

Negative learning

Michael Oppenheimer · Brian C. O'Neill ·
Mort Webster

Received: 12 January 2007 / Accepted: 7 February 2008 / Published online: 6 May 2008
© The Author(s) 2008

Abstract New technical information may lead to scientific beliefs that diverge over time from the *a posteriori* right answer. We call this phenomenon, which is particularly problematic in the global change arena, negative learning. Negative learning may have affected policy in important cases, including stratospheric ozone depletion, dynamics of the West Antarctic ice sheet, and population and energy projections. We simulate negative learning in the context of climate change with a formal model that embeds the concept within the Bayesian framework, illustrating that it may lead to errant decisions and large welfare losses to society. Based on these cases, we suggest approaches to scientific assessment and decision making that could mitigate the problem. Application of the tools of science history to the study of learning in global change, including critical examination of the assessment process to understand how judgments are made, could provide important insights on how to improve the flow of information to policy makers.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-008-9405-1) contains supplementary material, which is available to authorized users (ESM).

M. Oppenheimer
Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA

M. Oppenheimer (✉)
Department of Geosciences, Princeton University, Princeton, NJ, USA
e-mail: omichael@Princeton.edu

B. C. O'Neill
International Institute for Applied Systems Analysis, Laxenburg, Austria

M. Webster
MIT Joint Program on the Science and Policy of Global Change, Massachusetts Institute of Technology, Cambridge, MA, USA

Present address:

B. C. O'Neill
Institute for the Study of Society and Environment, National Center for Atmospheric Research,
Boulder, CO, USA

1 Introduction

An unstated assumption of nearly every quantitative analysis of decision making on environmental problems (and many other problems as well) is that more scientific research can be equated with (1) greater yields from prospecting, that is, awareness of potential problems that will ultimately require attention by policy makers, and (2) truer beliefs about the outcomes of problems already recognized, providing a superior basis for crafting solutions. In this article, we challenge the second assumption by examining the consequences of an alternate hypothesis particularly relevant to global change: As learning occurs, expert belief often diverges from the true outcome for time periods that are sufficiently long to affect the policy process in important ways. We refer to the phenomenon of new information leading to increasing or sustained divergence of current belief from the true outcome as *negative learning* (The rationale for this nomenclature is explained below).

There is a substantial literature that addresses the dynamics of expert learning in the context of decision-making. The focus of most of this literature has been on the *rate* of scientific learning and ways to incorporate it into formal decision models (Arrow and Fisher 1974; Kolstad 1996). However, the issue of the *direction* of learning (essentially, the dynamical effect of erroneous judgment) has, with rare exception (Small and Fischbeck 1999; Lange 2003; Keller and McInerney 2007), not been explicitly considered. This is surprising given the numerous salient cases where natural sciences (and social sciences) applied in a policy context have erred in ways that affected people when policy makers and practitioners implemented prescriptions based on faulty findings. Examples include sharp changes in medical recommendations with respect to ulcers (Abbott 2005) and errant projections of energy prices (Huntington 1994). Both modelers and decision makers have paid little attention to the consequences.

There has also been substantial research over several decades into the causes of erroneous judgments about uncertainty. For example, overconfidence is one well-known cognitive bias in human judgments about uncertainty that can lead to serious errors (Tversky and Kahneman 1994; Morgan and Henrion 1990). Methods have been proposed for improving uncertainty estimation within a variety of formal contexts for expert elicitation of judgment. However, studies of overconfidence and other cognitive biases generally have been framed as static issues. Even when overconfidence has been shown to persist over time (Henrion and Fischhoff 1986), the focus has not been on the consequences of its persistence in circumstances related to decision making on policy.

In other words, there have been studies of dynamics of learning in the policy decision context with little emphasis on error, and studies of error with little emphasis on dynamics. In this article, we formalize the concept of negative learning to unify these distinct domains. Negative learning is an umbrella concept, a description of an aspect of the large-scale, long-term dynamics of learning about technical information that results from the interaction of many well-known phenomena¹. Overconfidence is one likely cause of negative learning, but it is by no means the only one. The use of expert elicitation to assess knowledge and uncertainty among limited groups of experts sometimes involves reflexive revision of judgments that is known to consolidate beliefs, revealing some of the group interactions

¹ Additional context for introduction, definitions and conclusions, discussion of analogs of negative learning in other fields, and details of case histories, policy responses, and modeling are available as Electronic Supplementary Material (ESM).

that can lead to negative learning (Morgan and Henrion 1990, section 7.7). Reviews of large-scale assessments, such as those of the Intergovernmental Panel on Climate Change (IPCC), have taken note of such phenomena and their potential consequences (Funtowicz and Ravetz 1992; Patt 1999, 2007; Pielke 2001). There is also copious literature on structural uncertainty in models (Draper 1995), one cause of negative learning discussed extensively here.

Our interest is in science where it interacts with the policy process: we describe a phenomenon more particular than the general tendency of science to learn from mistakes. Learning timescales considered here are long compared to the timescale over which policy decisions are made. The mismatch of timescales arises from various lags and irreversibilities that are pervasive in global change (but absent from some decision arenas, e.g., option markets). We analyze learning about the “conventional wisdom” or “scientific consensus” on which policy is often based, rather than the beliefs of individual scientists, whose awareness of flawed assumptions may precede that of the wider community. Our study suggests that twenty years’ experience with large international assessments has failed to solve, and in some respects even aggravated, the problem posed by negative learning for policy makers.

To clarify our argument, we focus on changes in the estimated value of a particular modeled or observed quantity that was commonly used in official assessments or in the projection of a particular policy-relevant outcome. Where assessments are not available, we draw on information from formalized expert elicitations and research articles and reviews that are widely cited (see ESM). While we recognize that such an approach does not do full justice to the complexity of learning, it is also true that policy makers do not pay attention to the full complexity either. The scientific input to decisions is often flattened to a low-dimensional representation, e.g., rate of emissions growth, climate sensitivity, the range of projected warming or sea level rise, projected ozone depletion (see Crutzen and Oppenheimer 2008), or projected energy demand.

Discussion of the direction of learning inevitably leads to the question, “direction toward (or away from) what?”, e.g., a true state of belief. The question of whether a true state of belief exists has been addressed extensively but not conclusively for centuries (Kitcher 1993; see also ESM), and we cannot hope to resolve it here. Furthermore, there is ample literature on the question of whether progress toward truth in science exists at all (Kuhn 1962). Part of that discourse revolves around the issue of the uses of physical models and their relation to true belief (Oreskes et al. 1994). For current purposes, we will assume that a functionally or provisionally true (“good enough”) state exists and is eventually attained. This assumption allows us to speak of learning as having a direction. We also do not attempt to develop guidelines for determining, during the period before provisional truth is known, whether negative learning is occurring. Instead, we point to a research agenda following from our study that may make such judgments feasible in the future.

