When opposites attract: averting a climate catastrophe despite differing ethical views

Aurélie Méjean[1,2], Antonin Pottier[1,3], Stéphane Zuber[4,2] & Marc Fleurbaey[5]

[1] *CIRED, Centre International de Recherche sur l'Environnement et le Développement (CNRS, Agro ParisTech, Ponts ParisTech, EHESS, CIRAD)*

[2] *Centre National de la Recherche Scientifique (CNRS)*

[3] *Ecole des Hautes Etudes en Science Sociales (EHESS)*

[4] *Paris School of Economics (PSE)*

[5] *Woodrow Wilson School of Public and International Affairs, Princeton University*

Climate change is seen by economists as an issue of intertemporal consumption trade-off: consume all you want today and face climate damages in the future, or sacrifice consumption today to implement costly climate policies that will bring future benefits through avoided climate damages. If one assumes enduring technological progress, a controversial conclusion ensues: to reduce intergenerational inequalities, we should postpone climate policies and let future, richer generations pay. Growing evidence however suggests that the trade-off is more complex: abrupt, extreme, irreversible changes to the climate may cause discontinuities to socio-economic systems, possibly leading to human extinction. The most relevant trade-off would then be between present consumption and the mere existence of future generations. In this paper we show that when accounting even for a very small risk of catastrophic climate change, it is optimal to pursue stringent climate policies to postpone extinction. Our results conform with the well-known conclusion that tight carbon budgets are preferred when aver-

sion towards inequalities between generations is low. However, by contrast with previous studies, we show that stringent policies are also optimal when inequality aversion is high. This is because a higher inequality aversion makes the scenario of a small and relatively poor population (obtained when mitigation is low) especially unattractive. The size of the optimal carbon budget decreases with the social preference for large populations, but this parameter plays almost no role at extreme levels of inequality aversion. Our result thus demonstrates that views from opposite sides of the ethical spectrum in terms of inequality aversion converge in terms of climate policy recommendations, warranting immediate climate action. We therefore identify new spaces of compromise between contrasted ethical stances to set the ambition of climate policies, as new coalitions may emerge, bringing together opposing sides of international climate negotiations.

The risk of abrupt and irreversible changes to the climate is one of the five reasons for concern identified by the IPCC[1,2]. Extreme climate events, climate tipping points[3] - such as the shutoff of the Atlantic thermohaline circulation, the collapse of the West Antarctic ice sheet or the dieback of the Amazon rainforest - or climate-driven epidemic outbreaks may have indirect impacts, for instance through increased migration and conflicts[4,5], i.e. triggering a 'cycle of conflict and climate disaster', as stated by the UN security council[6]. Without appropriate policy responses, chain reactions could very well follow, possibly leading to general warfare, which would challenge the physical, political and social infrastructures of global society and possibly lead to its collapse. Taking catastrophes into account has been a daunting task of economics[7–9]. The discipline has traditionally mainly considered climate change as an issue of intertemporal consumption trade-off[10,11]. In reality, abrupt climate change may introduce an irreversible regime shift in the sense that post-catastrophe welfare is independent from pre-catastrophe actions[12,13], and could be zero. The possibility that social welfare may drop to zero can be interpreted as human extinction. The trade-off would then be not only between present and future consumption, but between present consumption and catastrophic risk reduction[14].

We evaluate the social welfare associated with 250 climate policies, corresponding to 250 carbon budgets. Each policy is first translated into a social outcome x using a climate-economy model linking the risk of extinction (or hazard rate) to temperature change. Social outcomes, i.e. the streams of consumption and extinction risk over time, are then translated into social welfare using a social welfare function[15]. This function embodies ethical views in terms of inequality aversion and social preference for large populations. The chosen function treats all generations in a symmetric way, giving no *a priori* preference to the present. Policies are then ranked according to social welfare. We examine the impact of inequality aversion (η), the social preference for large populations (β) and the risk of extinction (through the marginal hazard rate b) on the optimal policy. Policies are specified in terms of saving rate (fixed at a constant for all scenarios[16], set at 25.8%, which is consistent with the observed world average gross saving rate[17,18]) and carbon budget (spanning 400-5800 $GtCO_2$ from 2015 to 2250 across scenarios). The 400 $GtCO_2$ budget is consistent with the objective of limiting warming to 1.5°C, while the 5800 $GtCO_2$ budget corresponds to warming above 6°C.

