**Noname manuscript No.**
(will be inserted by the editor)

# Disconcerting learning on climate sensitivity and the uncertain future of uncertainty

**Alexis Hannart · Michael Ghil ·**
**Jean-Louis Dufresne · Philippe Naveau**

Received: date / Accepted: date

**Abstract** How will our estimates of climate uncertainty evolve in the coming years, as new learning is acquired and climate research makes further progress? As a tentative contribution to this question, we argue here that the future path of climate uncertainty may itself be quite uncertain, and that our uncertainty is actually prone to increase even though we learn more about the climate system. We term *disconcerting learning* this somewhat counter-intuitive process in which improved knowledge generates higher uncertainty. After recalling some definitions, this concept is connected with the related concept of *negative learning* that was introduced earlier by Oppenheimer et al. [2008]. We illustrate disconcerting learning on several real-life examples and characterize mathematically certain general conditions for its occurrence. We show next that these conditions are met in the current state of our knowledge on climate sensitivity, as shown physically on hand of an energy balance model of climate. Finally, we discuss the implications of these results on developing policy for adaptation and mitigation.

**Keywords** Climate change uncertainty · Knowledge evolution · Learning models

## 1 Introduction and motivation

Strong scientific consensus prevails over the fact that Earth's climate is currently warming and will be warming further over the coming decades, as a consequence of the radiative perturbations caused by anthropogenic greenhouse-gas (GHG) emissions. The conclusions of the IPCC's Fourth Assessment Report (AR4: [Solomon et al. (2007)], [AR4] hereafter) further buttress this consensus. There is, however, substantial uncertainty regarding the extent of future warming, as pointed out in the same report and in many of its references.

A. Hannart
IFAECI
Ciudad Universitaria
Buenos Aires, Argentina
Tel.: +5411-4787-2693
Fax: +5411-4788-3572
E-mail: alexis.hannart@cima.fcen.uba.ar

This uncertainty renders decision making on appropriate adaptation and mitigation steps more difficult. In addition, the uncertainty level regarding future climate evolution has not decreased significantly over the past decades. This observation paves the way for climate-warming naysayers; it is sometimes used as an argument to discredit climate science as a whole and to slow down action on this issue. Lively scientific debate continues on the extent and the reasons for the uncertainty. This debate motivates us to revisit here the question of the future evolution of uncertainties.

Uncertainties regarding future climate warming are usually divided into three categories [Hawkins and Sutton (2009)]: (i) those regarding GHG increase scenarios [AR4]; (ii) those arising from the climate system's internal variability [Ghil et al. (2008)]; and (iii) those inherent to the climate system's long-term response to a given forcing. Because contribution (i) is part and parcel of humankind's future course of action and the relative contribution of (ii) may vanish after a few decades, we focus on the third category, which we refer to henceforth simply as *climate uncertainty*. To quantify it, a widely used metric consists in the spread $\sigma_{\Delta T}$ associated with the probability density function (PDF) of *climate sensitivity*; the latter is defined here as the change $\Delta T$ in global equilibrium surface temperature $T$ associated with a doubling in atmospheric $CO_2$ concentration.

This metric stems from the fact that the diversity of plausible long-term future climate states for a given emission scenario is determined, to a large extent, by the range of climate sensitivity $\Delta T$. According to [AR4] — which compiled PDFs of $\Delta T$ obtained by various studies over the last few years — $\Delta T$ is likely to lie between 2°C and 4.5°C, a range which is still high. It is thus relevant for socio-economic and political decision making to ask how this range will evolve in the future, as climate research makes further progress.

To answer this question, one can find, on the one hand, numerous studies (e.g., [Stainforth et al. (2005),Knutti and Hegerl (2008),Roe and Baker(2007),Hannart et al. (2009)] and references therein) that focus on the reasons for the presently high range of $\Delta T$. These studies identified a number of key research areas — such as cloud processes (e.g., [Soden and Held (2006),Dufresne and Bony (2008)]) or oceanic variability and response [Dijkstra and Ghil (2005),Ghil et al. (2008)] — whose better understanding and modeling may potentially lead to a reduction of the uncertainty in $\Delta T$.

On the other hand, a vast body of literature addresses the question of learning at an epistemological level and that of uncertainty in the general context of scientific research. For instance, the very definitions of learning and scientific progress, as well as the question of the existence of truth, have been debated at length over millennia of philosophical tradition (e.g., [Aristotle (40 B.C.),Bacon (1605), Kuhn (1962)]). The interplay between learning, uncertainty, erroneous judgements and decision making has received increased attention in recent years, especially in the context of environmental policy (e.g., [Crutzen and Oppenheimer (2008), Keller and McInerney (2007),O'Neill et al. (2006)]). There are still but few studies, however, (e.g., [Oppenheimer et al. (2008),Webster et al. (2008)]) that address jointly the question of the uncertainty in $\Delta T$ — so often debated in the climatic literature — and the aforementioned, more general literature on learning and progress.

The [Oppenheimer et al. (2008)] paper ([ONW08] hereafter) not only included such a broader perspective, but also made several important points that we briefly recall here. First and foremost, [ONW08] challenged the intuitive, and hence pervasive view that usually enters into decision making on environmental problems, namely that "scientific research can be equated (...) with truer beliefs about the outcomes of problems (...) thus providing a superior basis for crafting solutions." In formulating their challenge, these authors introduced the broad concept of *negative learning* to describe any situation where "new technical information leads to scientific beliefs that diverge over time from the a posteriori right answer."

[ONW08] illustrated the concept of negative learning on hand of four prominent case histories, thus showing that negative learning did occur in the past. One of these case histories dealt with advances in the understanding of ozone depletion in the 1970's and 80's. In the latter case, the negative aspects of the learning touched upon important facets of the problem under study, for reasons that were similar to those involved in global warming, and did affect policy making. Finally, [ONW08] showed that negative learning on climate sensitivity could well occur in the future, for instance if an unknown radiative feedback is not incorporated into climatic models, i.e. if the latter are subject to structural error.

The present article pursues the same line of questioning as [ONW08]. While [ONW08] focused on the conditions of occurrence and on the damaging effects of negative learning, they did not examine the detailed dynamics of learning in the "non-negative" case, which they termed *progressive learning*. Progressive learning, though, may still be problematic when it comes to uncertainty. Our main point in the present paper is that, while progressive learning always leads to truer beliefs by definition, it does not systematically imply that these truer beliefs are less uncertain. We thus introduce the term *disconcerting learning* to describe this nonetheless counter-intuitive situation, in which new information leads to scientific beliefs that are closer to the a posteriori right answer, while still being marked by greater uncertainty. Conversely, we use the term *reassuring learning* for the more intuitive situation in which progressive learning does lead to less uncertainty. These four possibilities — of negative vs. positive, and of disconcerting vs. reassuring learning — are illustrated in Fig. 1, and are explained more precisely in the next section.