Given the recognition that the phenomenon of negative learning occurs, our objectives in this paper are to define negative learning, show that negative learning is prevalent in global change problems, and demonstrate that the consequences of negative learning can be modeled and may be significant in particular cases. To achieve these objectives, this paper makes the following five contributions:

1. The recognition of negative learning raises important definitional and conceptual questions, which we address in Section 2.
2. Despite the literature above on particular mechanisms, the issue of the dynamics of learning remains peripheral to the conduct of science, scientific assessments, and

- decision making. In Section 3, we use four prominent case histories to demonstrate that negative learning occurs for reasons that are common in global change and that it occurs about important aspects of issues, not just the details, and thus is potentially highly problematic for policy makers.
3. In Section 4, we modify a formal decision model for the climate problem to illustrate how negative learning could occur even given correct application of Bayes' Theorem, and be sustained through many decision periods. We also demonstrate that large potential social welfare losses can occur as a result. We employ a stochastic version of the DICE model for this illustration, which, despite the criticisms of this model, is of appropriate and sufficient complexity for the objective here.
 4. In the discussion, we draw on experience with global change assessments to date to recommend improvements in their structure that may minimize the likelihood and impact of negative learning.
 5. The recognition of negative learning raises questions that require further research. We argue, as is implicit in at least one previous study (Shlyakhter 1994), that science history has something important to offer to science, in our case the science of global change. Specifically, studies of why science and science assessments have gone awry can provide important insights about the pitfalls to be avoided in the future. Otherwise, the same misjudgments that led to negative learning in the past can be expected to be repeated with no lesser frequency. We also assert that new research is needed into formal decision models to elaborate the causes of negative learning and suggest whether any policy prescriptions could reduce its influence on decisions.

2 Definitions and context

Learning is the change in any aspect of the uncertainty of an outcome occurring as a result of theory development, modeling, observation, or experiment. *Uncertainty* is a “multifaceted concept” (Morgan and Henrion 1990); here we use the term to refer to imprecision in the characterization of any outcome. An *outcome* is any quantity, process, or structure (e.g., model) that may be of interest to policy makers.

Our definitions of both uncertainty and learning are intended to be broad. Characterizations of the uncertainty in an outcome can take a range of forms, from purely qualitative to fully probabilistic, with a variety of approaches occupying a middle ground, including those in which the probabilities themselves are imprecise. Imprecision in probability assignments can reflect conditions such as a weak information basis or the existence of alternative opinions on a controversial issue, and it can be incorporated into quantitative analyses (e.g., Hall et al. 2007) or be reflected in qualitative characterizations of the confidence placed on probability distributions (Moss and Schneider 2000).

Learning is a change in any of these aspects of uncertainty. Typical changes include updated estimates of the mean or variance of a probability distribution. But others could include a change in the confidence attached to the uncertainty characterization, or a change in the information basis for the uncertainty.

Negative learning is a *decrease* or sustained divergence in the correspondence between the true outcome and the uncertainty characterization (or belief) over time. In the case of probabilistic characterizations, this will occur when the probability density near the true outcome decreases over time, or when the distance increases between the true outcome and the outcome considered most likely. We choose the name “negative” to reflect this move

away from the true outcome. Negative learning can occur as a result of a variety of changes to probability distributions, including an increase in the variance of a distribution that is centered around the true outcome, a decrease in the variance of a distribution that is centered around an estimate sufficiently distant from the true outcome, or a shift in a distribution (without a change in variance) away from the true outcome.

Progressive learning is defined as an *increase* in the correspondence between the true outcome and the uncertainty characterization over time. Following Kolstad (1996), we define *learning history* as the path of uncertainty from an initial state of belief to the current state. A learning history may be characterized as negative learning only *a posteriori*. Until the true (or provisionally true) outcome is known (see ESM) all that can be detected is a change in correspondence between current belief and any of the possible outcomes.

3 Cases of negative learning

3.1 Ozone depletion

Scientific beliefs with regard to the causes of ozone depletion by halocarbon compounds, as well as the international and domestic policy framework based upon those beliefs, have been stable for over a decade, long enough for extensive retrospective analysis (Benedick 1998; Parson 2003; Christie 2001; Crutzen and Oppenheimer 2008; also see “[Electronic supplementary material](#)” for additional discussion of policy developments and references). Belief on key scientific aspects, e.g., the importance of heterogeneous chemical reactions, may be regarded as provisionally true even as science will continue to develop in several areas (e.g., with respect to the question of the effect of greenhouse gas concentrations on the stratosphere). We explore the history of learning about total column ozone depletion at a hypothetical steady state (in the distant future) for a constant halocarbon emissions level, a synoptic, policy-relevant output of photochemical models (NRC 1984; WMO 1986).

Figure 1 indicates a history of learning about this outcome from one-dimensional model calculations based on constant emissions at 1974 levels (uncertainty $\{1\sigma\}$ due to error in photochemical rates was estimated to be comparable to total depletion (NRC 1984; WMO 1986)). Projections of depletion were made from 1974 onward. The gradual inclusion of an increasingly complex gas phase chemical kinetic scheme in models, and greater precision in the determination of rate constants in the laboratory (Crutzen and Oppenheimer 2008; Parson 2003; NRC 1984; WMO 1986), led to revised estimates of depletion that eventually settled at relatively low values compared to the earliest projections. Knowledge of stratospheric chemistry was then transformed by the discovery of large, seasonal Antarctic depletion (the “ozone hole”) in 1985, the validation in 1987 of its origin in halogen photochemistry, and subsequent identification of depletion at the mid-latitudes and in the Arctic (WMO 1988, 1991; Rowland 1989). Various heterogeneous chemical reactions, discounted by most researchers years before and absent from nearly all model simulations (WMO 1986), were shown to be the missing photochemical processes required to explain observed depletion. Including such reactions in models substantially increased projected depletion for a given level of halocarbon emissions.

Learning, as measured by changes in modeled total depletion, was clearly negative for about a decade, during a critical period of policy formation. Negative learning resulted (counter intuitively) from progressive learning about one part of a complex system. The failure to account for heterogeneous reactions may be regarded as either the result of flawed models or inaccurate parameter determination: Early measurements on rates of some of

