Climate sensitivity: should the climate tail wag the policy dog?

Gerard H. Roe · Yoram Bauman

Abstract The small but stubbornly unyielding possibility of a very large long-term response of global temperature to increases in atmospheric carbon dioxide can be termed the fat tail of high climate sensitivity. Recent economic analyses suggest that the fat tail should dominate a rational policy strategy if the damages associated with such high temperatures are large enough. The conclusions of such analyses, however, depend on how economic growth, temperature changes, and climate damages unfold and interact over time. In this paper we focus on the role of two robust physical properties of the climate system: the enormous thermal inertia of the ocean, and the long timescales associated with high climate sensitivity. Economic models that include a climate component, and particularly those that focus on the tails of the probability distributions, should properly represent the physics of this slow response to high climate sensitivity, including the correlated uncertainty between present forcing and climate sensitivity, and the global energetics of the present climate state. If climate sensitivity in fact proves to be high, these considerations prevent the high temperatures in the fat tail from being reached for many centuries. A failure to include these factors risks distorting the resulting economic analyses. For example, we conclude that fat-tail considerations will not strongly influence economic analyses when these analyses follow the common—albeit controversial—practices of assigning large damages only to outcomes with very high temperature changes and of assuming a significant baseline level of economic growth.

G. H. Roe (✉)
Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA
e-mail: gerard@ess.washington.edu

Y. Bauman
Program on the Environment, University of Washington, Seattle, WA, USA

Y. Bauman
Institute of Global Low-Carbon Economy, University of International Business
and Economics, Beijing, China

 Springer

1 The fat tail of climate sensitivity

Climate sensitivity—the long-term response of global-mean, annual-mean surface temperature to a doubling of the atmospheric carbon dioxide concentration above pre-industrial values—has long been a benchmark by which to compare different estimates of the planet's climatic response to changes in radiative forcing. A doubling of carbon dioxide increases the radiative forcing by about $\Delta R_{2\times} \approx 4$ Wm^{-2}. In a remarkable analysis, Svante Arrhenius (1896) estimated that the equilibrium global-mean temperature would increase by $\Delta T_{2\times} \approx 5$ °C. A major reassessment came in 1979 with the National Research Council Charney Report. Reviewing the intervening advances in science, Charney (1979) estimated $\Delta T_{2\times}$ between 1.5 and 4.5 °C (described as the 'probable error'), a range that has changed only incrementally ever since.

Modern estimates provide more precise probability density functions (PDFs). The Intergovernmental Panel on Climate Change (IPCC 2007) report gives a 'likely' range (2-in-3 chance) of $\Delta T_{2\times}$ lying between 2 and 4.5 °C, and concludes it is 'very unlikely' (< 1-in-10) to be less than 1.5 °C. This summary is consistent with a peculiarity of a large number of other studies—the small but non-trivial possibility of $\Delta T_{2\times}$ being much larger than the canonical 1.5 to 4.5 °C range. One of the important achievements in recent climate science has been to establish great confidence in the lower bound, but the upper bound has proven much less tractable (e.g., Knutti and Hegerl 2008). Allen et al. (2006) have argued that this fat tail of possibly-high climate sensitivity is a fundamental feature of $\Delta T_{2\times}$ estimates from observations, and Roe and Baker (2007) have argued that it is a fundamental feature of $\Delta T_{2\times}$ estimates from numerical climate models. It will require improbably large reductions in uncertainties about the radiative forcing the planet has experienced—or, equivalently, in our uncertainty about physical feedbacks in the climate system—to substantially remove the skewness. Some studies (e.g., Annan and Hargreaves 2006) have combined multiple estimates in a Bayesian framework and argued this can yield narrower and less skewed distributions for $\Delta T_{2\times}$. However the answers are critically sensitive to how independent the different estimates are, and what the Bayesian prior assumptions are. Both factors are fiendishly elusive to pin down objectively. Knutti and Hegerl (2008) and Knutti et al. (2010) provide a good review of the scientific issues involved. Efforts continue in this direction (Annan and Hargreaves 2009) but, overall, it seems prudent to assume that estimates of $\Delta T_{2\times}$ will not change substantially for the foreseeable future.

Several macroeconomic analyses of the costs of climate change have argued that there is a sting in the fat tail of $\Delta T_{2\times}$ (e.g., Weitzman 2009a, b, 2010). This is seen as follows: if, as is reasonable to assume, climate damages increase nonlinearly with temperature, then the tail of the PDF is weighted more strongly than the middle of the PDF in calculations of the expected damages. Moreover, if damages are highly nonlinear with temperature, one could argue that even the very smallest chance of absolutely apocalyptic consequences might properly dominate a rational policy. A close analogy is the St. Petersburg paradox, a coin-flipping wager with an expected value dominated by low-probability, high-valued outcomes. The extent to which this "fat-tail" argument is a worry depends in part on how rapidly the global mean temperature might rise towards its equilibrium value. This is both because impacts and costs of adaptation depend on the rapidity of change and because economic analyses discount future damages. Millner (2011) performs a comprehensive analysis

of the interplay between temperature uncertainty, climate damages, and welfare under which Weitzman's conclusions hold.

The primary message in this paper is that there are important physical constraints on the climate system that limit how fast temperatures can rise. This does not mean that high temperatures cannot be reached on timescales that can impact society: if the system is forced strongly enough they are guaranteed. Nor does it mean that climate-change damages cannot be high: one can always pick a damage function that will achieve that. But it does mean that the uncertainties surrounding economic growth assumptions and climate damage functions are likely to be much more important than those associated with "fat-tailed" climate sensitivities.

2 The transient evolution of possible future climates

A key player in the physical system is the enormous thermal inertia represented by the deep ocean. The climate system cannot reach a new equilibrium until the deep ocean has also reached equilibrium. In response to a positive climate forcing (i.e., a warming tendency), the deep ocean draws heat away from the surface ocean, and so buffers the surface temperature changes, making them less than they would otherwise be. The deep ocean can absorb enormous amounts of heat, and not until this reservoir has been exhausted do the surface temperatures attain their full, equilibrium values.