Although the term "disconcerting learning" introduced here is novel, to the best of our knowledge, earlier works in statistics, probability and economics have already pointed out the existence of this situation (e.g., [Burdett (1996)] or [Zidek and van Eeden (2003),Bagnoli and Bergstrom (2005),Chen et al. (2010)] or [Chen (2011)]). These theoretical studies have also established a few rigorous results concerning the conditions of occurrence of such a situation, but only under some very restrictive conditions.

Hence, a general theory of disconcerting learning is lacking for the time being and [Chen et al. (2010)] have even described such a theory as elusive. In any case, research on this type of learning is still in its early days and more work is needed to improve its understanding. The importance of uncertainty regarding climate sensitivity motivates us to do so, and the relevance of this motivation will be made clear in Section 4. Thus, our main contribution here is to further illustrate and analyze why and how disconcerting learning occurs and to demonstrate that it is prone to occur in learning about $\Delta T$ in the future.

The paper is organized as follows. In Section 2, we recall the definitions of [ONW08] and introduce our own definitions and notation. In Section 3, disconcerting learning is illustrated based on two real-life, biomedical problems that are more insightful in our view than the climatological situation eventually at stake here. Then, we introduce a simple Bayesian model of progressive learning and we use it to study the general properties of and conditions for the occurrence of disconcerting learning. In Section 4, we return to the physics and sensitivity of climate and we show that disconcerting learning may occur in studying $\Delta T$, i.e. that climate uncertainty may persist or increase even though scientific research yields progressive results. We emphasize this finding in the idealized context of a linear energy balance model of climate and illustrate it more concretely with a real example. Section 5 comments on some policy implications of the present results, while Section 6 discusses some further aspects of our work and states our conclusions.

## 2 Definitions and notations

As in [ONW08], an *outcome* is any quantity, process or structure of interest, and we denote it by $x$. The state of knowledge on $x$, for a given observer at a given moment, consists in the set of informations relating to $x$ that are available to the observer at that moment. We denote this set by $\mathcal{I}$ and represent the state of knowledge on $x$ in probabilistic terms by using the pdf $p(x \mid \mathcal{I})$ of $x$ conditional on $\mathcal{I}$. Further considerations on the relevance of this probabilistic description of a state of knowledge, as well as on the underlying interpretation of probabilities, can be found in Supplemental Material A.

Learning on $x$ is thus defined here by a change in the pdf of $x$ subsequent to its update by some new information. Such changes may occur as a result of developments in theory, modeling, observations or experiments. We denote by $\mathcal{I}_0$, $\mathcal{I}$ and $\mathcal{I}_1 = \mathcal{I}_0 \cup \mathcal{I}$, respectively, the a priori information, the new information learnt, and the a posteriori information.

With this notation, learning can be formalized in the Bayesian framework as follows:

$$p_1(x) = \frac{p_0(x)\mathcal{L}(x \mid \mathcal{I})}{\int p_0(x)\mathcal{L}(x \mid \mathcal{I})\,\mathrm{d}x}\,. \tag{1}$$

In Eq. (1), the prior distribution $p_0(x) = p(x \mid \mathcal{I}_0)$ represents the initial state of knowledge on $x$ and is multiplied by a likelihood function $\mathcal{L}(x \mid \mathcal{I})$ that summarizes the new information. This product yields, after normalization, the a posteriori distribution $p_1(x) = p(x \mid \mathcal{I}_1)$.

In this probabilistic definition, the level of uncertainty on $x$ that is associated with a given state of knowledge $\mathcal{I}$ is easily quantified by using $p(x \mid \mathcal{I})$. We do so using the standard deviation $\sigma_{\mathcal{I}}$ of this pdf as a metric. Other metrics for uncertainty are possible, i.e. Shannon entropy, but this choice is not critical for the present discussion; see Supplemental Material B.

In this Bayesian setting, the definition of disconcerting learning given in Section 1 becomes simply

$$\sigma_1 > \sigma_0\,, \tag{2}$$

i.e. the uncertainty level on the outcome increases even though more information on $x$ was gained. Conversely, reassuring learning corresponds to a learning situation in which $\sigma_1 \leq \sigma_0$, while the definition of negative learning given by [ONW08] becomes

$$p_1(x^*) < p_0(x^*)\,, \tag{3}$$

whereas $p_1(x^*) \geq p_0(x^*)$ for progressive learning; here $x^*$ denotes the true value of $x$. Figure 1 describes the four possibilities associated with the pair of inequalities between prior and posterior variance and between prior and posterior bias.

Since the key idea associated with negative learning is to describe a situation in which "scientific beliefs diverge over time from the a posteriori right answer," one could consider an alternative, but very closely related definition of negative learning as an increase in bias — i.e. $\mid \mu_1 - x^* \mid > \mid \mu_0 - x^* \mid$; such a definition would more closely parallel the definition of disconcerting learning as an increase in uncertainty, where $\mu$ is the distribution mean. The purpose of this article is primarily to study the situation of disconcerting learning, $\sigma_1 > \sigma_0$, in the context of progressive learning, which we choose for simplicity to be defined as $p_1(x^*) \geq p_0(x^*)$.

## 3 Conditions of occurrence of disconcerting learning

### 3.1 Two illustrations of disconcerting learning

With the quantitative definitions formulated in Section 2 in hand, we now proceed to exhibit two typical situations of disconcerting learning that occur in the medical context, before turning to our main climatic applications in Section 4.

*Medical screening test.* Suppose one is interested in whether or not an individual is affected by a disease. The outcome $x$ here is a binary variable with $x = 1$ if the individual is affected by the disease and $x = 0$ if not. We assume that a medical screening test is available for the detection of the disease. The result of the test $z$ can also be treated as a binary variable with $z = 1$ if the test is positive and $z = 0$ if it is negative.

Our initial state of knowledge consists simply in the mean frequency of occurrence $q_0$ of the disease in the population. We thus have $p_0(x) = q_0^x(1 - q_0)^{1-x}$, and the a priori standard deviation is given by $\sigma_0 = \sqrt{q_0\,(1 - q_0)}$. Then, the medical screening test is conducted on the individual, and we assume that it gives a positive result, $z = 1$, thus suggesting illness. However, the test is known to be imperfect: it has a false positive frequency $q$ and a false negative frequency $q'$.

The new probability $q_1$ and standard deviation $\sigma_1$, after learning the test result, are equal to

$$q_1 = \{1 + (1 - q_0)/q_0\beta^2\}^{-1}, \qquad \sigma_1 = \sigma_0\{\beta q_0 + \beta^{-1}(1 - q_0)\}^{-1}\,, \tag{4}$$

where $\beta = \sqrt{(1 - q')/q}$. It thus follows immediately that, whenever the prior probability $q_0$ is smaller than $(1+\beta)^{-1}$, learning the positive result of the test leads to an increase of the uncertainty level, i.e. to disconcerting learning. Furthermore, the increase is largest for $q_0 = q = q'$.