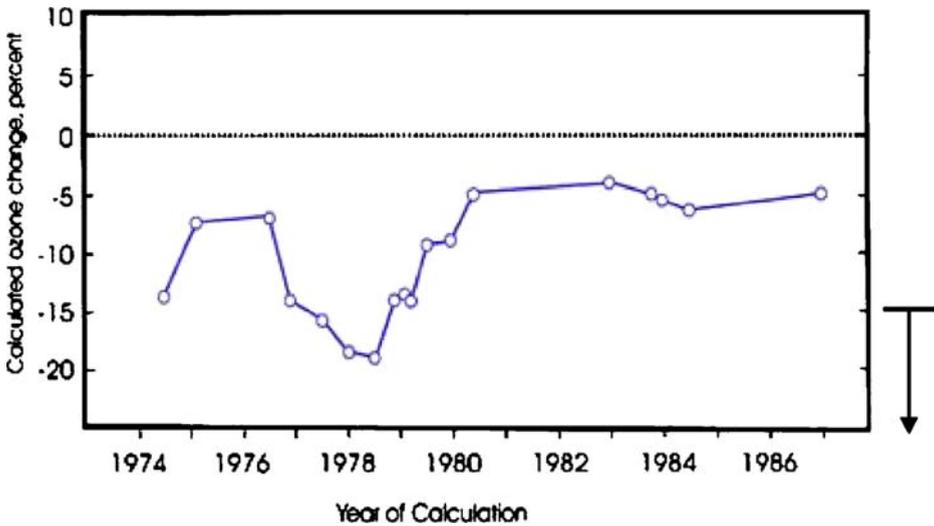


Fig. 1 History of projected depletion of total ozone from a one-dimensional model from Parson (2003), based on analysis by Don Wuebbles presented in NRC (1984) and WMO (1986) and references therein. Projected depletion shown at steady-state for constant CFC emissions at 1974 levels. In order to indicate where current belief lies, we have modified the figure to indicate an estimate ($\overline{\downarrow}$) of depletion that might have occurred at mid-latitude for constant 1974 emissions obtained by linear extrapolation of recent observations of depletion using known stratospheric chlorine levels (WMO 2006). Newer models that include heterogeneous chemistry would likely produce greater depletion than the horizontal bar due to nonlinear behavior at high chlorine levels, as indicated by *arrow*, but simulations are not available for such levels

these reactions were misleading in suggesting that they were insignificant while other reactions were thought to be potentially important but difficult to measure in the laboratory (WMO 1986). The potential implications of heterogeneous chemistry were of concern to some scientists (Rowland 1989) but largely discounted in assessments. In addition, the availability of stratospheric surfaces where such reactions could occur was not recognized.

Both the variations and the eventual decrease to low values for projected depletion (possibly in conjunction with the relative stability of the low values) may have reinforced the inclination in governments and industry to defer or oppose comprehensive regulation of halocarbon production after use of chlorofluorocarbons as propellants in aerosol spray cans was limited or banned in many jurisdictions beginning in 1977 (see Crutzen and Oppenheimer 2008, for a fuller account of views on the interaction of scientific and political histories of ozone depletion; also see ESM). Economic simulations indicate that relatively modest cost but large public health benefits would have accrued in the counterfactual case of imposing comprehensive regulation after 1975 (Ha-Duong et al. 2003). Negative learning was a component of a complex set of factors and judgments (see ESM) that led to a policy that in retrospect was sub-optimal.

3.2 West Antarctic ice sheet (WAIS)

Deglaciation of the West Antarctica ice sheet (WAIS) would raise global mean sea level by about 5 m (Vaughan 2008). In contrast to ozone depletion, learning about WAIS likely

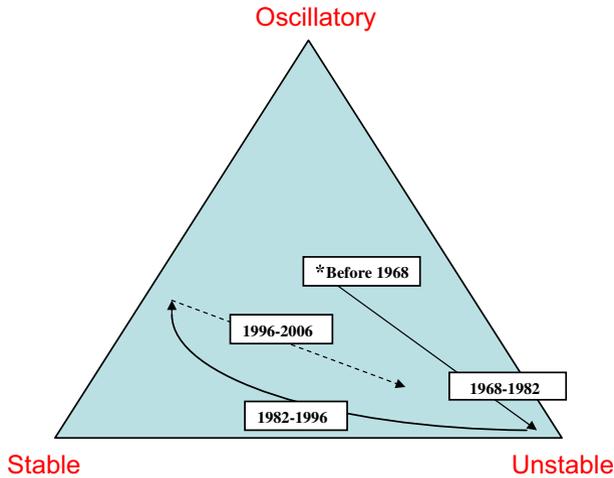


Fig. 2 Learning history for WAIS illustrated schematically by an information structure (Kolstad 1996; Jones and Ostroy 1984) based on three possible dynamic behaviors in a warming world. The stable and unstable ice sheets would have lifetimes of a few millennia and a few centuries, respectively. The oscillatory ice sheet would exhibit discharge of large fractions of its mass on random timescales between these two extremes, followed by some degree of re-growth. *Arrows* indicate changes over successive periods of learning beginning with a state of ignorance (*asterisk*) before 1968. The *final arrow (broken line)* is speculative because a community consensus on the significance of recent findings has yet to emerge. Proximity to each vertex indicates qualitatively the support for each view in expert opinion on the behavior of WAIS. The various assessments do not use identical assumptions about amount of warming attained in a given future year

remains very incomplete; its ultimate influence on policy is unknown, and only limited analysis within a science history or policy framework exists (see ESM). However, the question of WAIS' stability in a warming world has received extensive treatment in the literature (Vaughan 2008; Oppenheimer 1998; Oppenheimer and Alley 2004; Overpeck et al. 2006; Vaughan and Spouge 2002) which in turn has been reviewed in multiple assessments (Patt 1999) including those by the US National Research Council (NRC) and the Intergovernmental Panel on Climate Change (IPCC), and two expert elicitations (Vaughan and Spouge 2002; Lenton et al. 2008) spanning 30 years (see ESM).

In brief (see Oppenheimer 1998; Vaughan and Spouge 2002 for details), early theories that drew on paleoclimate analogs and simple models of parts of an ice sheet were developed in the wake of the International Geophysical Year. These assumed buttressing of grounded ice by floating ice shelves that were supposed to disintegrate upon warming, leading to instability and fast ice loss (on decadal timescales). Consideration of processes limiting the rate of iceberg formation extended the lifetime following loss of buttressing to at least 200 years. Expanded, but still limited, observations of ice streams and the advent of numerical whole-ice-sheet models, detailed analysis of the grounding line that separates land-based from floating ice, and general circulation models projecting slow southern circumpolar warming, shifted the weight of opinion toward the view that WAIS would be stable even absent ice shelves, and that deglaciation would most likely occur on timescales of a millennium or more, if at all. Nevertheless, a wide range of beliefs remained (Vaughan and Spouge 2002), including the possibility that the ice sheet could exhibit large oscillations in its volume independent of human forcing (MacAyeal 1992). Recent observational and paleoclimate studies, including critically important satellite radar and

airborne laser altimetry, combined with accelerating changes in the ice sheet, again point to the possibility of instability triggered by anthropogenic warming on a multi-century timescale (Alley et al. 2005; Hansen 2005; Overpeck et al. 2006). Models cannot yet explain the recent observations; both inadequate representation of dynamical processes and numerical deficiencies are likely to blame (Little et al. 2007).

The learning history for WAIS is summarized in Fig. 2 within a star-shaped information structure (Jones and Ostroy 1984; Kolstad 1996) using WAIS' stability as the policy-relevant characteristic (one that can be expressed as a parameter, the lifetime) and allowing three possible outcomes. The figure illustrates schematically how belief about the ice-sheet wandered relative to these states. Flawed models combined with a relatively low spatial and temporal density of observations to produce negative learning. Since the true outcome remains unknown it is not clear which branch of the diagram illustrates negative learning.

