they were then asked to write out why they might want to know about coastal flooding.

Experimental conditions Participants were randomly assigned to one of the three time frames [2020/2050/2100]. They were asked to explore the Risk Finder until they had enough information to tell their family about coastal flooding. After completing the tests, participants were asked (a) if they "ever lived in a coastal county" and, if so, if they were "familiar with the coastline where [they] lived."; (b) if they had "ever vacationed along the coast" or had "ever experienced a hurricane"; and (c) demographic questions.

Participants We recruited 149 U.S. adult participants through an advertisement on Amazon's Mechanical Turk (MTurk), an online subject recruitment tool providing diverse, although non-representative samples. According to their self-reports, participants' average age was 36.1 (*SD*=13.0), with 47.7 % being female, 79.2 % identifying as White or Caucasian, 48.4 % having at least a bachelor's degree, and 42.2 % with household incomes of at least $51K. Most reported being Democrats at 51 %, with 28.2 % Independents, 16.8 % Republicans, and 4 % Other or "Prefer not to answer." Although a minority lived on the coast (37.6 %), most had vacationed there (71.8 %). Many reported having some familiarity with the coastline (34.1 %), having experienced a hurricane (44.3 %) or flood or knowing someone who had (36.9 %).

A sample this size has statistical power of 80 % to reveal differences of 0.43 ft between the time frames, with α=.01. Although results with any convenience sample should be used cautiously, experimental studies comparing MTurk with other recruitment procedures have found few differences (Crump et al. 2013; Mason and Suri 2012).

4 Results

Time Frame: We found few statistically significant differences in responses to the three time frames. As a result, we pooled responses except where there were differences. (Note: We also observed no significant correlations between self-reported flood experience and any of our dependent measures (*p*>.05)).

Criterion 1: **Knowledge**.

> *Test 1 – pattern recognition.* Most participants (95 %) successfully recognized that higher flood heights are less likely than lower ones. Most responses to the open-ended question, "What happens to the flood height (ft) as you move the slider up and down?", also indicated understanding, with answers ranging from simple ("The flood height changes") to more precise ("It increases and decreases by 1 ft intervals, between 1 and 10 ft"). Participants showed similar understanding when asked, "What happens to the flood likelihood (%) as you move the slider up and down?", with answers again ranging from simple ("The percentage goes down as the flood height increases") to precise ("It starts to drop on the way up, but not until the six foot mark, and it increases on the way back down").
>
> Most participants (96 %) correctly reported the likelihood of a flood reaching each of four designated heights (2, 3, 4, 9 ft). Simple t-tests found no significant differences between reported and displayed values (α=.01).
>
> *Test 2 – land impacts.* Most participants (95 %) recognized that water covers more land as flood height increases, with responses ranging from general ("the map shows more of the county as under water") to precise ("a 10 ft flood means that 26 % of the land in St. Tammany County will be under water").

Test 3 – people and structure impacts. Almost all (99 %) recognized that more people and structures are affected as flood height increases, again with consistent explanations.

Criterion 2: **Consistency**.

Test 4 – government policy and willingness to move. The three judgments of the highest flood height that users would tolerate (*before* deciding not to move), using a scale from 1 = completely disagree to 7 = completely agree, were strongly correlated with one another (r^2 ranging from 0.71 to 0.87; $p<.001$). The absolute level of these judgments, however, depended on the question, $F(1.78,259.51) = 24.30$, $p<.001$. Immediately after using the tool, participants were moderately willing to move despite the flood risks ($M=4.82$, $SD=1.89$). That willingness declined after being asked to imagine that local government would take no action to prepare for the risks ($M=4.13$, $SD=1.89$), $t(148) = 6.19$, $p<.001$, and that it would do nothing about global warming ($M=4.34$, $SD=1.91$), $t(148) = 4.49$, $p<.001$. The last two judgments also differed significantly, $t(148) = -2.48$, $p<.01$.