In the present context, $q_0$ is typically small but nonzero — i.e., illness is a priori possible, but remains the exception and health the rule — and so are $q$ and $q'$, since medical tests are reasonably trustworthy, although not completely so. We are thus often in a situation in which the condition $q_0 < (1+\beta)^{-1}$ could be met, and where $q_0 \simeq q \simeq q'$ is also perfectly plausible. For instance, [Humphrey et al. (2002)] gives $q_0 = 0.06$, $q = 0.13$ and $q' = 0.02$. For these values, learning a positive result almost doubles the standard deviation from 0.23 to 0.45, i.e. it is an instance of strongly disconcerting learning.

It should be emphasized that disconcerting learning in such a situation is possible, but not unavoidable. For instance, learning a negative test result ($z = 0$) will result in a sharp decrease in the standard deviation, from 0.23 to 0.16, i.e. reassuring learning. Learning the positive test result ($z = 1$) could also be reassuring in two situations.

First, if the test were much more reliable, the resulting probability of illness $q_1$ would be closer to one and the standard deviation closed to zero. In the above example, this would require $q < 0.004$; for $q = 0.001$, the standard deviation $\sigma_1$ would then decrease from 0.23 to 0.12. Second, if the prior probability of illness $q_0$ was close to 0.5, the prior uncertainty level would nearly equal its maximal value and its posterior value would thus necessarily decrease. In the above example, this would require $q_0 > 0.26$ for the given test reliability.

To summarize the insights gained from this example, one can state the following: For a binary outcome with contrasted a priori probabilities, $0 < q \ll 0.5 \ll 1-q < 1$, new information that favors the unexpected modality tends to be disconcerting, as long as the new information is not conclusive. One can thus speculate that in general, disconcerting learning occurs when surprising but inconclusive evidence is found.

*Disease incidence rate.* We focus next on a slightly different, but connected, real-life situation. We are interested this time in the frequency of occurrence $x$ of the disease in a given population. In this case, learning is obtained by observing whether a new individual is ill or not.

Let us suppose that this new individual is found to be affected by the disease, i.e. $z = 1$. Consider, for definiteness, that at the time $z$ is observed, $n = 15$ individuals were already observed and that $k = 3$ of them are ill and $n - k = 12$ are healthy. Finally, suppose that prior to this initial observation of $n$ individuals, $x$ was assumed to be uniform on $[0, 1]$. In this situation, we find from Eq. (1) that $p_0(x)$ is the beta distribution $\mathcal{B}(k, n-k)$ and that the posterior $p_1(x)$, after learning $z = 1$, is $\mathcal{B}(k + 1, n - k)$. (Supplemental Material C). Hence:

$$\sigma_1 = \sigma_0\{(1 - (k+1)^{-1})(1 + n^{-1})(1 + 2\,n^{-1})\}^{-\frac{1}{2}} \,. \tag{5}$$

Equation (5) yields $\sigma_1/\sigma_0 = 1.05$ and we find ourselves in a situation of disconcerting learning as well.

As in the previous example, disconcerting learning happens here because the new information is simultaneously surprising — i.e., the observation of a new case of disease was rather unexpected, due to the fact that most previous observations were of healthy people — and yet inconclusive — i.e., one extra case of disease is insufficient to properly estimate the frequency of disease occurrence over the population. Conversely, had the observation been unsurprising, i.e. had $z$ been equal to its expected modality of zero, $z = 0$ for $n = 16$, the spread would have decreased. Likewise, had the observation been surprising but conclusive, i.e. had we observed a very large number (say 500) of cases of disease instead of one single case, the spread would also have decreased.

## 3.2 Disconcerting learning and shape of the prior distribution

In this subsection, we now focus on scalar, continuous outcomes $x$, and we address the following two questions: are there characteristics inherent to the prior pdf $p_0(x)$

that increase the chances for disconcerting learning to occur; and if so, which? We address these questions based on a review of results available in recent literature and on hand of a detailed simulation study designed for this purpose.

The simulation study relies on a Bayesian learning model applied to a variety of prior distributions that combine several shape features, namely: skewness, from fully symmetric to pronounced asymmetry; kurtosis, from leptokurtic to platykurtic; tail size, from bounded to heavy-tailed; and multimodality, from one to two modes. Distributions combining these features were generated based on the Pearson family. The literature review as well as the simulation study are described and illustrated in detail in Supplemental Material D and E, respectively.

This exploration yielded three main findings. First, there are essentially two characteristics that enhance the likelihood of disconcerting learning to occur, namely that $p_0(x)$ (i) is highly skewed, and (ii) that it possesses heavy tails, cf. Figs. 2 and SM1; when combining these two characteristics, the incidence of disconcerting learning tends to increase substantially. This result sheds light on the illustrative examples given in Section 3.1: in each instance, when disconcerting learning occurred, the prior distribution did indeed have significant skewness. For instance, the skewness of the Bernoulli prior $p_0(x) = q_0^x\,(1-q_0)^{1-x}$ was equal to 3.7 for $q_0 = 0.06$, and that of the Beta distribution $\mathcal{B}(3, 12)$ was equal to 1.5; for comparison purposes, the skewness of the highly asymmetric exponential distribution is equal to 2.

Second, disconcerting learning in this model is always associated with a large swing in the value of the mean; see Figs. 2 and SM2. This result further supports the validity of the speculation in Section 3.1, according to which disconcerting learning occurs whenever surprising evidence is found, as shown by large shifts in the expected value of the outcome. Note that the two findings summarized so far are perfectly consistent with, and shed light on, each other. Indeed, skewed and heavy-tailed distributions share a property that symmetric, light-tailed distributions do not have: They assign high probabilities to the occurrence of values that are remote from the "central core" of the distribution — i.e., unexpected values — which are precisely those that give rise to large swings and disconcerting learning.

Third, disconcerting learning is systematically associated with a large dispersion of the trajectories of the uncertainty (Figure SM2b). This finding can be understood qualitatively by considering the fact that no distribution can generate surprises in a systematic manner — otherwise they would not be surprises. In other words, a distribution that is compatible with the occurrence of surprises— i.e., that is skewed or heavy tailed or both — still generates unsurprising evidence most of the time. Accordingly, a distribution that is compatible with the occurrence of disconcerting trajectories still generates reassuring trajectories most of the time, resulting in a widespread range of trajectories.