A second key player is the inherent relationship between feedbacks and adjustment timescale. If it transpires that we do, in fact, live on a planet with a high climate sensitivity, it will be because the planet has strong positive feedbacks. In other words, the net effect of all of the dynamic processes (clouds, water vapor, ice reflectivity, etc.) is to strongly amplify the planet's response to radiative forcing. In this event, it would mean the planet is inefficient in eliminating energy perturbations: a positive feedback reflects a tendency to retain energy within the system, inhibiting its ultimate emission to space, and therefore requiring a larger temperature response in order to achieve energy equilibrium. Moreover, it is generally true that, all else being equal, an inefficient system takes longer to adjust than an efficient one. This behavior is absolutely fundamental and widely appreciated (e.g., Hansen et al. 1985; Wigley and Schlesinger 1985): as time progresses, more and more of the ocean abyssal waters become involved in the warming, and so the effective thermal inertia of the climate system increases. Hansen et al. (1985) solve a simple representation of this effect and show that the adjustment time of climate is proportional to the square of climate sensitivity. The equilibration process is an asymptotic one, so the adjustment time characterizes the time it takes to achieve some fraction (typically $1 - 1/e$, or ~ 63 %) of its equilibrium value. The relationship between response time and climate sensitivity means that if it takes 50 yrs to achieve that fraction with a climate sensitivity of 1.5 °C, it would take 100 times longer, or 5,000 yrs, to achieve that fraction if the climate sensitivity is 15 °C. Although Nature is of course more complicated than this (see, e.g., Gregory 2000), the basic picture described here is reproduced in models with a more realistic ocean circulation (Held et al. 2010).

In the context of the PDF of $\Delta T_{2\times}$, the relationship with the response time has been reviewed in Baker and Roe (2009). Their model represents both atmospheric feedbacks and the uptake of heat by the upper and deep ocean. A schematic

illustration is shown in Fig. 1, and is very briefly described here. The evolution of global-mean, annual-mean surface temperature, as represented by the ocean mixed layer, is governed by:

$$\rho C h \frac{dT}{dt} + \frac{T(1 - f_a)}{\lambda_0} - \kappa \frac{\partial T_{th}}{\partial z}\big|_{z=0} = \Delta R_F(t), \tag{1}$$

where ρ and C are the density and specific heat capacity of water, h is the mixed-layer depth, and f_a is the sum of atmospheric feedbacks. λ_0 reflects the climate sensitivity in the absence of atmospheric feedbacks (i.e., if $f_a = 0$), following the definitions given in Roe (2009). κ is the thermal conductivity governing the heat flux across the thermocline (at $z = 0$), which also depends on the vertical temperature gradient.

Reflecting uncertainties in observations, uncertainties in climate feedbacks are assumed to be Gaussian distributed (e.g., Allen et al. 2006; Roe and Armour 2011). We take a mean value for f_a of 0.65, and a standard deviation of 0.13. This range is fully consistent with observed estimates of $\Delta T_{2\times}$ (e.g., Allen et al. 2006), with the multi-thousand model experiments of the innovative *climateprediction.net* program (Stainforth et al. 2005), and with the IPCC ranges of $\Delta T_{2\times}$. For instance, the range means that about 75 % of possible $\Delta T_{2\times}$s are less than ~ 4.5 °C (as seen in, for example, Fig. 2).

Equation 1 assumes that the feedbacks are linear (i.e., $f_a \neq f_a(T)$). Several studies have explored how well this assumption holds. In one illustrative case, Colman and McAvaney (2009) apply an extreme forcing of thirty-two times pre-industrial CO_2, and find a 40 % decrease in $\Delta T_{2\times}$. However, results across different climate models are not consistent, even about the sign of the change in $\Delta T_{2\times}$. In a survey of a dozen studies, Roe and Armour (2011) find that the magnitude of df_a/dT is less (and typically much less) than 0.04 K^{-1}. It would be straightforward to include such nonlinearities in the framework presented here, though Roe and Armour (2011) conclude that the impact on the shape of the PDF of $\Delta T_{2\times}$ was small.

Temperature in the deep ocean, T_{th}, is governed by a balance between diffusion and upwelling:

$$\frac{\partial T_{th}}{\partial t} = \chi \frac{\partial^2 T_{th}}{\partial z^2} - w \frac{\partial T_{th}}{\partial z}, \tag{2}$$

Fig. 1 Schematic illustration of the simple global-mean climate model of Baker and Roe (2009)
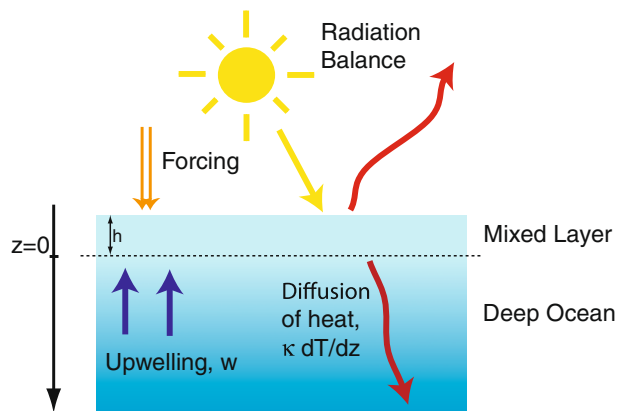
where w is the upwelling rate and χ is the diffusivity. The model is virtually identical to a host of other equivalent models, which have been shown to be fully capable of emulating more complete numerical models, and also historical observations at the global scale (e.g., Raper et al. 2001). Such models are regularly used to make long-term climate predictions (e.g., IPCC 2007). Their flexibility and computational efficiency means that parameters can be easily varied and that uncertainties can be fully explored. The equations and parameters follow those of Baker and Roe (2009), which matches best current estimates of heat uptake and climate feedbacks. The only change from Baker and Roe (2009) is a χ of $2.5 \times 10^{-4}\,\mathrm{m^2 s^{-1}}$ instead of $2.0 \times 10^{-4}\,\mathrm{m^2 s^{-1}}$, as it was found to produce slightly better agreement for historical trends. Results are insensitive to plausible variations in these climate parameters.