We show that with an exogenous risk of extinction (marginal hazard rate $b = 0$), i.e. when the risk of extinction is independent from temperature and from the carbon budget, large carbon budgets are optimal if the social planner is averse to inequalities among generations (high η). This is a standard result in the economics literature: as future generations are assumed to be richer thanks to technological progress, a high aversion towards inequalities across generations means postponing climate policy (i.e. choosing larger carbon budgets) to spare the present, poorer generation[19]. This is illustrated in Fig. 1, where carbon budgets leading to warming of at least 3°C (red and purple) are preferred for an inequality aversion $\eta \geq 1.7$. This result changes dramatically if the risk of extinction depends on temperature. In that case, even with a very small endogenous risk of extinction (e.g. $b = 10^{-7}$ per °C, Fig. 2), it is optimal to achieve tigher carbon budgets than in the case where the risk of extinction does not depend on temperature. This result holds for all combinations of inequality aversion (η) and preference for large populations (β), although to various extents depending on the combination. For instance in the case ($\eta = 2.5, \beta = 0.5$), with $b = 10^{-7}$ per °C, the optimal carbon budget is 1800 $GtCO_2$ lower than in the case of an exogenous risk of extinction (from 4600 $GtCO_2$ in the exogenous case to 2800 $GtCO_2$ in the endogenous case).

With a strong preference for large populations (large β), it is optimal to implement tight carbon budgets. This is shown on Fig. 2, where for a given inequality aversion (say $\eta = 3$), increasing β from 0 to 1 (i.e. switching from the average utilitarian to the total utilitarian case) reduces the optimal carbon budget from 5000 $GtCO_2$ - corresponding to a stabilization above 3°C - to less than 1600 $GtCO_2$ - bringing the temperature increase close to the 2°C objective. This result is intuitive, as a tighter carbon budget delays extinction due to climate change, which leads to a larger number of individuals across generations. The impact of inequality aversion

(η) on the optimal carbon budget depends on the social preference for large population. With a relatively weak preference for large populations ($\beta = 0$), we find the standard result that a high aversion to inequalities across generations leads to postpone climate policy to spare the present generation. This directly translates into a larger carbon budget. In that case, only low inequality aversion (close to 1) commands carbon budgets small enough to be compatible with the objective of limiting global warming to 2°C compared to the pre-industrial level. With a stronger preference for large populations ($\beta > 0$), both low inequality aversion (η close to 1) and high inequality aversion ($\eta > 3$) command tight carbon budgets which may limit global warming to 2°C (blue, Fig. 2). This is because with a low inequality aversion, the benefit of preserving a large and on average richer population makes mitigation attractive, whereas a high inequality aversion makes the scenario of a short-lived and relatively poor population particularly unattractive.

This pattern remains as the risk of extinction due to climate change increases from 10^{-7} to 10^{-4} per °C. However, in those cases, tight carbon budgets are optimal for a wider range of ethical views: the results show a larger number of combinations of inequality aversion (η) and preference for large populations (β) for which tight carbon budgets are warranted (Fig. 3). For a higher risk of extinction (b between 10^{-2} and 10^{-3} per °C), the most stringent carbon budget (400 $GtCO_2$) is optimal for all social preferences on population size and inequality aversion across generations. Such a carbon budget is compatible with limiting global warming to 1.5°C above the pre-industrial level. For an even higher risk of extinction ($b = 0.1$ per °C), a doom effect[20] occurs: it is then optimal to loosen the carbon budget in order to maximize present consumption, as future generations are very unlikely to exist.

Our message is twofold. First, the possibility that climate change may drive human extinction, as improbable as it may be, changes the terms of the trade-off to solve the climate puzzle. An additional risk of extinction due to climate change as small as 10^{-7} per annum per $^\circ$C above 1°C commands much tighter carbon budgets than in the case where that risk is not accounted for. An additional risk of extinction due to climate change of 10^{-3} per annum per degree above 1°C commands strong, immediate action to limit global temperature increase to 1.5°C for the whole range of ethical values. Second, accounting for the risk of extinction reveals new spaces of compromise between opposing ethical stances to set the ambition of climate policies. Coalitions could therefore arise between opposite sides of the ethical spectrum in terms of inequality aversion. This result may have far-reaching consequences for international climate negotiations.