## 4 Disconcerting learning and climate sensitivity

### 4.1 Implications from recent PDFs of climate sensitivity

At present, most PDFs obtained for climate sensitivity $\Delta T$ are skewed and heavy-tailed (Fig. 3). There is ongoing debate and discussion on the reasons for the redundancy of skewness in these PDFs (see for instance [Allen et al.(2006)] or

[Zaliapin and Ghil (2010),Roe and Baker(2011)]) but these interesting discussions are beyond our scope, which is merely to analyze the implications of these factual features for disconcerting learning. In this purpose and as a starting point, we first applied directly the general learning model used in Section 3.2 to several PDFs of climate sensitivity sampled from recent studies (references in Supplemental Material). Doing so, we obtained a set of future trajectories for the uncertainty in climate sensitivity (Fig. 3). Unsurprisingly, we find that (i) disconcerting learning on climate sensitivity is prone to occur in the future; that (ii) it is most severe when the prior distribution is highly skewed; and that (iii) the future trajectory of the uncertainty is itself quite uncertain.

The interpretation of these findings is quite straightforward by using the insights gained from Sections 3.1 and 3.2: because of skewness, high values of climate sensitivity are unlikely but cannot be discarded altogether thus our present state of knowledge allows surprises to occur as we learn more. More specifically, skewness here implies that the shape of the PDF is flatter for high values than it is for medium values, i.e. high values tend to be more evenly distributed than medium values. This means that our knowledge is more imprecise in the upper range than it is in the medium range. Therefore, if a new piece of information shifts our beliefs with respect to climate sensitivity upwards, this will take us into a domain of values about which we know less. The new information will thereby raise more questions and doubts than it will bring answers and certainties. Uncertainty will thus increase and the learning will be disconcerting.

## 4.2 An illustration of disconcerting learning on climate sensitivity

We now discuss and illustrate more concretely how learning on climate sensitivity may occur. For this purpose, we adapt our general Bayesian learning framework to the case of climate sensitivity, by following an approach similar to that of [Kelly and Kolstad (1999),Leach (2007),Webster et al. (2008)] in which the new knowledge on climate sensitivity $\Delta T$ is obtained from a new temperature observation $T$. The latter is interpreted in terms of climate sensitivity by means of a climate model that establishes a probabilistic connection between $\Delta T$ and $T$, summarized by $p(T \mid \Delta T)$. Then, the Bayesian update Equation (1) yields:

$$p_1(\Delta T) = \frac{p_0(\Delta T)\, p(T \mid \Delta T)}{\int p_0(\Delta T)\, p(T \mid \Delta T)\, \mathrm{d}T}. \tag{6}$$

Equation (6) describes a learning process on $\Delta T$ which combines two types of information: (i) a climate observation $T$ indirectly linked to $\Delta T$; and (ii) a climate model representing the available physical knowledge with respect to the indirect link between $T$ and $\Delta T$. In the following, we focus on the observational learning (i) only. Note, though, that an improvement of our theoretical understanding of climate physics subsequently affecting the model (ii) can also be accounted for using this framework. But in any case, even though the learning process is assumed to be purely observational here, the climate model is at the core of this process, because it entirely defines the distribution $p(T \mid \Delta T)$ of the new observation conditional on climate sensitivity used in Equation (6).

Any choice of climate model — deterministic, stochastic, from low to high complexity — is in theory fit within this framework (Supplemental Material F).

For the present, learning-theoretic purposes, we chose a stochastic version of a zero-dimensional energy balance model, in discrete time:

$$\kappa \frac{\delta \overline{T}_t}{\delta t} = -\frac{\Delta R_0}{\Delta T}\, \overline{T}_t + R_t \ , \quad T_t = \overline{T}_t + \varepsilon_t. \tag{7}$$

where $R_t$ is the radiative forcing at $t$; and $\kappa$, $\Delta R_0$ and $\sigma$ are climate parameters that are assumed to be known, i.e. $\Delta T$ is assumed to be the only uncertain parameter here. The model and its assumptions are described in detail and solved in Supplemental Material G; it yields a closed form expression of $p(T_t \mid \Delta T)$, the distribution of $T_t$ seen from time $t-1$ conditional on $\Delta T$. For observed trajectories of the forcing $R_t$ and of the temperature response $T_t$, we can thus use $p(T_t \mid \Delta T)$ and Eq. (6) to perform iterative updates and obtain the successive PDFs $p_t(\Delta T)$ of $\Delta T$ at each instant $t$, as well as the corresponding successive values of the standard deviations $\sigma_t$ — i.e. the trajectories of the uncertainty in $\Delta T$.

We applied the latter procedure by using an initial prior distribution $p_0(\Delta T)$ that synthesizes the [AR4] inferences on PDFs of climate sensitivity, i.e. a mean equal to 3.2°C, a likely range of 2°C–4.5°C, and a positive skewness that we assume equal to 0.7. We stopped this updating process at present time $t$, and we studied the sensitivity of $p_{t+1}(\Delta T)$ to $T_{t+1}$, the new temperature observation for year 2013, which is assumed to be still unknown at time $t$. We found that, for the simulated value $T_t$, whenever $T_{t+1} - T_t \leq 0.24$°C — i.e., for a $\delta T$ in Eq. (7) that corresponds to either cooling or to a moderate warming between $t$ and $t+1$ — we will always have $\sigma_{t+1} \leq \sigma_t$ and, if so, the new observation corresponds to reassuring learning. Conversely, when $T_{t+1} - T_t > 0.24$°C, i.e. for a more intense warming between $t$ and $t+1$, then $\sigma_{t+1} > \sigma_t$ and the new observation corresponds to disconcerting learning; see Fig. 4.

These findings match our previous results and conclusions. Indeed, according to the information available at $t$, a moderate warming between $t$ and $t+1$ is to be expected. If a moderate warming materializes, this will be in line with the expected value of climate sensitivity and will confirm this value; thus learning in this case will be reassuring. Conversely, if an intense warming materializes, this will be unexpected and come as a surprise. This surprising observation will tend to indicate that climate sensitivity is higher than expected. The indication, though, will be inconclusive: first, because the unexpected observation could still be explained to a large extent by short-term fluctuations caused by internal variability, rather than characterizing the climate system's long-term response; and second, because the a priori PDF $p_0(\Delta T)$ is skewed towards high values. The role of the latter skewness is particularly important here. To further emphasize it, we applied the same update procedure of Eq. (7) on the same simulated values of temperature, but initializing this time with a Gaussian, symmetric prior distribution. We find that under such an a priori, learning is always reassuring no matter the value of the new observation $T_{t+1}$; this is the case even for unexpected, intense warming between $t$ and $t+1$ (not shown).