2.1 A numerical example: a doubling of CO_2

We begin by presenting the climate response to an idealized, instantaneous doubling of carbon dioxide. In such a scenario the envelope of possible temperature responses must ultimately evolve in time into the equilibrium climate sensitivity distribution. The reasons for choosing this initial scenario are two-fold. Firstly, climate sensitivity has been identified as a key uncertainty in integrated assessment models (IAMs) (e.g., IAG 2010) and this idealized step-function forcing clearly shows the long timescales involved in the development of the fat tail of climate sensitivity. This clarity can be obscured when looking at more complicated scenarios. Secondly, it has itself been used in some economic evaluations (e.g., Weitzman 2010). Later we consider a more realistic forcing scenario.

The dramatically long timescale for the development of the fat tail is seen in Fig. 2a, which shows the evolution of the envelope of possible climate trajectories. Note the logarithmic time axis after 500 years. The shading represents the one-, two-,
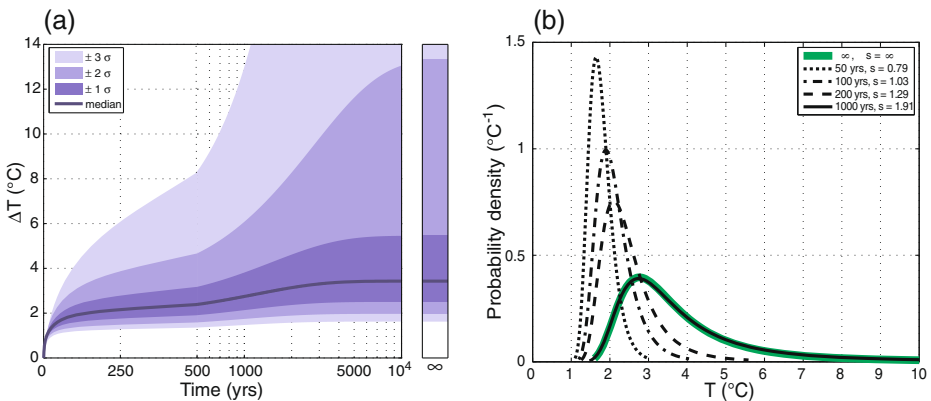

Fig. 2 a The time evolution of uncertainty in global temperature in response to an instantaneous doubling of CO_2 at $t = 0$, and for standard parameters. The *shading* reflects the range of feedbacks considered (symmetric in feedbacks, but not in climate response), as explained in the text. Note the change to a logarithmic *x*-axis after t = 500 yr. The panel illustrates that for high climate sensitivity it takes a very long time to come to equilibrium. **b** The shape of the distribution at particular times. The skewness of the distributions are also shown in the legend; as described in the text, the upper bound on possible temperatures is finite at finite time, limiting the skewness

and three-standard deviation ranges for climate sensitivity, encompassing 68.3, 95.5, and 99.7 % of possibilities, respectively. The Gaussian uncertainty in feedbacks leads to uncertainties in the evolving climate response that are also Gaussian initially, but which become highly skewed over time (e.g., Baker and Roe 2009).

The extreme limits of the range we consider would imply that the climate system is unstable. Our 3σ value for f_a is 1.04. For $f_a < 1$, the second term in Eq. 1 acts as a sink of energy, allowing a finite temperature response to a finite climate forcing: for $f_a = 1$, no sink exists, and the climate sensitivity is infinite; for $f_a > 1$ the second term changes sign and becomes a source of energy to the system, which as a result is unstable. Figure 2a therefore illustrates that even for a planet that is formally headed
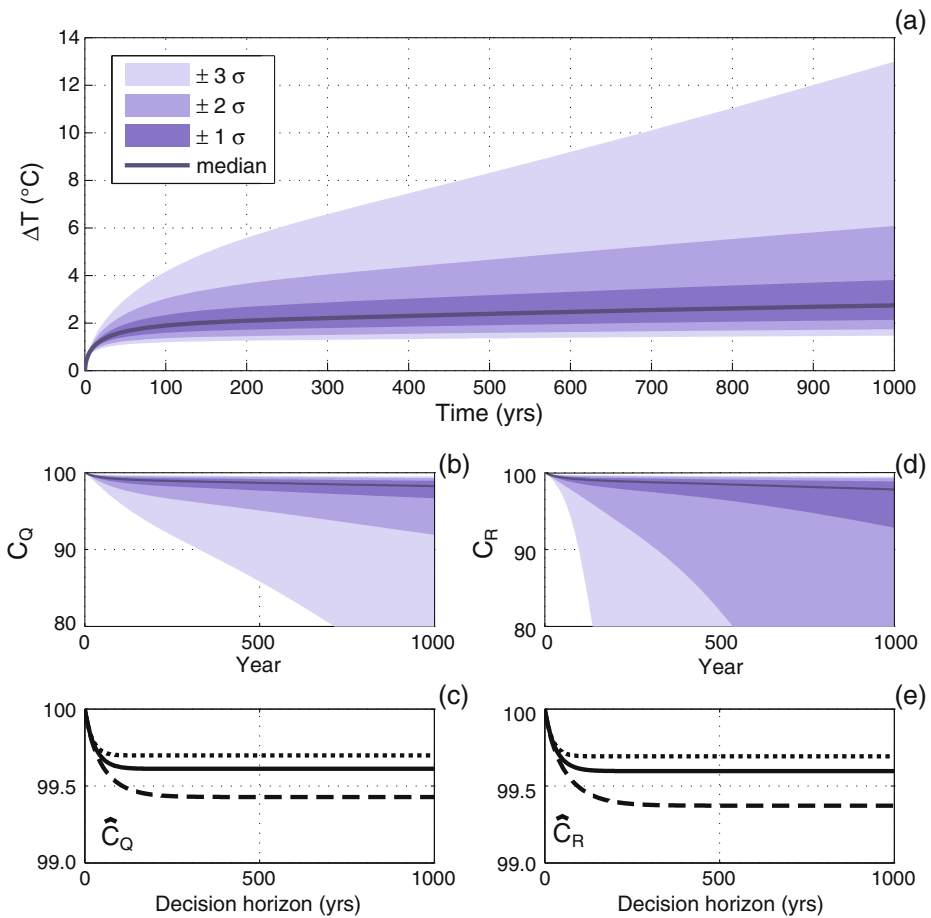