3.3 Population and energy projections

Projections of demographic change and energy system characteristics have had significant influence over public and private decision-making (see *ESM*). Although the history of projection has been characterized by progressive learning in many instances, such projections have often gone awry, sometimes spectacularly so. The list of contributing causes is long; we focus on how breaks in socio-economic trends can lead not just to a single poor forecast but to negative learning, i.e., projections that become less accurate over some significant period of time. See the “[Electronic supplementary material](#)” for a discussion of the similarities and differences between negative learning in socio-economic and physical systems.

The demographic transition is the most fundamental population-related structural change countries experience, involving a shift from high to low levels of fertility and mortality thought to be driven by a complex mix of economic, social, and policy factors (Bongaarts and Bulatao 2000; see also “[Electronic supplementary material](#)”). This transition, anticipated by demographic theory, has occurred at different times and rates in different parts of the world, but experience in forecasting fertility has been similar: a tendency to underestimate how soon and how fast the transition would occur. Projections tend to be relatively accurate before the transition, when the total fertility rate is 6 births per woman or higher, less accurate during the transition, and more accurate again in the post-transition phase when fertility is between 2 and 2.5 births per woman (Fig. 3). The period of negative learning during the transition can last for several decades, sufficient time to substantially affect long-term projections of population size and age structure (O'Neill et al. 1999). Errors persist even in the face of accumulating observations, an example of “*assumption drag*” (Keilman 1999).

Figure 3 also indicates that if a second transition (van de Kaa 1987) to fertility below the replacement level of about two births per woman occurs, errors grow again as forecasts fail to anticipate this additional structural change. The assumption that fertility would stabilize at replacement level was nearly ubiquitous in projections for decades and has been revised only recently (Lutz et al. 2001; UN Department of Economic and Social Affairs, Population Division 2005).

Unforeseen events (Schneider et al. 1998) have also initiated periods of negative learning in demography as the significance of trend breaks were both over- or underestimated and corrections made only slowly. For example, during the post-World War II baby boom in the U.S., fertility increased from less than 2.5 to above 3.5 births per woman over the period 1940–1960. U.S. Census Bureau projections consistently anticipated a

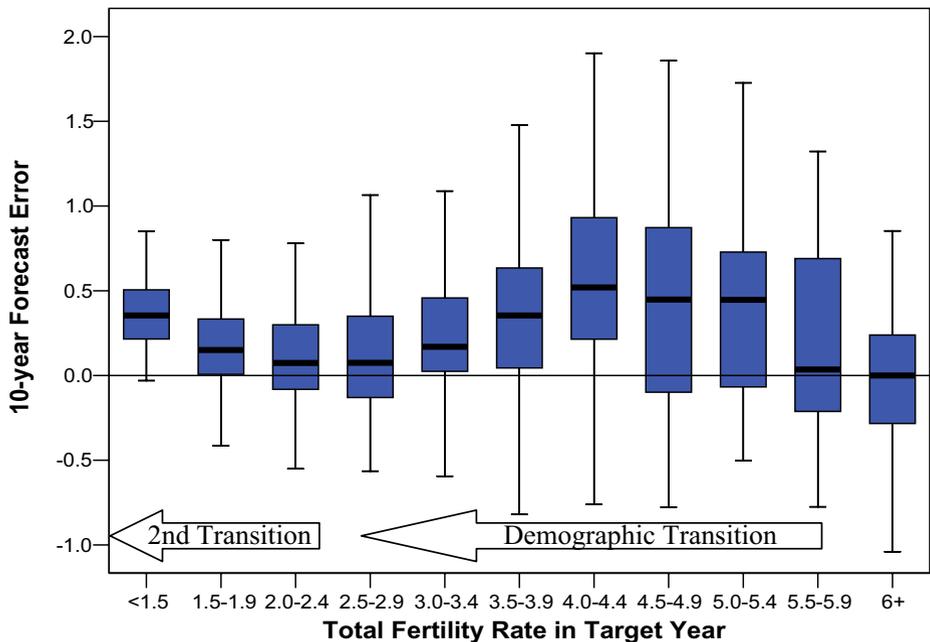


Fig. 3 Distributions of errors in 10-year forecasts of total fertility rates (TFR, births per woman) at the country level versus actual TFR in the year forecasted. Fertility in individual countries typically declines over time as countries pass through the demographic transitions described in the text. Actual fertility for all countries and time periods is taken from the most recent UN estimates (UN Department of Economic and Social Affairs, Population Division 2005). Forecasts are the medium variants from 11 global UN population projections with base years between 1965 and 1995 ($N=1,319$) and have been corrected for errors in base year data (see “[Electronic supplementary material](#)” for details on data and methodology). Boxes show medians and interquartile ranges, whiskers extend to a maximum of $1.5 \times$ the interquartile range. Forty outliers have been excluded from the plot

continuation of high fertility, when in fact fertility fell back to pre-war levels by the early 1970s (Lee et al. 2000).

In energy systems, the supply crisis involving OPEC in the early 1970s led to negative learning as projections of U.S. total energy consumption and electricity generation made for nearly a decade afterward failed to anticipate a potentially permanent shift toward greater energy efficiency of the economy (e.g., Craig et al. 2002; see also “[Electronic supplementary material](#)”). As a result, they were almost uniformly too high. The unanticipated 1979–1980 spike in world oil prices led major energy forecasting activities to predict that high prices would be a persistent feature of the market when in fact prices fell by over 50% by 1986 (Huntington 1994; see also “[Electronic supplementary material](#)”). Similarly, the development of nuclear power for electricity generation beginning in the 1950s led to excessive optimism over its potential and inaccurate projections that persisted for decades (see [ESM](#)).

4 Modeling negative learning about climate sensitivity

We complement the above case studies with a demonstration of how negative learning could occur about climate sensitivity, a key uncertainty in projecting future climate change.

Negative learning can result from any of several different problems, including structural error in models, error in a parameter not subjected to learning, and misspecification of the error structure. We illustrate the first possibility by conducting a model simulation of a hypothetical climate system that includes an “unknown” feedback on radiative forcing, i.e., one not incorporated in the model. This illustration is exploratory and not posed to argue that the particular feedback exists. The main point of this illustration is that given a model with structural error, correctly updating priors using observations according to Bayes’ Theorem will result in flawed projections, and that these erroneous projections can have significant costs.