We end this subsection with an example of an actual recent observation that, in line with our above illustration, may arguably be considered as disconcerting. In 2007, the yearly minimal extent of Arctic sea ice has started to decline abruptly and faster than expected by climate models ([Stroeve et al. (2012)]). All observations after 2007 were consistent with an abrupt change, especially the latest to date

(September 2012). Such a situation is indeed surprising, yet it is clearly inconclusive: On the one hand, it tends to indicate that climate sensitivity may be higher than expected. It might even suggest that the climate system has passed a tipping point ([Lenton et al. (2008), Abbot et al.(2011), Livina and Lenton (2012)]), even though this possibility is still actively debated ([Tietsche et al. (2011)]). On the other hand, internal variability is high in the polar regions [Ghil et al. (1987), Darby and Mysak (1993)], and it could explain this situation without requiring a high sensitivity. So, it can be argued that the recent decline in Arctic sea ice raises more questions than it provides answers, and it is therewith a disconcerting observation. To settle the matter would require applying the procedure described in Section 4.2 to a more detailed model than our linear model of Eq. (7). Such a model would have to explicitly represent sea ice and allow for the presence of tipping points [Ghil (2001)]; it is thus beyond the scope of the present paper.

## 5 Policy implications

The key finding of the previous section is that the future trajectory of uncertainty with respect to climate sensitivity is itself uncertain and that this uncertainty could well increase. Such a finding may have implications for the development of climate change mitigation policy. In the present section, we merely discuss whether and how disconcerting learning may affect policy, but stop short of any recommendations on this matter.

It is clear that climate change may seriously affect humankind's socio-economic well-being in the future. The extent and cost of any future damages, though, are quite uncertain, in particular because of uncertainty concerning climate sensitivity. It is also clear that uncertain future damages can be mitigated by actions taken today — e.g., a $CO_2$ abatement achieved by various means, including a carbon tax, for instance. Unlike the cost of climate damages, which lies in future, these mitigation actions have a cost that is immediately incurred and is also fairly accurately known.

The crucial issue at the heart of mitigation policies is thus one of defining the right trade-off between uncertain, future damages and certain, present costs. The issue can thus be posed as a risk management problem, and there is an abundant literature in which it is tackled within this framework, using the concepts and tools of decision under uncertainty ([Arrow and Fisher (1974)]). In spite of the common analysis framework used to tackle mitigation policy design, policy recommendations range from very substantial ([Rahmstorf (1999)]) to very low ([Tol (1997)]) near-term $CO_2$ abatement. Such a degree of divergence may relate to the fact that optimal policy design depends critically on a number of key assumptions built into both the economic and the climate model involved in the analysis. A lack of consensus prevails on these assumptions, e.g. on the assumption of reversibility used in the climate model as well as in the target criterion ([Keller et al. (2004)]).

In the framework of decision making under uncertainty, one wishes to hedge against an undesirable future outcome. Thus the present level of uncertainty on the future outcome obviously influences the hedge level chosen at the present time. In a learning situation in which uncertainty is expected to evolve in the future, it is not as obvious whether and how the expected trajectory of uncertainty should affect the present decision.

Several studies ([Keller et al. (2004),Webster et al. (2008)]) have explicitly approached this question by comparing the optimal policy found under static uncertainty (i.e., no learning) and under decreasing uncertainty (i.e., reassuring learning, in the terminology proposed herein). In spite of differences in assumptions and methods, the conclusions of these studies are qualitatively consistent: the expectation of a future decrease of uncertainty is found to influence policy when its aim is to avoid a dangerous threshold, and to have a negligible influence when the cost-benefit objective function is smooth and has no such threshold. In the former case, the level of abatement is significantly reduced if the uncertainty is expected to decrease.

So far, though, there have been no studies that investigated the influence on optimal mitigation policy of an expectation of increasing uncertainty — i.e., of disconcerting learning in the present terminology — or even given a more general expectation of uncertain future uncertainty. Nevertheless, it is quite plausible, given the results of previous studies, that such expectations should greatly affect optimal policy design.

Indeed, since we might expect — given a continued IPCC learning process with a 6–7-year cycle — that the uncertainty may get higher before it gets lower, one could argue that it is worthwhile to "buy some time" for this hectic learning process to reach its final target of full certainty, at which time wiser, optimally informed decisions are more likely to be reached. In the present context, buying some time could, however, mean one of two things: either (i) enforcing higher abatements so as to delay an irreversible climate catastrophe [Ghil (2001),Lenton et al. (2008), Zaliapin and Ghil (2010)]; or, (ii) to the contrary, take care of other, possibly more urgent problems while the learning goes on, with still-growing or fluctuating uncertainties [Hillerbrand and Ghil (2008)]. Given the divergence of opinions on such a momentous decision, it is imperative to go beyond the speculative reasoning in this section and apply systematically the learning-theoretical framework introduced in Sections 2–4, in combination with the risk-management type of analyses cited in the present section.

## 6 Discussion and conclusions

[Oppenheimer et al. (2008)] (cited as [ONW08] throughout the present paper) introduced a probabilistic definition of learning in the context of scientific research on environmental problems. These authors showed that learning does not necessarily lead to truer beliefs, a situation they termed *negative learning.*

We have extended this analysis here to show that learning does not necessarily lead to more certain beliefs either, a situation for which we introduced the term of *disconcerting learning.* Negative learning corresponds to an increase in PDF bias, disconcerting learning corresponds to an increase in PDF dispersion. We have shown that the latter differs from, and is not tied to, the occurrence of the former. In other words, learning may well result in a state of knowledge which is closer to the truth and yet more uncertain, cf. Fig. 1.

We have shown that this rather counter-intuitive situation typically arises when a surprising but inconclusive piece of evidence is found. In Section 3.1, we used the simple example of a medical screening test that gives a positive result as an illustration of this fact. Such medical evidence is definitely informative but it is surprising because a negative result is a priori more likely; at the same time, it is also inconclusive because false positives are quite common in screening tests. Thus the patient, once informed of the test result, definitely knows more about his or her health but is still left with greater uncertainty than before the test.

Motivated by this simple example, we introduced in Section 3.2 a probabilistic model based on reasonable assumptions about learning, and used it to confirm that disconcerting learning in general occurs as a result of surprising but inconclusive evidence at a particular step in the learning process. Furthermore, we narrowed in on this situation arising when the PDF that reflects the state of knowledge is asymmetric or has heavy tails (Fig. 2). We have shown that the dispersion of the trajectories of uncertainty as learning occurs — i.e. the uncertainty on the uncertainty — is high when disconcerting learning is prone to happen.

Finally, because pronounced asymmetry appears to be a pervasive feature of the PDF of climate sensitivity in our current state of knowledge [AR4], climate uncertainty is thus prone to remain high or to increase — even if and as climate science makes steady progress — and thus its future trajectory is itself highly uncertain. Whether or not this is good news remains to be seen.