Fig. 3 A closer look at the next thousand years, for an instantaneous doubling of CO_2 at $t = 0$: **a** Evolution of the PDF of possible future climates, $h(\Delta T(t))$; **b** evolution of the PDF of possible climate damages, using the quadratic damages function, $h(C_Q)$; **c** welfare-equivalent consumption, $\hat{C}_Q$, as a function of the decision horizon. For all curves, $g = 2 \%$ yr^{-1} and calculations using $\eta = 2, 3, 4$ are shown as *dashed*, *solid*, and *dotted lines*, respectively. Panels **d** and **e** are the same as panels **b** and **c**, but using the reactive damages function, C_R. Even for the 3σ possibility of climate sensitivity, and for the reactive damages function, the willingness-to-pay remains at less than 1 % of equivalent consumption. *Shading* denotes confidence intervals as in Fig. 2a

to oblivion, it can take a very long time to get there because of the ocean's capacity to absorb heat.

The consensus of the climate science community would be that the likelihood we are actually on a runaway greenhouse trajectory (i.e., $f_a \geq 1$) is vanishingly small (e.g., Solomon et al. 2010). In other words, we are effectively considering a maximum possible upper bound on the probabilities residing within the fat tail. Because of this, we truncate the PDF of feedbacks beyond this $\pm 3\sigma$ range. Climate feedbacks exceeding the upper bound of this range should not be seriously entertained, and climate feedbacks less than the lower bound make no impact of any consequence on the analyses.

Let $h(\Delta T(t))$ stand for the PDF of possible future global mean temperatures, $\Delta T(t)$, as they evolve over time. The abiding impression from Fig. 2a is that low ΔT trajectories rapidly adjust to their equilibrium values over a few decades or a century, whereas those at the high end take thousand of years even to approach their equilibrium values. For the reasons given above, climate trajectories stay relatively tightly bunched over the course of the first few centuries, diverging only slowly thereafter, and all of that divergence occurs toward higher temperatures.

Time slices of $h(\Delta T)$ are shown in Fig. 2b. At $t = 0$ the shape is actually Gaussian; it acquires skewness only gradually (e.g., Baker and Roe 2009). At $t = \infty$, the skewness of the distribution is infinite, but even 1000 years after doubling CO_2, the skewness remains less than two. For any finite time, there is an upper limit on the temperature set by how much energy has been accumulated within the system, which can be found by integrating Eqs. 1 and 2. For the highest value of f_a that we consider, $f_a = 1.04$ (which as already noted would imply a slightly unstable climate system), this upper limit is the upper edge of the outermost shaded regions in Figs. 2a, 3a, and 4b. The ultimate, though utterly absurd, upper limit is set by the rate at which the Earth receives energy from the Sun. Infinite temperature cannot be realized in finite time and so by the strict statistical definition, that the PDF asymptotes as a power law in the limit of $T \to \infty$, there can be no fat tail at finite time.

3 Economic implications

How does the evolution of the climate tail affect calculations of the cost of damages associated with climate change? The analysis that follows draws heavily on Weitzman (2010), which provides a simple and incisive framework for calculating the costs of avoiding or insuring against climate change. In order to explore the significance of the fat tail, Weitzman suggests comparing two different "climate damages" functions. The first is a quadratic function of global mean temperature:

$$C_Q = \frac{1}{1 + \alpha \Delta T^2}, \tag{3}$$

C_Q is the "welfare equivalent" consumption as a fraction of what consumption would have been were $\Delta T = 0$. Weitzman takes the rather precise value of $\alpha = 2.388 \times 10^{-3}$ in order to match the damage function of Nordhaus (2008), for which $C_Q(2\,°C) \approx 99\,\%$. He also considers a "reactive" damages function with an extra term:

$$C_R = \frac{1}{1 + \alpha \Delta T^2 + \beta \Delta T^\gamma}. \tag{4}$$

By choosing $\beta = 5.075 \times 10^{-6}$ and $\gamma = 6.754$, Weitzman thus creates a damages function that has significant consequences for higher global-mean temperatures ($C_R(6\ °C) = 50\ \%$, and $C_R(12\ °C) = 1\ \%$). The functional choices and parameters are chosen purely for illustrative purposes.[1]

What should be our "willingness to pay" to avoid the climate damages associated with either C_Q or C_R, or to insure against them? Weitzman defines utility, or welfare, as a simple function of consumption, $U(C)$, with initial consumption normalized to 1; and also assumes an underlying long-term exponential economic growth rate, g. In a world experiencing climate change, consumption accelerates because of the long-term growth rate but is also retarded by the climate damages that accompany warming. Thus for every possible trajectory of future temperature, $\Delta T(t)$, there are accompanying trajectories for consumption, $C(t)$, and welfare, $U(t)$.

Weitzman then compares this world to one in which consumption is reduced by a constant fraction $(1 - \hat{C})$ that is potentially used to avoid or insure against climate changes. In this world, the fractionally reduced consumption $\hat{C}$ still grows at rate g, but does not suffer climate damages. $\hat{C}$ can be found by calculating the integrated welfare of these two worlds up to a decision horizon ($\equiv t_h$, the duration of time into future that is relevant for decision making), and equating them:

$$\int_0^{t_h} U(\hat{C}e^{gt})dt = E\left[\int_0^{t_h} U(C_{Q,R}(\Delta T(t))e^{gt})dt\right], \tag{5}$$

where E[ ] denotes the statistical expectation operator. In principle then, $1 - \hat{C}$ ought to be the upper bound on our willingness to pay, in terms of a fraction of consumption, in order to avoid or insure against climate damages. Following reasonably standard practices, Weitzman sets $U(C) = C^{1-\eta}/(1 - \eta)$, and chooses $\eta = 3$ and $g = 2\ \%\ yr^{-1}$. The rate of pure time preference, ρ, is set to zero, meaning that the welfare of future generations (within the time horizon of the model, t_h) are given equal weight to the current. The effect of these choices is that future reductions in consumption are discounted at an exponential rate of $\eta g = 6\ \%\ yr^{-1}$ because future generations are richer than the current generation and therefore suffer a smaller loss in utility from reduced consumption at the margin.