Equilibrium climate sensitivity, defined as the equilibrium temperature change from a doubling of carbon dioxide concentrations, captures the response of geophysical feedbacks on temperature that reduce or amplify the effect of the change in radiative forcing. Previous studies have discussed the evolution of estimates of climate sensitivity (Somerville et al. 2007), examined how long it might take to reduce the uncertainty (Kelly and Kolstad 1999; Webster et al. 2008), and what policies should be undertaken in the near-term (Yohe et al. 2004; Kolstad 1996; Webster 2002), assuming that the true value is revealed after several decades. Viewed retrospectively, the learning histories in each case were explicitly or implicitly progressive. Here we extend the model to include three multi-decade periods of learning. In the period 2000–2040, decisions are made under a prior for climate sensitivity developed (Andronova and Schlesinger 2001) and used (Yohe et al. 2004) previously. In 2040, the prior is updated based on observed decadal temperature changes over that period using Bayes’ Theorem to obtain a new posterior distribution for climate sensitivity. Decisions in the second period, 2040–2100, are made under the posterior distribution. In 2100, the true value of climate sensitivity becomes known with certainty.²

One critical assumption in most Bayesian analyses is that the underlying model – and therefore the likelihood function, which plays a key role in updating the prior – is correct. However, if the likelihood-generating model omits a structural feature (Draper 1995; Small and Fischbeck 1999), attempts to learn about an uncertain parameter are likely to narrow in on an incorrect value. Projections based on posteriors derived from an incorrect model will be subject to both model error and error in the estimated parameter, with synoptic outcomes dependent on the balance of the two effects.

We simulate negative learning, its impact on policy decisions and its cost implications using a modified version of the DICE model (Nordhaus and Boyer 2000). DICE is a Ramsey growth model augmented with simple representations of the carbon cycle, radiative forcing, and the heat balance of the climate system. It has been used in numerous studies of climate policy (e.g., Yohe et al. 2004; Popp 2004; Keller et al. 2004). The DICE model has been criticized on many grounds, for example that a single decision-maker model is not a useful assumption (Morgan et al. 1999), that standard discounting approaches are inappropriate for long time horizons (Frederick et al. 2002; Schelling 1994), that distributional consequences are obscured, and that assumed climate damages are too low and its optimal mitigation path is too lax (Kaufmann 1997; Joos et al. 1999; Fussel 2007). We use the DICE model to demonstrate negative learning not because we believe it renders an accurate projection of the future in detail, but simply as a laboratory for demonstrating how negative learning occurs in the presence of structural uncertainty in spite of the correct

² This experimental design is based on a current exercise in the uncertainty subgroup of the Energy Modeling Forum 22.

application of Bayes' Theorem. The stochastic optimization methods employed require a relatively simple model structure to be computationally feasible. The illustration of negative learning and its impacts is robust to the above critiques of DICE, and we argue that the impacts of negative learning would be at least as serious in the real world as in the model.

We use a stochastic version of DICE (Yohe et al. 2004), defining different possible outcomes (or states-of-the-world, SOWs) by different climate sensitivities, with decisions made to maximize the expectation of utility across SOWs. The prior and posterior distributions for climate sensitivity are approximated with discrete four-valued distributions (Table 1) as in Yohe et al. (2004). Depending on which climate sensitivity is true, different temperature changes will be observed over 2000–2040, which are also a function of emissions and other sources of error and variability. For each possible set of observations, there is a corresponding revised posterior for climate sensitivity. Four possible values of climate sensitivities and four possible observations yield a total of 16 SOWs. One difference from Yohe et al. (2004) is that we maximize discounted utility, assuming a social rate of time preference of 3% per year in 1995, declining to 2.3% per year by 2100 (see “[Electronic supplementary material](#)”) instead of a social rate of time preference of zero.

For this example, we assume that the perceived model is incorrect due to an additional feedback on radiative forcing that is unknown to the observer/decision-maker. As in Small and Fischbeck (1999), we modify the original equation in DICE for radiative forcing $F(t)$ as follows:

$$\text{If } \Delta T(t) > \beta, F^*(t) = F(t) + \alpha F^*(t-1). \quad (1)$$

When temperature change exceeds some threshold β , the true radiative forcing $F^*(t)$ differs from current belief about forcing by some fraction of the previous period's forcing. This model is meant to capture feedback resulting from, for example, an unknown aerosol-cloud interaction. We explore the effects of an unknown positive feedback ($\alpha=0.3$) and also of an unknown negative feedback ($\alpha=-0.5$), which result in an equilibrium temperature change from CO₂ doubling of +40% and -30%, respectively. We assume a threshold temperature of $\beta=1.0$ degree above pre-industrial, and phase in the feedback using a logistic form over three decades.

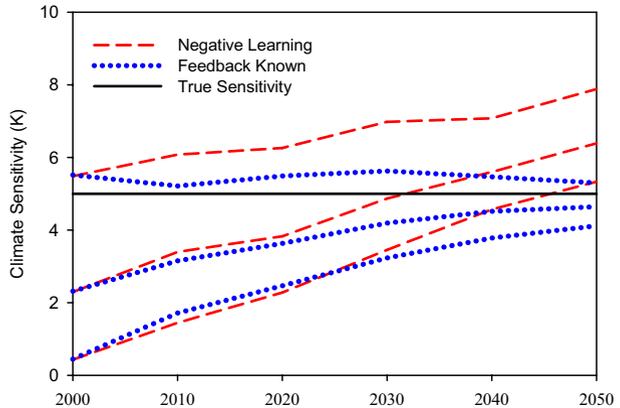
Bayesian learning is calculated using Monte Carlo integration, similar to Kelly and Kolstad (1999). Observed temperatures are generated from the true model, DICE modified with Eq. 1. We calculate posteriors for climate sensitivity twice for each observation: once using the original DICE to simulate negative learning, and again using the modified version that produced the observations. More details on the methods used can be found in the “[Electronic supplementary material](#)”.

The results for a positive feedback on radiative forcing indicate that the observed temperature changes over 2000–2040 are interpreted as resulting from a climate sensitivity significantly higher than the true one (Fig. 4). Nonetheless, the nonlinear feedback dominates overestimated sensitivity so that the posteriors lead to projections of temperature change with

Table 1 Prior distribution for climate sensitivity

Climate sensitivity	Probability	Cumulative probability
1.5	0.25	0.25
3.0	0.45	0.7
5.0	0.15	0.85
8.0	0.15	1.0

Fig. 4 Median and 90% bounds on the updated posterior distribution of climate sensitivity. Case shown is where the true climate sensitivity is 5.0 K. Bayesian updating occurs each decade, using observations from the true model with the additional feedback. For simplicity, DICE model uses prior until 2040, and posterior from 2040 update for 2040–2100 decisions



95% intervals that lie entirely below the actual warming (Fig. 5). The results for a negative feedback on radiative forcing (not shown) are analogous in the opposite direction.