At first, the news that substantial research efforts dedicated to improving our understanding of the climate system could potentially result in an increased uncertainty on the outcome of future climate change may sound rather discouraging. On the other hand, the present article also provides a rational justification for the fact that constant or even increasing uncertainty is perfectly compatible with steady scientific progress and improved knowledge of the climate system. In other words, our results suggest that the uncertainty on climate sensitivity should not be considered as an appropriate metric to monitor progress in climate science, as has sometimes been suggested.

Our discussion here emphasizes two characteristics of disconcerting learning. First, disconcerting learning is a possibility, not a fatality. Since the occurrence of a surprise is by nature unexpected — and thus unlikely — so is the occurrence of disconcerting learning. Second, when it does occur, disconcerting learning is a transient state that eventually ends, at least in our model of Section 3.

Indeed, the initial increase of uncertainty is caused by the inconclusive nature of the surprising evidence. As more reassuring evidence confirms what was at first a surprise, uncertainty will eventually decrease. Still, the process of uncertainty getting worse before it gets better is intrinsic to the progressive learning model that we introduced: a sufficiently large surprise can occur only once during progressive learning, because a second surprise of large amplitude cannot occur unless expectation moves away from the true value, i.e. unless we are engaged in negative learning. In other words, one cannot be surprised twice without being wrong at least once.

Once we allow for the possibility of negative learning, though, successive surprises — progressive and negative — become possible, causing the repetition in time or the lengthening of disconcerting learning episodes. While this was beyond our scope here, studying how uncertainty will evolve when taking negative learning into account is both interesting and relevant. As Fig. 4 shows, the model introduced in Section 4 is capable of exhibiting such behavior; it may thus be a good starting point to investigate the more complex learning dynamics that can occur when negative learning is a possibility.

Finally, we have considered in Section 4 a situation in which climate sensitivity is the only uncertain quantity at stake in the evolution of climate, and in which learning is driven by the mere observation of global temperature. It turns out that surprises may occur in the future evolution of our assessment of climate sensitivity, even in such a simple situation of linear deviations from a radiative equilibrium.

This being said, the framework applied in Section 4 here for illustration purposes only can be made more realistic. Indeed, in more detailed climate models there are various uncertain parameters and processes that are either independent of climate sensitivity — e.g., ocean heat take-up or aerosol forcing — but do influence the climate response, or actually determine climate sensitivity itself, e.g. cloud-radiative feedbacks.

On the other hand, global temperature is certainly not the only variable one can use to constrain climate sensitivity and additional observations should be added into the learning-process analysis. It would therefore be of interest to investigate in a probabilistic framework, like that of Section 4 here, how the combined and possibly nonlinear effect of simultaneous learning on these various uncertain quantities, by means of multiple observations, plays out. It is quite possible that the results of such studies might affect our conclusions with respect to the uncertain future of climate uncertainty, as well as lead to more definitive policy implications, as discussed in Section 5.

**Acknowledgements**
The authors would like to thank the French Centre National de la Recherche Scientifique (CNRS) and the Argentinean CONICET for their support of this collaboration. MG's work is partially supported by NSF grant DMS-0934426 and by US Department of Energy grant DE-SC0006694, while PN also acknowledges the ANR-funded AssimilEx and FP7 ACQWA projects. We thank Dr. Andres Farall for pointing out relevant references, and we are grateful to the Editor, to James Risbey and to two anonymous reviewers for their suggestions that helped us improve the original manuscript.

## References

[Abbot et al.(2011)] Abbot, D.S., Silber, M., Pierrehumbert, R.T. (2011) Bifurcations leading to summer Arctic sea ice loss. *Journal of Geophysical Research* 116, D19120.

[Allen et al.(2006)] Allen MR, Andronova N, Booth B, Dessao S, Frame D, et al. (2006). Observational constraints on climate sensitivity. In *Avoiding Dangerous Climate Change*, ed. HJ Schellnhuber, W Cramer, N Nakicenovic, T Wigley, G Yohen, pp. 281–89. Cambridge, UK: Cambridge Univ. Press

[Aristotle (40 B.C.)] Aristotle (40 B.C.) Posterior Analytics, in *Organon*, translated by Mure, G. R. G. (2007), The University of Adelaide Library, Adelaide, Australia.

[Arrow and Fisher (1974)] Arrow, R.K., A.C. Fisher (1974) Environmental preservation, uncertainty, and irreversibility, *Quart. J. Econ.* 88, 312–319.

[Bacon (1605)] Bacon, F. (1605) *The Advancement of Learning.*

[Bagnoli and Bergstrom (2005)] Bagnoli, M., Bergstrom, T. (2005) Log-concave probability and its applications. *Economic Theory*, 26, 455–469.

[Burdett (1996)] Burdett, K. (1996) Truncated means and variances. *Economics Letters*, 52, 263–267.

[Chen (2011)] Chen, J. (2011) A partial order on uncertainty and information. *J Theor Probab*, doi: 10.1007/s10959-011-0375-2.

[Chen et al. (2010)] Chen, J., van Eeden, C., Zidek, J.V. (2010) Uncertainty and the conditional variance. *Probability & Statistics Letters*, 80, 1764–1770.

[Crutzen and Oppenheimer (2008)] Crutzen, P., M. Oppenheimer (2008) Learning about ozone depletion. *Clim. Change*, 64:1–10.

[Darby and Mysak (1993)] Darby, M.S., and L.A. Mysak (1993) A Boolean delay equation model of an interdecadal Arctic climate cycle, *Climate Dyn.*, 8:241–246.

[Dijkstra and Ghil (2005)] Dijkstra, H. A., M. Ghil (2005) Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys*, 43, RG3002, doi:10.1029/2002RG000122.

[Dufresne and Bony (2008)] Dufresne, J. L., and Bony, S. (2008), An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *J. Climate*, 21, 5135–5144.

[Duong et al. (1997)] Duong M.H., M. Grubb, and J.-C. Hourcade (1997), Influence of socioeconomic inertia and uncertainty on optimal CO2-emission abatement, *Nature*, 390, 270–273.

[Ghil (2001)] Ghil, M. (2001) Hilbert problems for the geosciences in the 21st century, *Nonlin. Processes Geophys.,* 8, 211–222.

[Ghil et al. (1987)] Ghil, M., A.P. Mullhaupt, and P. Pestiaux (1987) Deep water formation and Quaternary glaciations, *Climate Dyn.*, 2, 1–10.

[Ghil et al. (2008)] Ghil, M., M.D. Chekroun, and E. Simonnet (2008) Climate dynamics and fluid mechanics: Natural variability and related uncertainties *Physica D*, 237, 2111–2126, doi:10.1016/j.physd.2008.03.036.

[Hannart et al. (2009)] Hannart A., J.-L. Dufresne, P. Naveau (2009) Why climate sensitivity may not be so unpredictable, *Geophys. Res. Lett.*, 36, L16707.