3.1 A numerical example: the impact of a CO_2 doubling

Section 2 reviewed the strong physical constraints on how $h(\Delta T(t))$ evolves with time. From Eqs. 3 to 5 the evolving PDFs of $C_{Q,R}$ can be calculated, and are shown in Fig. 3b and d. Because the median of $h(\Delta T)$ remains below 3 °C over the whole millennium, there is little difference between the medians of C_Q and C_R. At the far end of the fat tail of possibilities, the 3σ trajectory reaches 6 °C at around 250 years which, were it to transpire, would cause both C_Q and C_R to be significantly reduced (for instance, $C_R(3\sigma) = 47\ \%$ at year 250).

When cast in terms of welfare-equivalent consumption (i.e., $\hat{C}_{Q,R}$), we reach two important conclusions. Firstly, we see from Fig. 3c and e that there is only a small difference between $\hat{C}_Q$ and $\hat{C}_R$. Because of the exponential discounting of the

---

[1]Note that in all cases ΔT is measured in °C, and so the units of α and β depend on the choice of exponents. This could be avoided by dividing T by a normalizing temperature.

future, both $\hat{C}_Q$ and $\hat{C}_R$ asymptote to a constant value, and change little for decision horizons beyond 150 years (reflecting the exponential decay of the weighting given to future damages). The results hold for the three different values of η considered ($\eta = 2, 3$, and 4).

The second and arguably more important conclusion is that willingness-to-pay is under 1 % for even the most extreme climate possibilities. This is largely because long periods of time are required to reach high temperatures regardless of the long-run climate sensitivity. Saying that $C_R(3\sigma) = 47$ % at year 250 sounds impressive, but with consumption growing at 2 % per year this simply means that climate change will reduce consumption from 141 times current consumption to only 66 times current consumption. Under the assumption of diminishing marginal utility of consumption it is not surprising that intertemporal optimization comes to the conclusion that current generations should be unwilling to sacrifice more than 1 % of consumption for the sake of future millionaires.
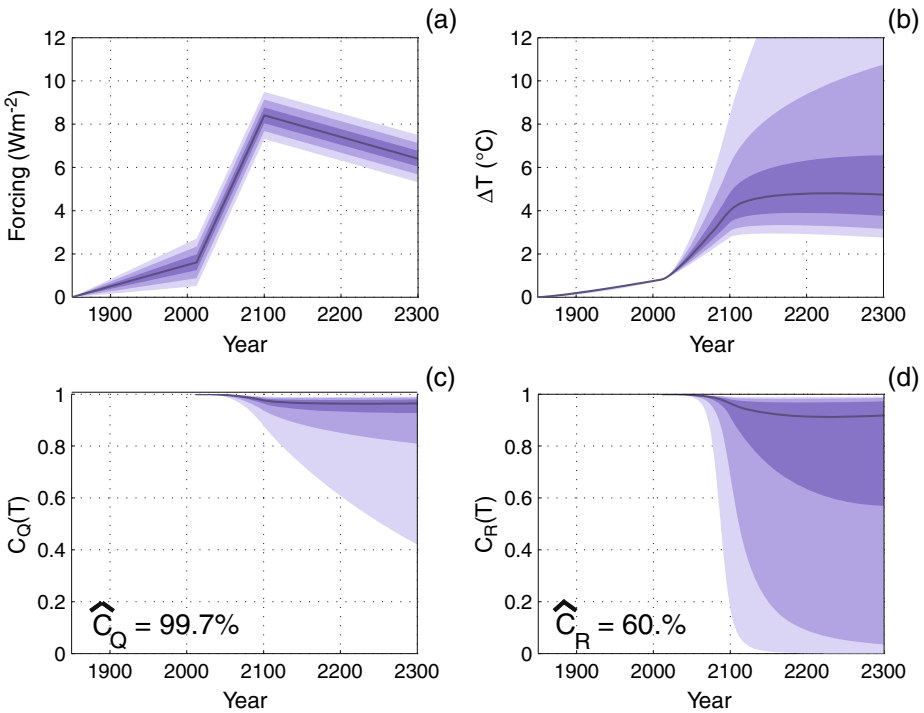

Fig. 4 As for Fig. 3, but for a more realistic climate forcing scenario. Starting at 1850, past forcing (i.e., pre-2011) is consistent with observations given in IPCC AR4 (IPCC 2007). Future forcing through 2100 is approximately consistent with the IPCC A2 scenario (IPCC 2007). Thereafter climate forcing declines at a rate chosen to approximately stabilize temperatures for the median value of climate sensitivity. Note that uncertainty in past forcing must be matched to uncertainty in climate feedbacks, in order to match the well-constrained temperature record. Because of this complementary uncertainty in forcing and climate sensitivity, a lower wedge of forcing uncertainty matches with an upper wedge of uncertainty in climate response. See text for more details. Damages in panels **c** and **d** are calculated with respect to a baseline of 2011. *Shading* denotes confidence intervals as in Fig. 2a

4 A more realistic scenario

Up to this point we've used a doubling of CO_2 as an illustrative climate forcing because of its long history as a benchmark measure of climate change, and because of its relationship to climate sensitivity. We next consider a more realistic forcing scenario that, while still idealized, captures important aspects of the real situation: (1) past climate forcing is approximately consistent with past observations and their uncertainty, (2) future forcing is based on a do-nothing, business-as-usual scenario for the 21st century, similar to the IPCC A2 scenario, (3) economic analyses and climate damages are calculated using the present day (as of writing, 2011) as a baseline, and (4) the climate model is correctly initialized for the current state of the climate system, accounting for the correlated uncertainty in forcing and climate sensitivity. This last aspect, which is omitted in all IAMs that we are aware of, contributes to how soon high temperatures can be realized in the future. Again, the essential result is that the fat tail of climate sensitivity plays only a weak role in climate projections on economically relevant time scales.