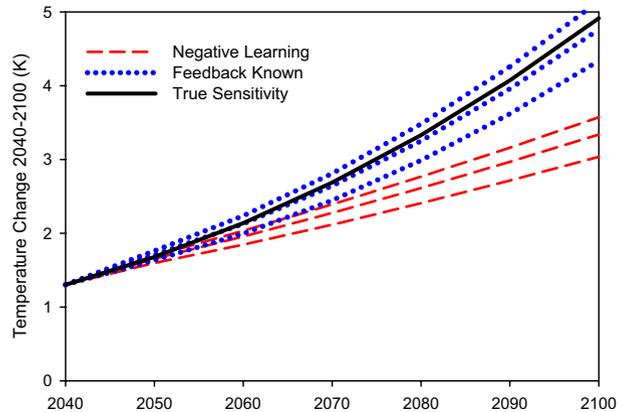
The costs of negative learning, measured as losses in consumption, are calculated by comparing the negative learning case to the counterfactual where the true model with the additional feedback was known all along. With an unknown positive feedback, negative learning leads to too little abatement during 2040–2100. After 2100 when the true model and sensitivity are learned, costs increase rapidly due to additional controls needed to compensate and the additional climate damage from earlier emissions (Fig. 6a). Negative learning in the presence of a negative feedback leads to more abatement than would be optimal if the true model were known. Consumption losses occur in the near-term from excessive abatement; for some SOWs there are net gains in the long run from both avoiding future emissions reductions and decreased damage (Fig. 6b). The difference in the discounted present value between the correct and incorrect model projections for any one SOW can be either positive or negative, depending on the climate sensitivity. For the negative feedback case, future damages are larger for higher sensitivities, so the gains from overcontrol are larger and there is a net benefit in those cases. The expectation of the discounted net present value of lost consumption³ is approximately \$1 trillion for the positive feedback and \$10 billion for the negative feedback. The asymmetry in the costs of missing a positive vs. a negative feedback may be a general feature of the climate problem, adding further support to precautionary policies, or may be an artifact of this particular illustration; additional study of this question is needed. While the precise numbers of these costs would likely differ in reality from those in the DICE model, the point is the same: the implications of negative learning in the context of climate change are significant welfare losses, because of the long lifetime of policy decisions.

5 Conclusions

The cases discussed are far from exhaustive – learning histories of aerosol and orbital forcing of climate, and the relative importance of smog precursors, exhibit some similarities

³ We calculate the present value of lost consumption using the PVC_{ex} method described in Füssler (2007), using the declining social rate of time preference and the growth rate from the reference no control solution to DICE-99. See Electronic Supplementary Material.

Fig. 5 Median and 90% bounds on global mean surface temperature projections made in 2040 based on posterior distributions for climate sensitivity. *Solid line* shows actual temperature change realization. True climate sensitivity is 5.0 K as in Fig. 4



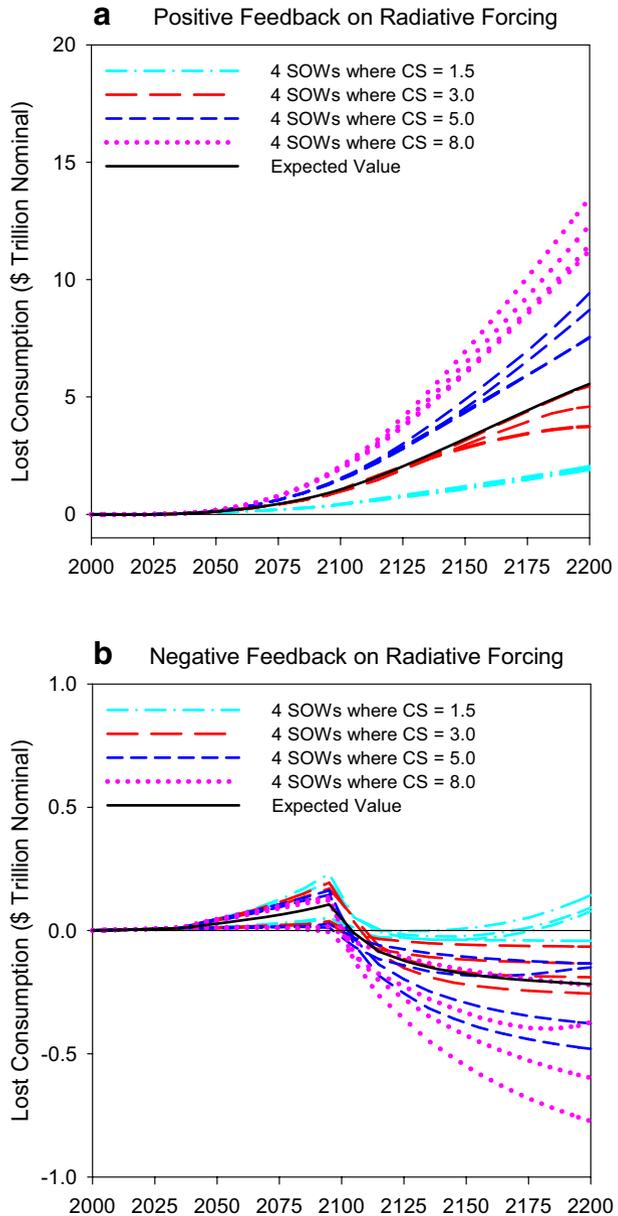
to the above, suggesting that negative learning may be pervasive in global change research and may continue to be a contributing factor to policies entailing substantial welfare losses to society. Decision analysis has evolved to incorporate more realistic descriptions of uncertainty and learning. Addressing the consequences of negative learning would be an appropriate next step. Progress depends on understanding its causes. While diverse, they share common elements. For example the inability of models to project outcomes in unexplored regions of physical, technological, or economic space may be viewed as arising from structural error in models, analogous to the structural error in the hypothetical climate sensitivity illustration. Methods such as Bayesian Model Averaging may be applied where alternative plausible models are available (which is often not the case, as for WAIS discussed above). However, the distinction between structural uncertainty and parameter uncertainty can be ambiguous. A very low parameter value and an absent process can produce nearly identical outcomes, and represent a similar level of misunderstanding (see ESM).

Furthermore, social dynamics of science and science assessment processes can enhance the potential for negative learning to occur and to influence policy makers (Patt 2007; Morgan and Henrion 1990, Section 7.7). The well-known inhibitory effect of disciplinary walls is inimical to progress in global change, and arguably played an important role in retarding understanding of ozone depletion (Christie 2001) and the behavior of WAIS. Patt (1999) has categorized assessments by purpose, identifying three types: consensus, advisory, and advocacy. Each has limitations, e.g., consensus assessments, like those of the IPCC, may tend to downplay or omit plausible extremes of the distribution of outcomes (Oppenheimer et al. 2007).

Given past experience, we recommend that the assessment process should be overhauled so that characterizing uncertainty (Moss and Schneider 2000) becomes a co-equal partner with consensus building. Attention should be devoted to the implications of poorly understood or hypothetical but plausible processes and alternative model structures. Recommendations for improvements that would minimize the possibility of negative learning in the production and use of such assessments include avoiding uncertainty assessment based only on model inter-comparison (Oppenheimer et al. 2007) and explicit reporting of disagreements among assessment authors (Patt 2007).

In addition, research funding from mission-oriented agencies also bears the potential to reinforce existing assumptions. Funding could be explicitly allocated to research that

Fig. 6 Economic impacts of negative learning over time, measured as lost consumption in the negative learning case relative to the case where true model was known. Effect of a positive feedback on radiative forcing is shown in **a** and of a negative feedback in **b**



explores alternative processes or assumptions not present in current models, and also targeted to identify and carry out those observations most likely to elucidate flaws in existing state-of-the-art models.