[Hawkins and Sutton (2009)] Hawkins E., R. Sutton (2009) The potential to narrow uncertainty in regional climate predictions, *Bull. Amer. Meteorol. Soc.*, 90, 1095–1107.

[Hillerbrand and Ghil (2008)] Hillerbrand, R., and M. Ghil (2008) Anthropogenic climate change: Scientific uncertainties and moral dilemmas, *Physica D*, 237, 2132–2138, doi:10.1016/j.physd.2008.02.015.

[Humphrey et al. (2002)] Humphrey L.L., M. Helfand, B.K.S. Chan, S.H. Woolf (2002) Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force, *Ann. Intern. Med.*, 137, 347–360.

[Keller et al. (2004)] Keller K, Bolker BM, Bradford DF (2004), Uncertain climate thresholds and optimal economic growth. *J. Environ. Econ. Mgmt.*, 48:723–741

[Keller and McInerney (2007)] Keller K., D. McInerney (2007) The dynamics of learning about a climate threshold. *Clim. Dyn.* 30:321–332.

[Kelly and Kolstad (1999)] Kelly DL, Kolstad CD (1999) Bayesian learning, growth, and pollution. *J. Econ. Dyn. Control* 23:491–518.

[Knutti and Hegerl (2008)] Knutti R., G. C. Hegerl (2008), The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nature Geoscience*, 1, 35–743.

[Kuhn (1962)] Kuhn T. S. (1962) *The Structure of Scientific Revolutions.* Univ. of Chicago Press, Chicago.

[Leach (2007)] Leach AJ (2007), The climate change learning curve. *J Econ Dyn Control*, 31(5):1728–1752.

[Lenton et al. (2008)] Lenton, T. M., et al. (2008) Tipping elements in the Earth's climate system, *Proc. Natl. Acad. Sci. USA*, 105, 1786–1793.

[Livina and Lenton (2012)] Livina, V. N. and Lenton, T. M. (2012) A recent bifurcation in Arctic sea-ice cover, *The Cryosphere Discuss.*, 6, 2621–2651.

[O'Neill et al. (2006)] O'Neill, B. C., et al. (2006) Learning and climate change. *Climate Policy*, 6, 5, 585–589.

[Oppenheimer et al. (2008)] Oppenheimer M., B. C. ONeill, M. Webster (2008) Negative learning, *Climatic Change*, 89,155–172.

[Rahmstorf (1999)] Rahmstorf S., Shifting seas in the greenhouse?, *Nature*, 399, 523–524.

[Roe and Baker(2007)] Roe, G. H., M. B. Baker (2007) Why is climate sensitivity so unpredictable?, *Science*, 318, 629–632.

[Roe and Baker(2011)] Roe, G.H., and M.B. Baker, (2011). Comment on Another look at climate sensitivity. *Nonlinear Processes in Geophysics*, 18, 125–127.

[Soden and Held (2006)] Soden, B. J., and Held, I. M. (2006), An Assessment of Climate Feedbacks in Coupled Ocean-Atmosphere Models, *J. Climate*. 19, 3354.

[Solomon et al. (2007)] Solomon, S., et al., Eds. (2007), *Climate Change 2007: The Scientific Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Eds., Cambridge University Press, Cambridge, UK, and New York, NY, USA.

[Stainforth et al. (2005)] Stainforth, D.A., et al. (2005), Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433, 403406.

[Stroeve et al. (2012)] Stroeve, J. C., V. Kattsov, A. P. Barrett, M. C. Serreze, T. Pavlova, M. M. Holland, and W. N. Meier (2012), Trends in Arctic sea ice extent from CMIP5, CMIP3 and observations, *Geophys. Res. Lett.*, 39, L16502.

[Tietsche et al. (2011)] Tietsche, S., Notz, D., Jungclaus, J.H., Marotzke, J. (2011), Recovery mechanisms of Arctic summer sea ice. *Geophysical Research Letters* 38 (2), L02707.

[Tol (1997)] Tol R.S.J. (1997), On the optimal control of carbon dioxide emissions: an application of FUND, *Environ Model. Assess.*, 2, 151–163.

[Webster et al. (2008)] Webster M. D., L. Jakobovits, J. Norton (2008) Learning about climate change and implications for near-term policy. *Clim Change*, 89, 67–85.

[Zaliapin and Ghil (2010)] Zaliapin I., M. Ghil (2010), Another look at climate sensitivity, *Nonlin. Processes Geophys.*, 17, 113–122.

[Zidek and van Eeden (2003)] Zidek, J.V., van Eeden, C. (2003). Uncertainty, entropy, variance and the effect of partial information. In *Mathematical Statistics and Applications: Festschrift for Constance van Eeden*. Eds. M. Moore, S. Froda, and C. Leger. Lecture Notes–Monograph Series, Institute of Mathematical Statistics, 42, 155–167.

| | $p_1(x^*) < p_0(x^*)$ ~ increasing bias | $p_1(x^*) > p_0(x^*)$ ~ decreasing bias |
|---|---|---|
| $\sigma_1 < \sigma_0$ = decreasing uncertainty | Reassuring Negative Learning | **Reassuring Progressive Learning** |
| $\sigma_1 > \sigma_0$ = increasing uncertainty | Disconcerting Negative Learning | **Disconcerting Progressive Learning** |

**Fig. 1** Schematic diagram of the four different learning situations that result from the definitions introduced in Oppenheimer et al. [2008] — i.e., negative vs. progressive — and in the present article — i.e., disconcerting vs. reassuring. These four situations are mapped here with respect to the evolution of the bias and of the uncertainty in the probability density function (PDF), as they reflect the state of knowledge, according to Eqns. (2) and (3).