In order to describe this more realistic scenario, we must first detour briefly into the causes of uncertainty in predictions of future climate. The main points of this paragraph are summarized in Roe (2010) and Armour and Roe (2011), and are based on, and entirely consistent with, the IPCC AR4, and prior studies such as Forest et al. (2002) and Knutti et al. (2002). Uncertainty about the future state of the climate can be decomposed into two main factors: (1) uncertainty in future emissions, and (2) uncertainty in how the climate will respond to those emissions. For the latter, the community uses climate models that have been calibrated against past observations of global climate which, in turn, must satisfy the observed global energy balance. Many parts of this global energy balance are well constrained, including the ocean heat uptake, the global-mean temperature change, and the radiative forcing due to increases in CO_2 and other greenhouse gases. The two largest sources of uncertainty are the aerosol radiative forcing and the climate sensitivity. Anthropogenic aerosols—airborne particles, either solid or liquid—have a direct impact on the Earth's radiative budget and also an indirect impact via their influence on clouds. The effect of aerosols is almost certainly negative, but the uncertainty is large: IPCC AR4 states the aerosol forcing as $\sim -1.3 \pm 0.5(1\sigma)\mathrm{Wm}^{-2}$ for the change since the pre-industrial. This negative forcing masks a significant portion of the positive forcing from the better constrained greenhouse gases ($\sim 2.9 \pm 0.2(1\sigma)\mathrm{Wm}^{-2}$). The key point is that uncertainties in climate sensitivity and climate forcing are inherently equivalent: if aerosol forcing has been strongly negative, the net climate forcing has been weak, and therefore the climate sensitivity must be high in order for us to have experienced the warming we have; on the other hand, if aerosol forcing has been weak, net forcing has been more strongly positive and a low value of climate sensitivity is implied. Thus climate forcing and climate sensitivity must be treated as complementary pairs: if one is high the other is correspondingly low.

Figure 4a shows a linear approximation to past forcing from 1850 to the present, including forcing uncertainty. The net climate forcing for the present day (radiative forcing minus ocean heat uptake) is assumed to be Gaussian distributed. For each possible forcing trajectory, a complementary climate sensitivity is chosen (or equivalently, f_a in Eq. 1) such that the climate model reproduces the well-constrained temperature history from 1850 to the present (Fig. 4b). Between 2011 and 2100, the

climate forcing is chosen to emulate the IPCC AR2 scenario, which implies a strong increase to about 8 Wm^{-2} by 2100. Thereafter a decline in forcing is assumed at a rate which approximately stabilizes the temperature for the median trajectory. Calculations are continued through to 2300. Uncertainty in aerosol forcing is assumed to be constant though, in principle, future improvements in understanding would enable us to narrow the uncertainty in forcing and hence in the accompanying climate sensitivity. Figure 4b shows the temperature response to this scenario. Consistent with the relationship between climate sensitivity and adjustment time, trajectories with low climate sensitivity equilibrate rapidly to changes in forcing, and those with high sensitivity respond much more slowly, remaining far from equilibrium. The median temperature change is 4 °C at 2100, but uncertainty in climate sensitivity leads to a diverging envelope of possible future temperatures after 2011, with a 3σ range of about 3 to 8 °C in 2100.

Should the temperatures towards the upper part of this range be considered as part of a "fat tail" of the climate response? Strictly speaking, the term fat tail has a formal statistical definition, which we've already shown cannot apply at any finite time. However the term is also used more colloquially as short-hand for the possibility of very large temperature changes in response to anthropogenic forcing. In this regard it is very important to clearly distinguish two aspects of the problem. As already noted "climate sensitivity" is defined as the equilibrium response to the (comparatively weak) climate forcing of a CO_2 doubling. For such a scenario it is reasonable, for instance, to refer to temperatures above the IPCC "likely" range (i.e., > 4.5 °C) as lying in the "fat tail" of Fig. 2 or 3a. In contrast, the business-as-usual scenarios that lead to the possibility of similarly high temperatures by 2100 (as shown in Fig. 4b) are very different: they are the near-term response to a very different and much larger climate forcing. Furthermore, as can be seen in Fig. 4b, the PDFs applying to such scenarios are not particularly skewed. For the term "fat tail" to retain its value, it should be reserved for, at minimum, much more strongly skewed PDFs.

In terms of climate damages, the quadratic damages function, C_Q, grows relatively slowly. The 3σ trajectory of climate damages reaches 20 % of consumption (i.e., $C_Q = 0.8$) at around 2120. This slow growth and the exponential discounting of future welfare produces a $\hat{C}_Q$ of 99.7 %, or in other words a 0.3 % willingness to pay. For the reactive damages function, damages obviously grow much quicker, with the 3σ trajectory reaching 20 % damages forty years earlier, at about 2080. The possibility of continued warming after 2100 (despite declining forcing) occurs along high sensitivity trajectories because of a continuing energy imbalance, and damages grow commensurately. These large damages are given strong weighting in the welfare function, and so even though the median estimate for climate damages never exceeds about 10 % (i.e., $C_R = 0.9$) the expected welfare-equivalent consumption is 60 %, or a willingness-to-pay of around 40 % of present consumption.

It is important to drive home the point that this large (though not unbounded) willingness-to-pay does not come from the fat tail of high climate sensitivity for the simple reason that the fat tail cannot physically be realized on these timescales. Used in a more colloquial sense, however, the fat tail of climate sensitivity—referring not just to extreme equilibrium values of $T_{2\times}$ but also to the accompanying temperature trajectories that lead to those extremes—does play a role in the relatively large willingness-to-pay of 40 % shown in Fig. 4d. Economic analyses are driven by the interplay between temperature changes, climate damage functions, and baseline

economic growth, and in this case the three factors combine to generate a large willingness-to-pay. In the literature to date most of the attention has focused on the "fat tail of climate sensitivity", but we would argue that an even greater amount of uncertainty surrounds projections for economic growth and climate damage functions. In our view it is the "fat tails" relating to these parameters, which are much more poorly constrained than the temperature trajectories, that are most likely to wag the policy dog.

4.1 The importance of correctly initializing the climate state for economic calculations

Economic analyses for policy purposes typically take the present day as the baseline for decision making. In other words, choices made now are based on how things might change into the future, and not by the "sunk cost" changes that have already occurred. The purpose of this short section is to highlight that, in projecting possible future climate changes, it is very important that economic models are correctly initialized with the current climate state and its uncertainties. The planet's climate change experiment is already more than a hundred years old, and two factors are of particular significance: (1) the deep ocean warming lags behind the surface ocean warming, and the resulting thermodynamic imbalance is driving heat into the deep ocean, and (2) as noted above, we are uncertain about climate sensitivity because we are uncertain about the current energy balance of the planet.