The treatment of ice sheets and sea level rise in IPCC's Fourth Assessment Report (AR4) provides a prime example of the need to reform current practice, and a challenge to policy makers considering responses in light of this assessment. Specifically, numerical values of uncertainty for 21st century sea level rise were based on inter-comparison of

models with known structural deficiencies in the representation of ice sheets, leading to an unrealistic decrease in the range and lowering of the maximum value of projected sea level rise compared to the Third Assessment Report. Policy makers should refrain from anchoring their beliefs in the quantitative uncertainty assessment while paying close attention to the often ignored qualitative caveats also reported in AR4.

In considering the design for a fifth assessment, particular attention should be devoted to developing a broader basis for quantitative uncertainty estimation for sea level rise including, e.g., employing empirical methods (Rahmstorf 2007) and expert elicitation (Vaughan and Spouge 2002; Lenton et al. 2008) on a parity with model-based evaluation. IPCC should sponsor workshops aimed specifically at developing means to integrate these diverse approaches. Funding agencies that support either physical modeling or research on decisions under uncertainty should consider this arena to be a funding priority.

The fact that negative learning can result in surprises with large impacts that persist for multi-decade timescales or longer (such as the ozone hole) suggests that adaptive management (Lempert et al. 1996, 2003) would have limited effectiveness. Waiting to learn, or to observe outcomes, can commit us to costly outcomes. Experience with ozone depletion is a prime example. Rather than take a fully precautionary approach in the 1970s by regulating production of CFCs for uses such as refrigeration, air conditioning, and solvents, policy makers, after instituting limits on use of CFCs in aerosol spray cans, deferred further action. Essentially, adaptive management failed because the phenomenon of negative learning places a premium on improving scientific learning and assessment skills proactively, rather than relying too heavily on approaches that involve learning from changes and responding to them.

But to make further recommendations, more information is needed and several important new lines of inquiry are suggested by our exploration of negative learning. One area of needed research is the application of history of science to this problem. Despite retrospective analyses of how projections performed against reality (Van Vuuren and O'Neill 2006; Rahmstorf et al. 2007), almost no effort has been devoted to understanding why learning in the global change arena, whether in basic science or assessment, sometimes goes awry. It would be timely to perform critical reviews of particular assessment case histories, not just to compare predictions to outcomes but to understand how specific judgments were made. Accordingly, IPCC working group discussions should become much more transparent so that the basis for particular decisions might be understood by non-participants and participants alike. Scientific assessment was developed ad hoc, at least in the global change arena, developing rapidly to meet a perceived need. But with more than three decades of experience in hand, the scientific community should apply the same strict standards of scholarship to examining how assessments perform and how they might be improved that it applies to its own research. More broadly, interaction between science and science history may promote an understanding of the range of plausible learning histories and their potential interaction with policy. Experience combating established mindsets ought to be imported from other fields (such as study of accidents, see ESM) and applied in research, in assessment, and in decision making.

New research questions in formal and quantitative decision theory are also raised by the consideration of negative learning. Because persistence of a direction of learning over years or even decades does not necessarily indicate progress toward true belief, negative learning may increase the value of precautionary approaches, such as hedging against low-probability outcomes (Yohe et al. 2004), in cases characterized by asymmetric responses as in the climate example above. But whether a bias in one direction of decisions is robust in the face of negative learning requires more extensive analysis than the illustration given

here. Other useful research would include a detailed exploration of the multiple causes of negative learning, perhaps leading to a set of indicators for the risk of negative learning that could be applied to particular current issues; and development of improved formal decision models that account for its possibility.

The experience gained to this point from learning about global change might well be borne in mind by policy makers as they evaluate a variety of hazards, such as that posed by the potential disintegration of the major ice sheets. Sometimes, half-measures don't avoid anything like half of the potential damages.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abbott A (2005) Gut feeling secures medical Nobel for Australian doctors. *Nature* 437:801
- Alley RB, Clark PU, Huybrechts P, Joughin I (2005) Ice-sheet and sea-level changes. *Science* 310:456–460
DOI 10.1126/science.1114613
- Andronova NG, Schlesinger ME (2001) Objective estimation of the probability density function for climate sensitivity. *J Geophys Res* 106(D19):22,605–22,611
- Arrow KJ, Fisher AC (1974) Environmental protection, uncertainty, and irreversibility. *Q J Econ* 88:312–319
- Benedick R (1998) Ozone diplomacy: new directions in safeguarding the planet. Harvard Univ Press, Cambridge, MA
- Bongaarts J, Bulatao RA (eds) (2000) Beyond six billion: forecasting the world's population. National Academy Press, Washington, DC
- Christie M (2001) Ozone layer: a philosophy of science perspective. Cambridge Univ Press, Cambridge, UK, New York
- Craig PP, Gadgil A, Koomey JG (2002) What can history teach us? A retrospective examination of long-term energy forecasts for the United States. *Annu Rev Energy Environ* 27:83–118
- Crutzen P, Oppenheimer M (2008) Learning about ozone depletion. *Clim Change* (this issue). DOI 10.1007/s10584-008-9400-6
- Draper D (1995) Assessment and propagation of model uncertainty. *J R Stat Soc Ser B Stat Methodol* 57:45–97
- Frederick S, Loewenstein G, O'Donoghue T (2002) Time discounting and time preference: a critical review. *J Econ Lit* XL:351–401
- Funtowicz SO, Ravetz JR (1992) The emergence of postnormal science. In: von Schomberg R (ed) *Science, politics, and morality*. Kluwer Academic, Dordrecht, pp 85–123
- Füssel H-M (2007) Methodological and empirical flaws in the design and application of simple climate-economy models. *Clim Change* 81:161–185
- Ha-Duong M, Mégie G, Hauglustaine D (2003) A pro-active stratospheric ozone protection scenario. *Global Environ Change* 13:43–49
- Hall J, Fu G, Lawry J (2007) Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties. *Clim Change* 81:265–281
- Hansen JE (2005) Slippery slope: How much global warming constitutes “dangerous anthropogenic interference”? *Clim Change* 68:269–279
- Henrion M, Fischhoff B (1986) Assessing uncertainty in physical constants. *Am J Phys* 54:791–797
- Huntington HG (1994) Oil price forecasting in the 1980s: what went wrong? *Energy J* 15:1–22
- Jones RA, Ostroy JM (1984) Flexibility and uncertainty. *Revs Econ Studies* L1:13–32
- Joos F, Müller-Fürstenberger G, Stephan G (1999) Correcting the carbon cycle representation: how important is it for the economics of climate change? *Environ Model Assess* 4:133–140
- Kaufmann R (1997) Assessing the dice model: uncertainty associated with the emission and retention of greenhouse gases. *Clim Change* 35:435–448
- Keilman N (1999) How accurate are United Nations population forecasts? In: Lutz W, Vaupel JW, Ahlburg DA (eds) *Issue supplement: frontiers of population forecasting*, *Pop Develop Rev* 24:15–41