Methods

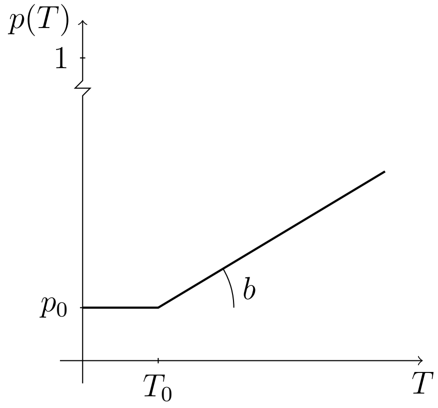
We evaluate 250 climate policies following two steps. Policies are specified in terms of saving rate (fixed at a constant rate for all scenarios) and carbon budget (spanning 400-5800 $GtCO_2$ across scenarios). First, each policy is translated into a social outcome x by a climate-economy model linking the risk of extinction (or hazard rate) to temperature change. Social outcomes are defined in terms of consumption per capita c and hazard rate p over time[a]. Due to this risk, the time horizon is unknown ex-ante. Second, social outcomes are evaluated using a social welfare function and

---

[a] Population (conditional on existence) is exogenous, the generation size stream n does not depend on the policy (it is thus common to all social outcomes). Time horizon T depends on the social outcome. Generation t is of size n_t enjoys consumption per capita c_t. Consumption is assumed to be equally distributed among individuals of the same generation.

policies are ordered according to social welfare.

The climate economy model RESPONSE[21–23] belongs to the tradition of compact integrated assessment models[11,24,25], combining a simple representation of the economy and a climate module. The economic module is a Solow-like growth model with capital accumulation and exogenous population. It includes climate mitigation costs and a climate damage function that is chosen to be equivalent to that of DICE. The climate module describes the evolution of global temperature increase and radiative forcing as a function of emissions. The temperature increase feedbacks on the economy through the damage function. We model extinction as an abrupt event that occurs over the course of a decade. This is a harsh simplification, as the processes involved would be gradual, with feedbacks between institutions, consumption, mortality and fertility, and would probably take decades to unfold. We consider that after the catastrophe occurs, the course of events is inevitably set and cannot be acted upon[12,13]. Whether such a catastrophic chain of events would actually lead to the extinction of all human beings is controversial and may be considered as highly unlikely. Given the large uncertainty surrounding the probability of such events, we attempt at covering a wide range of possible values (as described below). We assume that the hazard rate p depends on the temperature increase T only, and not on wealth or adaptative capacity. We therefore disregard any factor, social or natural, that may mitigate the effect of climate change on the risk of extinction. More precisely, we model the hazard rate p as a linear function of temperature increase T, above a temperature threshold T_0. The hazard rate at date t, p_t, is truly $p(T_t)$, a function of temperature increase T_t at date t. We assume that this function is valid in a range of temperature increase spanning from 1°C to 10°C, which includes the values of temperature increase presented in the analysis.



$p(T)$	hazard rate as a function of temperature increase T
p_0	minimum hazard rate, set at 10^{-3} per annum
T	temperature increase compared to pre-industrial levels (°C)
T_0	temperature increase above which the hazard rate starts rising with temperature, set at 1 °C
b	marginal hazard rate per °C above T_0

The survival probability at t is $P_{\geq t} = \prod_{\tau=0}^{t-1}(1 - p_\tau)$ (probability that the time horizon is greater than or equal to t). The probability that humanity becomes extinct after generation t is $P_t = p_t P_{\geq t}$ (horizon probability): it is the probability that the time horizon is t. Given the very nature of the risk considered, there is no available data about its realization from which the parameters of the hazard rate function could be inferred. The calibration can thus only be illustrative and tentative. The calibration does not rely on the frequentist approach to probability but on the Bayesian or subjectivist approach, where the probability reflects informed though subjective beliefs about the links between events. We build on The Stern Review[26] which treats generations in a symmetric way, yet uses a non-zero time discount rate, set at 10^{-3} per annum to account for the possible extinction of humanity. We follow the argument and set p_0 at 10^{-3} per annum. Although the calibration of the marginal hazard rate (b) is impossible due to the inherent lack of data, we attempt at providing a range by considering all values that cannot be reasonably excluded, building on the evidence provided in the latest IPCC assessment report[27]. The maximum value of b can be derived from the upper bound of the temperature increase above T_0 at which there is no certain extinction due to climate change. We cannot affirm with certainty that life on Earth would be impossible for the human species if the temperature increase T reached $10°\mathrm{C}$ above T_0: evidence suggests that with a temperature increase of $11°\mathrm{C}$ to $12°\mathrm{C}$, metabolic heat dissipation would become impossible in most of today's inhabited areas[28]. Therefore, we know that b should be lower than 10^{-1} per °C. We thus set the maximum value for b, b_{max}, at 10^{-1} per °C. In order to define a minimum value for b, we argue that the additional hazard rate above p_0 due to an increase of temperature of $1°\mathrm{C}$ above T_0 would be at the minimum a few orders of magnitude below p_0. We retain $b_{min} = p_0 \cdot 10^{-4} = 10^{-7}$ per °C. With this minimum value, the survival probability after a hundred years at a sustained temperature increase of $10°\mathrm{C}$ becomes 0.90471, instead of 0.90479 for $b = 0$. This shows that b_{min} would have little impact on the hazard rate and is therefore appropriately small.