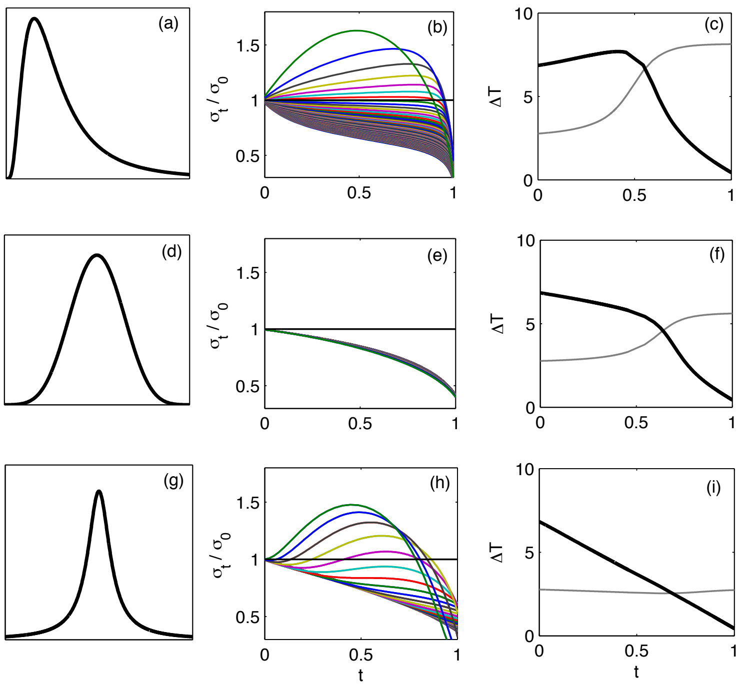


**Fig. 2** Evolution of the PDF $p(x)$ for three typical shapes of the initial PDF $p_0(x)$ — shown in panels (a, d, g) — to final convergence, when the spread vanishes. The spread $\sigma$ of $p(x)$ is normalized in panels (b, e, h) to its initial value $\sigma_0$ as learning occurs. For a given initial PDF, each trajectory is associated with a different true value towards which the progressive learning model of Eq. (SM3) converges; one hundred trajectories are thus plotted for each initial PDF, using its percentiles as true values. (a, b) Gamma PDF, asymmetric, with exponential tail; (d, e) Gaussian PDF, symmetric, with thin tails; and (g, h) Cauchy PDF, symmetric, with heavy tails. Evolution of the PDF parameters is plotted for the initial PDFs of Frame [2005] (c, f, i): percentile range 5%–95% (solid line) and median (gray line) of the climate sensitivity $\Delta T$, for three different true values $\Delta T^*$: (c) $\Delta T^* = 8°C$, (f) $\Delta T^* = 5.5°C$, and (i) $\Delta T^* = 3°C$.

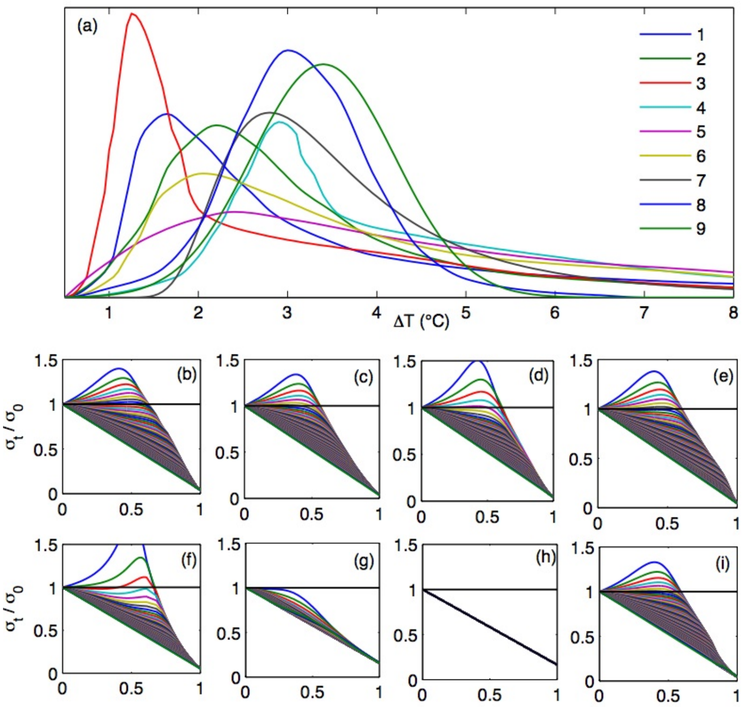


**Fig. 3** (a) Nine PDFs of climate sensitivity reported by [AR4], obtained by different teams and using diverse data and methods: (1 through 5) PDFs constrained by the transient evolution of the atmospheric temperature, radiative forcing and ocean heat uptake; (6,7) constrained by present-day climatology; and (8, 9) unweighted or fitted distributions from different models or from perturbing parameters in a single model; see text for details. Evolution of the PDFs $p(x)$ for the nine initial PDFs $p_0(x)$; same treatment as in Fig. 2. For a given initial PDF, each trajectory is associated to a different true value towards which the progressive learning model of Eq. (SM3) converges. One hundred trajectories are plotted for each initial pdf using its percentiles as true values. (b, 1) [Forster and Gregory (2006)], (c, 2) [Gregory et al. (2002)], (d, 3) [Frame et al. (2005)], (e, 4) [Hegerl et al. (2006)], (f, 5) [Andronova and Schlesinger (2001)], (g, 6) [Forest et al. (2006)], (h, 7) [Roe and Baker (2007)], (i, 8) [Knutti et al. (2002)], (9) [Raisanen (2005)].

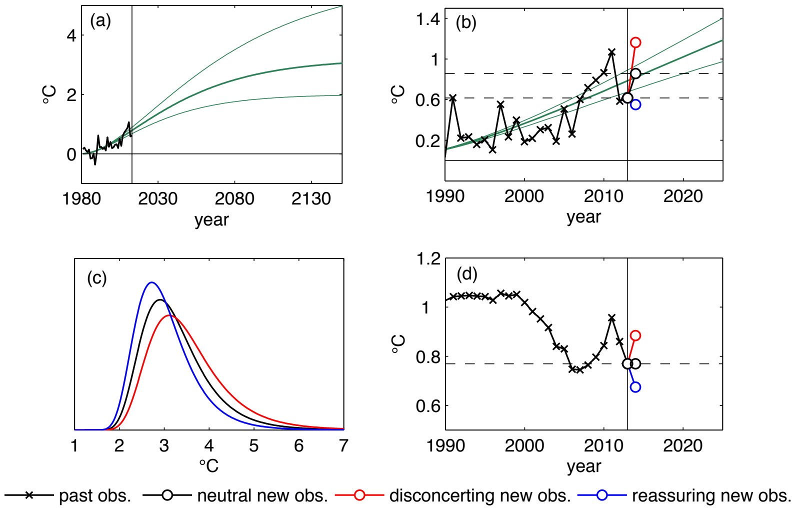


**Fig. 4** (a) Trajectories of Earth's global temperature anomaly: smooth component $\overline{T}_t$ simulated by the energy balance model of Eq. (7) over the time interval 1980–2150, for $\Delta T = 2°$C, 3.2°C and 6°C (green lines); and observations $T_t$ simulated over the interval 1980–2012 for $\Delta T = 3.2°$C (black line). (b) Same as (a) zoomed on the time period 1980-2025, with three possible new observations added in 2013: strong warming (red line and circle, disconcerting learning), moderate warming (dark line and circle, reassuring learning), moderate cooling (blue line and circle, reassuring learning). (c) Posterior PDFs of climate sensitivity after updating based on each of the three new 2013 observations. (d) Trajectory of the uncertainty on climate sensitivity $\sigma_t$ over the observational interval 1990–2012 (black line), and new uncertainty value, after updating based on each of the three new 2013 observations.