We now demonstrate the effect of incorrectly initializing economic models by repeating the calculations shown in Fig. 4, but neglecting to represent these two factors. The effect is to cause the range of possible future temperatures to spread out more quickly than should be the case. The reasons are as follows. Firstly, in the present-day climate state heat is currently being drawn into the deep ocean because it has yet to come into balance with the current forcing. To do so would, in fact, take centuries to millennia. If this imbalance is ignored in setting up the initial conditions of a model integration starting at the present day, it is equivalent to assuming that this ocean adjustment has already happened, meaning that heat which should be going into the deep ocean is instead available to heat the atmosphere. This leads to more rapid warming than should be the case. Secondly, and more importantly, is the fact that if climate sensitivity is high then climate forcing has been low. If this complementary pairing is neglected, it means that all trajectories (representing all possible climate sensitivities) are given the same starting point for present-day climate forcing. Hence, higher temperatures on high sensitivity trajectories are realized more quickly than is physically possible.

Figure 5 shows the effect of neglecting these two factors. For the calculation using C_Q, there is an increase of about 25 % in willingness to pay (i.e., 1-$\hat{C}_Q$), though it is still small, at around 0.5 %. There is an even bigger impact for $\hat{C}_R$. The 3σ trajectory reaches 20 % damages at 2050 (about thirty years earlier than it should), and willingess-to-pay increases from 40 to 80 %. These particular numbers are not the main point, but they illustrate that when the focus is on the high damages that might accompany the possibility of large temperature changes, it is important to correctly capture the physics of how such high temperatures might be attained.

Furthermore, in fully-coupled global climate models there is a complex spatial pattern to how quickly different regions warm, and to how quickly different regions
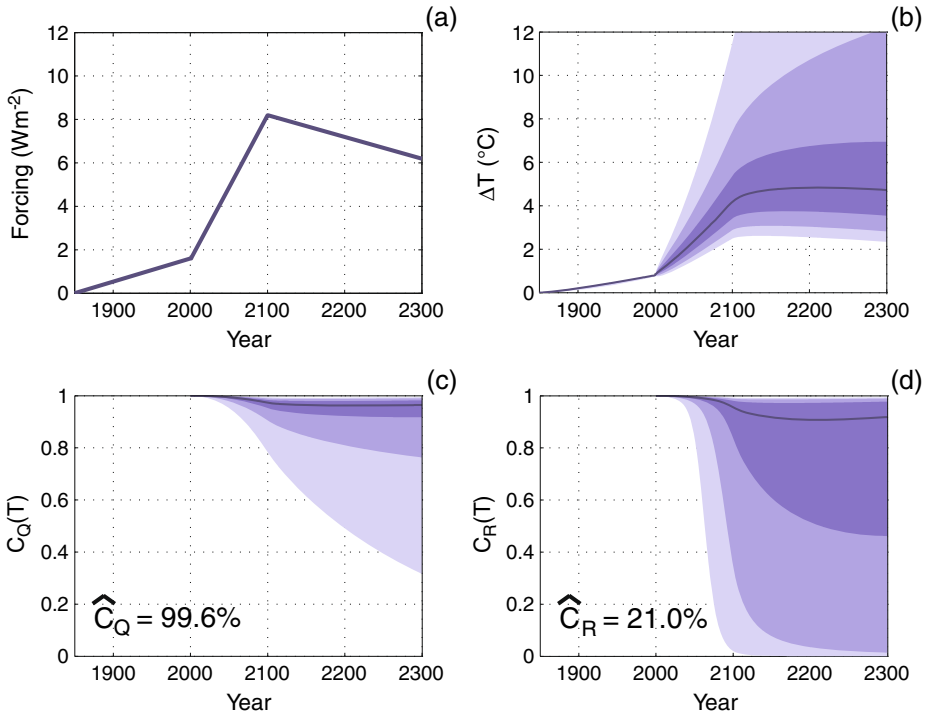
Fig. 5 The impact of an incorrectly initialized climate model. As for Fig. 4, but neglecting the relationship between forcing uncertainty and climate sensitivity, and also the thermodynamic imbalance between the surface ocean and the deep ocean. As a result, temperatures (panel **b**) spread out more quickly than in Fig. 4b. Large damages associated with C_R are realized more quickly, leading to a significantly reduced $\hat{C}_R$. The figure illustrates the importance of correctly initializing the climate state at the starting time of an economic analysis

come into equilibrium with the climate forcing. Since economic analyses often amalgamate regional assessment into a global damages function (e.g., Nordhaus 2008), IAMs ought to consider the impact of these differential regional warming rates. As of now, they do not do so.

5 Discussion

The primary point of this study is to emphasize that geophysical constraints exist that have an important impact on economic analyses of climate change policy. While the equilibrium climate sensitivity is characterized by a highly skewed uncertainty distribution (i.e., a fat tail), if it does transpire that we live on a planet with an extreme climate sensitivity, that fact will not become fully manifest for many centuries because of the enormous thermal inertia of the ocean.

Even in the absurd limit of an infinite climate sensitivity, the amount of warming that can occur is limited by the rate at which the system can accumulate energy (found by setting $f_a = 1$ and integrating Eq. 1). Therefore, from the standpoint of the

formal statistical definition—that the tail declines as a power in the limit of $T \to \infty$—there is no fat tail to the climate response at finite time. This is not an esoteric point: simple analytical damage functions can be highly sensitive to the integration limits. In idealized calculations of economic impacts, it is important to confront the physical realism of those limits.