Test 5 – flood tolerance and willingness to move. As another measure of consistent preferences, participants who would tolerate a lower flood height (*before* deciding not to move) while using the tool would also be less willing to move immediately after using it ($B=.25$, $se=.05$), $t(148)=4.58$, $p<.001$, after being asked to imagine that government does nothing to prepare for flooding ($B=.24$, $se=.06$), $t(148)=4.20$, $p<.001$, and after asked to imagine that government does nothing about climate change ($B=.22$, $se=.06$), $t(148)=3.86$, $p<.001$. Arguably, risk tolerance should depend on how long people expect to be living in a location. However, using age as a surrogate for tolerance, we found no significant correlations with any of our dependent measures ($p>.05$).

Criterion 3: **Active Mastery**.

Test 6 – extrapolation above 10 ft. All participants in the 2020 and 2050 conditions saw the probability of flooding above 10 ft as lower than the probability for 10 ft, as did almost all (86 %) in the 2100 condition (for group difference, $\chi^2(1,149) = 14.54$, $p<.01$).

Test 7 – estimation at different time periods. An analysis of variance with a linear contrast revealed a significant increasing trend, $t(89.71) = 7.02$, $p<.001$, $\eta^2=.26$, with participants expecting higher flood heights for 2100 ($M=6.84$, $SD=2.38$) than for 2020 ($M=3.92$, $SD=1.74$), $p=.04$. Estimates for 2050 ($M=6.04$, $SD=1.98$) were between those for 2020 and 2100, but not significantly different from either.

5 Discussion

Many groups have developed aids intended to help inform people's decisions regarding threats posed by climate change (Bottrill 2007; Chatterton et al. 2009; Meinke and Stone 2005). We offer a practical method for evaluating such aids' contribution to three forms of usability: (a) knowledge, (b) consistency, and (c) active mastery.

Applied to the Surging Seas Risk Finder, these tests find success of all three kinds. (a) Users learned about coastal flooding risk, recognizing the relationships between the height of flooding and its probability and impacts. (b) Users who said that they would tolerate lower

flood risks were also less willing to move to the focal location, both during and after using the aid, as well as when told about government inaction. (c) Users could extrapolate the chances of a flood above 10 ft (the maximum in the aid) and the trend in probabilities over time.

In almost all respects, users responded similarly to displays with the three time horizons (2020, 2050, 2100). In contrast, Spence et al. (2012) found greater concern with shorter time horizon. One possible explanation for the difference in our results is that participants' immersion in the aid overwhelmed any framing effect induced by the time period. A second possible explanation is that in the context of more specific decisions time period had cancelling effects, with the longer period revealing greater risks but also seeming more distant. These determinants of concern and action bear further attention, as do others, such as uncertainty about the estimates (Fischhoff and Davis 2014). In terms of usability, though, the three time horizons were equally effective and showed no evidence of framing effects.

Our method can be adapted for other climate change decision aids. The knowledge measures can capture the information that aid designers would most like their audiences to grasp (e.g., the probability, nature, and consequences of droughts of varying severity). The consistency measures can reflect logically related preferences related to the aid's topic (e.g., tolerance for drought severity under a range of contexts). Active mastery measures can reflect relevant factual inferences (e.g., the impacts of drought of varying temporal or geographic scope). Creating such measures requires the designers of an aid to make their aims explicit, a task that could improve the focus of their work. Once created, the measures can be administered to online samples like the one used here or members of specific target groups. Such evaluation costs little relative to the investments in aids and the stakes riding on them. Arguably, it should be the standard for any communication (Fischhoff 2013). These present findings revealed Surging Seas to be a decision aid that helps lay users understand coastal flooding risks. Subsequent to our research, it was featured at the Obama Administration's rollout of http://www.data.gov/climate/.

Author attributions All authors contributed to the conception and the design of the tests and the experiment used within the work presented. G.W.P. ran the experiment and analyzed the data. G.W.P. and B.F. led in the writing of the paper, with input from B.S.

Competing financial interests The authors declare no competing financial interests.