- Keller K, McInerney D (2007) The dynamics of learning about a climate threshold. *Clim Dyn* 30:321–332
DOI 10.1007/s00382-007-0290-5
- Keller K, Bolker BM, Bradford DF (2004) Uncertain climate thresholds and optimal economic growth. *J Environ Econ Manage* 48:723–741
- Kelly DL, Kolstad CD (1999) Bayesian learning, growth, and pollution. *J Econ Dyn Control* 23:491–518
- Kitcher P (1993) *The advancement of science: science without legends, objectivity without illusions*. Oxford Univ Press, New York
- Kolstad CD (1996) Learning and stock effects in environmental regulation: the case of greenhouse gas emissions. *J Environ Econ Manage* 31:1–18
- Kuhn TS (1962) *The structure of scientific revolutions*. Univ of Chicago Press, Chicago
- Lange A (2003) Climate change and the irreversibility effect – combining expected utility and maximin. *Environ Res Econ* 25:417–434
- Lee R, Tuljapurkari S, Lee RD, Auerback AJ (eds) (2000) *Demographic change and fiscal policy*. Cambridge Univ Press, Cambridge, UK
- Lempert RJ, Schlesinger ME, Bankes SC (1996) When we don't know the costs or the benefits: adaptive strategies for abating climate change. *Clim Change* 33:235–274
- Lempert RJ, Popper SW, Bankes SC (2003) *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*. Rand Corporation, Santa Monica
- Lenton TM et al (2008) Tipping elements in the earth's climate system. *PNAS* 105:1786–1793
- Little C et al (2007) Toward a new generation of ice sheet models. *Eos* 88:578–579
- Lutz W, Sanderson W, Scherbov S (2001) The end of world population growth. *Nature* 412:543–545
- MacAyeal DR (1992) Irregular oscillations of the West Antarctic ice sheet. *Nature* 359:29–32
- Morgan MG, Henrion M (1990) *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge Univ Press, Cambridge, New York
- Morgan MG, Kandlikar M, Risbey J, Dowlatabadi H (1999) Why conventional tools for policy analysis are often inadequate for problems of global change. *Clim Change* 41:271–281
- Moss RH, Schneider S (2000). In: Pachauri R, Taniguchi T, Tanaka K (eds) *Guidance papers on the cross cutting issues of the Third Assessment Report of the IPCC*. IPCC, Geneva
- Nordhaus WD, Boyer J (2000) *Warming the world: economic models of global warming*. MIT Press, Cambridge, Mass
- NRC (1984) *Causes and effects of changes in stratospheric ozone: update 1983*. National Academy Press, Washington, DC
- O'Neill BC, Scherbov S, Lutz W (1999) The long-term effect of the timing of fertility decline on population size. *Pop Develop Rev* 25:749–756
- Oppenheimer M (1998) Global warming and the stability of the West Antarctic ice sheet. *Nature* 393:325–332
- Oppenheimer M, Alley RB (2004) The West Antarctic ice sheet and long term climate policy. *Clim Change* 64:1–10
- Oppenheimer M, O'Neill BC, Webster M, Agrawala S (2007) The limits of consensus. *Science* 317:1505–1506
- Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641–646
- Overpeck JT, Otto-Bliesner BL, Miller GH et al (2006) Paleoclimatic evidence for future ice-sheet instability and rapid sea-level rise. *Science* 311:1747–1750
- Parson E (2003) *Protecting the ozone layer: science and strategy*. Oxford Univ Press, Oxford, New York
- Patt AG (1999) Extreme outcomes: the strategic treatment of low probability events in scientific assessments. *Risk Decis Policy* 4:1–15
- Patt AG (2007) Assessing model-based and conflict-based uncertainty. *Glob Environ Change* 17:37–46
- Pielke RA Jr (2001) Room for doubt. *Nature* 410:151
- Popp D (2004) ENTICE: endogenous technological change in the DICE model of global warming. *J Environ Econ Manage* 48:742–768
- Rahmstorf S (2007) A semi-empirical approach to projecting future sea-level rise. *Science* 315:368–370
- Rahmstorf S, Cazenave A, Church JA, Hansen JE, Keeling RF, Parker DE, Somerville RCJ (2007) Recent climate observations compared to projections. *Science* 316:709
- Rowland FS (1989) Chlorofluorocarbons and the depletion of stratospheric ozone. *Am Sci* 77:36–45
- Schelling TC (1994) Intergenerational discounting. *Energy Policy* 23:395–402
- Schneider SH, Turner BL, Garriga HM (1998) Imaginable surprise in global change science. *J Risk Res* 1:165–185
- Shlyakhter AI (1994) An improved framework for uncertainty analysis: accounting for unsuspected errors. *Risk Anal* 14:441–447

- Small MJ, Fischbeck PS (1999) False precision in Bayesian updating with incomplete models. *J Hum Ecol Risk Assess* 5(2):291–304
- Somerville R, Le Treut H, Cubasch U, Ding Y, Mauritzen C, Mokssit A, Peterson T, Prather M (2007) Historical overview of climate change. In: Solomon S, Qin D et al (eds) *Climate change 2007: the physical science basis. Contribution of working group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ Press, Cambridge, UK, New York
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131
- UN Department of Economic and Social Affairs, Population Division (2005) *World population prospects: the 2004 revision*. United Nations, New York
- van de Kaa DJ (1987) Europe's second demographic transition. *Population Bulletin* 42(1), The Population Reference Bureau, Washington, DC
- van Vuuren D, O'Neill BC (2006) The consistency of IPCC's SRES scenarios to 1990–2000 trends and recent projections. *Clim Change* 75:9–46
- Vaughan DG (2008) West Antarctic ice sheet collapse – the fall and rise of a paradigm. *Clim Change*, in press
- Vaughan DG, Spouge JR (2002) Risk estimation of collapse of the West Antarctic ice sheet. *Clim Change* 52:65–91
- Webster MD (2002) The curious role of learning: should we wait for more data? *Energy J* 23:97–119
- Webster MD, Jakobovits L, Norton J (2008) Learning about climate change and implications for near-term policy. *Clim Change* (this issue). DOI [10.1007/s10584-008-9406-0](https://doi.org/10.1007/s10584-008-9406-0)
- WMO (1986) *Atmospheric ozone 1985: assessment of our understanding of the processes controlling its present distribution and change*. Rep No 16, WMO, Geneva
- WMO (1988) *Report of the international ozone trends panel-1988*. Global Ozone Research and Monitoring Project, Rep No18, WMO, Geneva
- WMO (1991) *Scientific assessment of stratospheric ozone 1991*. Rep No 25, WMO, Geneva
- WMO (2006) *Scientific assessment of ozone depletion: 2006, executive summary*. WMO and UN Environment Program, Geneva
- Yohe G, Andronova NG, Schlesinger ME (2004) To hedge or not against an uncertain climate future? *Science* 306:416–417