Recent economic analyses have explored how uncertainty in the physical climate predictions affect policy guidance. Because such analyses invariably include discounting, the results can depend dramatically on how quickly this uncertainty grows into the future: the possibility of large, relatively near-term damages matters more than if those same damages occur in several centuries. We've isolated three physical factors that matter in this regard: the quasi-diffusive nature of ocean heat uptake, the correlated uncertainty between climate forcing and climate sensitivity, and the correct representation of the current climate state. Proper representation of each of these factors tends to reduce the rate at which uncertainty grows with time, and our example calculations with a simple economic model suggest that the impact on economic analyses is significant. In one example of an application of such analyses, none of the three IAMs used in a recent interagency report on the social cost of carbon (IAG 2010) represent any of these three physical effects.

Many economic models represent a more complicated interplay between growth rate, climate damages, and emissions than we have considered here, and the exact impacts of the effects we have presented will depend on the details of the formulation of the economic equations. Nonetheless, the inferences drawn using such models still depend on fundamental trade-offs between the rate at which future marginal consumption is discounted (because of economic growth), and the rate at which uncertainty in climate projections grows and becomes more skewed with time. Though the magnitude of the climate response depends on the magnitude of the climate forcing, Fig. 4 illustrates how physical uncertainty in the climate system affects the shape of the uncertainty in the climate response, no matter what the economic assumptions.

High temperatures can of course be reached if the climate system is forced strongly enough, and the illustrative C_R damage function would predict the distinct possibility of apocalyptic consequences at century's end for a do-nothing scenario (Fig. 4). Such highly skewed distributions can always be created by an economist making an appropriate choice of damage function, but we emphasize that the skewness is primarily a result of that economist's choice and not the physical uncertainty in the climate system.

The welfare impacts of climate change ultimately depend on the interplay between three factors: the temperatures resulting from a given carbon emissions profile, the damage functions resulting from those temperatures, and discounting issues that in our view are closely tied to assumptions about baseline economic growth. Fix any two of these and vary the third, and the results are likely to span the range of policy prescriptions from "don't worry about it" to "drastic action required immediately". Discussions of the fat tail of climate sensitivity have focused attention on the uncertainties surrounding the first of these three factors, but the uncertainties surrounding the other two factors are arguably even more profound. The most pressing questions are not about the likelihood of 10 or 20 °C temperature increases, but about the impacts of 3–6 °C temperature increases and about the ability of economic growth to cushion climate impacts in the decades and centuries ahead.

References

Allen MR et al (2006) Observational constraints on climate sensitivity. In: Schellnhuber HJ (ed) Avoiding dangerous climate change. Cambridge University Press, Cambridge, U.K.

Annan JD, Hargreaves JC (2006) Using multiple observationally-based constraints to estimate climate sensitivity. Geophys Res Lett 33:L06704. doi:10.1029/2005GL025259

Annan JD, Hargreaves JC (2009) On the generation and interpretation of probabilistic estimates of climate sensitivity. Clim Change. doi:10.1007/s10584-009-9715-y

Armour K, Roe GH (2011) Climate commitment in an uncertain world. Geohpys Res Lett 38. doi:10.1029/2010GL045850

Arrhenius S (1896) On the influence of carbon acid in the air upon the temperature of the ground. Philosoph Mag 41:237–276

Baker MB, Roe GH (2009) The shape of things to come: why is climate change so predictable? J Climate 22:4574–4589

Charney JG (1979) Carbon dioxide and climate: a scientific assessment. National Academy of Science, p 22

Colman R, McAvaney B (2009) Climate feedbacks under a very broad range of forcing. Geophys Res Lett 36. doi:10.1029/2008GL036268

Gregory JM (2000) Vertical heat transports in the ocean, and their effects on time-dependent climate change. Clim Dyn 16:501–515

Forest CE, Stone PH, Sokolov AP, Allen MR, Webster MD (2002) Quantifying uncertainties in climate system properties with the use of recent climate observations. Science 295:113–117

Hansen J, Russell G, Lacis A, Fung I, Rind D, Stone P (1985) Climate response times: dependence on climate sensitivity and ocean mixing. Science 229:857859

Held IM, Winton M, Takahashi K, Delworth T, Zeng F, Vallis GK (2010) Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. J Climate 23:2418–2427

Interagency Working Group on Social Cost Carbon (2010) Technical support document: social cost of carbon for regulatory impact analysis under executive order 12866. Available at www.epa.gov/otaq/climate/regulations/scc-tsd.pdf

IPCC (2007) Physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge Univ. Press

Knutti R, Stocker TF, Joos F, Plattner G-K (2002) Constraints on radiative forcing and future climate change from observations and climate model ensembles. Nature 416:719–723

Knutti R, Hegerl GC (2008) The equilibrium sensitivity of the Earth's temperature to radiation changes. Nat Geosci 1(11):735–743

Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl GA (2010) Challenges in combining projections from multiple climate models. J Climate 23:2739–2758

Millner A (2011) On welfare frameworks and catastrophic climate risks. Philosophy & Methodology of Economics. doi:10.2139/ssrn.1799481

Nordhaus WD (2008) A question of balance: economic modeling of global warming. Yale Press

Raper SCB et al (2001) Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. Clim Dyn 17:601613

Roe GH, Baker MB (2007) Why is climate sensitivity so unpredictable? Science 318:629–632

Roe GH (2009) Feedbacks, time scales, and seeing red. Annu Rev Earth Planet Sci 37:93–115

Roe GH (2010) Knowability and no ability in climate change projections. National center for environmental economics report no. 0564. Available at http://yosemite.epa.gov/ee/epa/eerm.nsf/vwAN/EE-0564-117.pdf

Roe GH, Armour KC (2011) How sensitive is climate sensitivity? Geophys Res Lett 38. doi:10.1029/2011GL047913

Solomon S et al (2010) Climate stabilization targets: emissions, concentrations and impacts over decades to millennia. National Research Council, National Academy of Sciences

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403–406

Weitzman ML (2009a) Additive damages, fat-tailed climate dynamics, and uncertain discounting. Economics 3:2009–39

Weitzman ML (2009b) On modeling and interpreting the economics of catastrophic climate change. Rev Econ Stat 91:1–19

Weitzman ML (2010) GHG targets as insurance against catastrophic climate damages discussion paper 2010-42. Harvard Project on International Climate Agreements, Cambridge, Mass.

Wigley TML, Schlesinger ME (1985) Analytical solution for the effect of increasing CO_2 on global mean temperature. Nature 315:649–652