References

Bekker HL, Hewison J, Thornton JG (2004) Applying decision analysis to facilitate informed decision making about prenatal diagnosis for Down syndrome: a randomised controlled trial. Prenat Diagn 24(4):265–275

Bottrill, C (2007) Internet-based carbon tools for behaviour change. University of Oxford Environmental Change Institute, Oxford (available online at http://www.eci.ox.ac.uk/research/energy/downloads/botrill07-calculators.pdf)

Braddock CH III, Edwards KA, Hasenberg NM, Laidley TL, Levinson W (1999) Informed decision making in outpatient practice. JAMA 282(24):2313–2320

Chatterton TJ, Coulter A, Musselwhite C, Lyons G, Clegg S (2009) Understanding how transport choices are affected by the environment and health: views expressed in a study on the use of carbon calculators. Public Health 123(1):e45–e49

Crump MJ, McDonnell JV, Gureckis TM (2013) Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 8(3):e57410

de Best-Waldhober M, Daamen D, Faaij A (2009) Informed and uninformed public opinions on CO_2 capture and storage technologies in the Netherlands. Int J Greenh Gas Control 3(3):322–332

Dietz T, Gardner GT, Gilligan J, Stern PC, Vandenbergh MP (2009) Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. PNAS 106(44):18452–18456

Fischhoff B (2013) The sciences of science communication. PNAS 110(Supplement 3):14033–14039

Fischhoff, B, Davis, AL (2014) Communicating scientific uncertainty. PNAS

Fischhoff B, Brewer N, Downs JS (eds) (2011) Communicating risks and benefits: an evidence-based users' guide. Food and Drug Administration, Washington

Füssel HM (2007) Adaptation planning for climate change: concepts, assessment approaches, and key lessons. Sustain Sci 2(2):265–275

Kahneman D (2011) Thinking, fast and slow. Farrar Giroux & Strauss, New York

Lichtenstein SL, Slovic P (eds) (2006) The construction of preferences. Cambridge University Press, New York

Lusardi, A (2008) Financial Literacy: An Essential Tool for Informed Consumer Choice? (No. w14084). National Bureau of Economic Research

Maglio SJ, Trope Y, Liberman N (2013) The common currency of psychological distance. Psychol Sci 22(4): 278–282

Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 44(1):1–23

Matthies M, Giupponi C, Ostendorf B (2007) Environmental decision support systems: Current issues, methods and tools. Environ Model Softw 22(2):123–127

Meinke H, Stone RC (2005) Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Clim Chang 70(1–2):221–253

Nicholson-Cole SA (2005) Representing climate change futures: a critique on the use of images for visual communication. Comput Environ Urban Syst 29(3):255–273

Noar SM, Benac CN, Harris MS (2007) Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 133(4):673–693

O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, McPherson R, Bunn H, Graham I, Drake E (1998) A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. Patient Educ Couns 33(3):267–279

Olson GM, Olson JS (2003) Human-computer interaction: psychological aspects of the human use of computing. Ann Rev Psychol 54(1):491–516

Padgett JP, Steinemann AC, Clarke JH, Vandenbergh MPA (2008) Comparison of carbon calculators. Environ Impact Assess Rev 28(2):106–115

Salvendy G (2012) Handbook of human factors and ergonomics. Wiley, Indiana

Simao A, Densham PJ, Haklay M (2009) Web-based GIS for collaborative planning and public participation: an application to the strategic planning of wind farm sites. J Environ Manag 90(6):2027–2040

Spence A, Poortinga W, Pidgeon N (2012) The psychological distance of climate change. Risk Anal 32(6):957–972

Strauss BH, Ziemlinski R, Weiss JL, Overpeck JT (2012) Tidally adjusted estimates of topographic vulnerability to sea level rise and flooding for the contiguous United States. Environ Res Lett 7(1):014033

Szwajcer EM, Hiddink GJ, Koelen MA, van Woerkum CM (2009) Written nutrition communication in midwifery practice: what purpose does it serve? Midwifery 25(5):509–517

Tebaldi C, Strauss BH, Zervas CE (2012) Modelling sea level rise impacts on storm surges along US coasts. Environ Res Lett 7(1):014032

Ver Steegh N (2002) Yes, no, and maybe: Informed decision making about divorce mediation in the presence of domestic violence. Wm Mary J Women L 9:145

Volk RJ, Spann SJ, Cass AR, Hawley ST (2003) Patient education for informed decision making about prostate cancer screening: a randomized controlled trial with 1-year follow-up. Ann Fam Med 1(1):22–28

von Winterfeldt D (2013) Bridging the gap between science and decision making. PNAS 110(Supplement 3): 14055–14061

Wong-Parodi, G, Strauss, B (2014) Team science for science communication, PNAS (in press)