We follow most of the existing literature and write aggregate welfare W from outcome x as a generalized utilitarian form, which can be written as an average utility and a population weight N_T^{β}.

Definition 1 (Expected number-dampened utilitarian social welfare functions) *A social welfare function is an* expected number-dampened utilitarian social welfare function *(ENDU SWF) if there* *exist real numbers* $\beta \in [0,1]$, $\underline{c} \in \mathbb{R}_{++}$ *and* $\eta \in \mathbb{R}_{+}$ *such that:*

$$\begin{aligned} W(x) &= \sum_{T=0}^{\infty} P_T \left(N_T^{\beta} \sum_{t=0}^{T} \frac{n_t}{N_T} \left[\frac{c_t^{1-\eta}}{1-\eta} - \frac{\underline{c}^{1-\eta}}{1-\eta} \right] \right) && (1) \\ &= \sum_{T=0}^{\infty} P_T N_T^{\beta} A U_T && (2) \\ &= \sum_{t=0}^{\infty} \underbrace{\left(\sum_{T=t}^{\infty} P_T N_T^{\beta-1} \right)}_{\theta_t} n_t \left[\frac{c_t^{1-\eta}}{1-\eta} - \frac{\underline{c}^{1-\eta}}{1-\eta} \right]. && (3) \end{aligned}$$

where $N_T = \sum_{t=0}^{T} n_t$ is the total (cumulated) population that comes into existence in social outcome x. $\frac{c_t^{1-\eta}}{1-\eta} - \frac{\underline{c}^{1-\eta}}{1-\eta}$ is the utility enjoyed by an individual of generation t. There is no pure time discounting, i.e. contrary to the standard approach, we treat generations in a symmetric way[b]. θ_t is tantamount to a time discount factor on the utility of generation t. This discount factor stems from the uncertainty about the horizon. It depends on the hazard rate and on the attitudes towards population size as embodied in the population ethics coefficient β. The social welfare function embodies the preferences of a social planner through two ethical parameters.

- Parameter η is the intergenerational inequality aversion coefficient. It determines the marginal utility of individual consumption. A more inequality averse social welfare function means that the social planner is willing to sacrifice more to equalize consumption levels across generations. We test a wide range for this parameter, i.e. from 1 to 4.5, similar to the range of inequality aversion values reviewed in the latest IPCC assessment report[15].
- Parameter β is the social preference for large populations. The social welfare function embeds well-known views of utilitarianism when population size varies. *Total* or *classical*

---

[b] Ramsey and Stern provide arguments on why we should treat generations in a fair way[26,29].

utilitarianism values the total sum of utilities ($\beta = 1$). *Average utilitarianism* values average utility ($\beta = 0$). We vary β between 0 and 1, spanning cases between total and average utilitarian views.

The social welfare function also includes a consumption per capita threshold, $\underline{c}$. In the case of total utilitarianism ($\beta = 1$), it is the *critical level of consumption*, i.e. the level of consumption such that, if enjoyed by an additional individual, total welfare is left unchanged when that individual is added to the population. When $\beta \neq 1$ however, the critical level is not constant and depends on population size and average utility. Parameter $\underline{c}$ still influences aggregate welfare, the value of changing population size, and thus of the risk of extinction. In this framework, extinction is therefore theoretically equivalent to a situation where consumption per capita reaches a level at which life is barely worth living, i.e. a situation where consumption per capita falls below the critical level of consumption. In the case of total utilitarianism, the catastrophe can therefore be understood as a situation where humanity lives below the $\underline{c}$ threshold. The interpretation is less straightforward in the case of average utilitarianism, as in that case the critical level of consumption is not equal to $\underline{c}$.

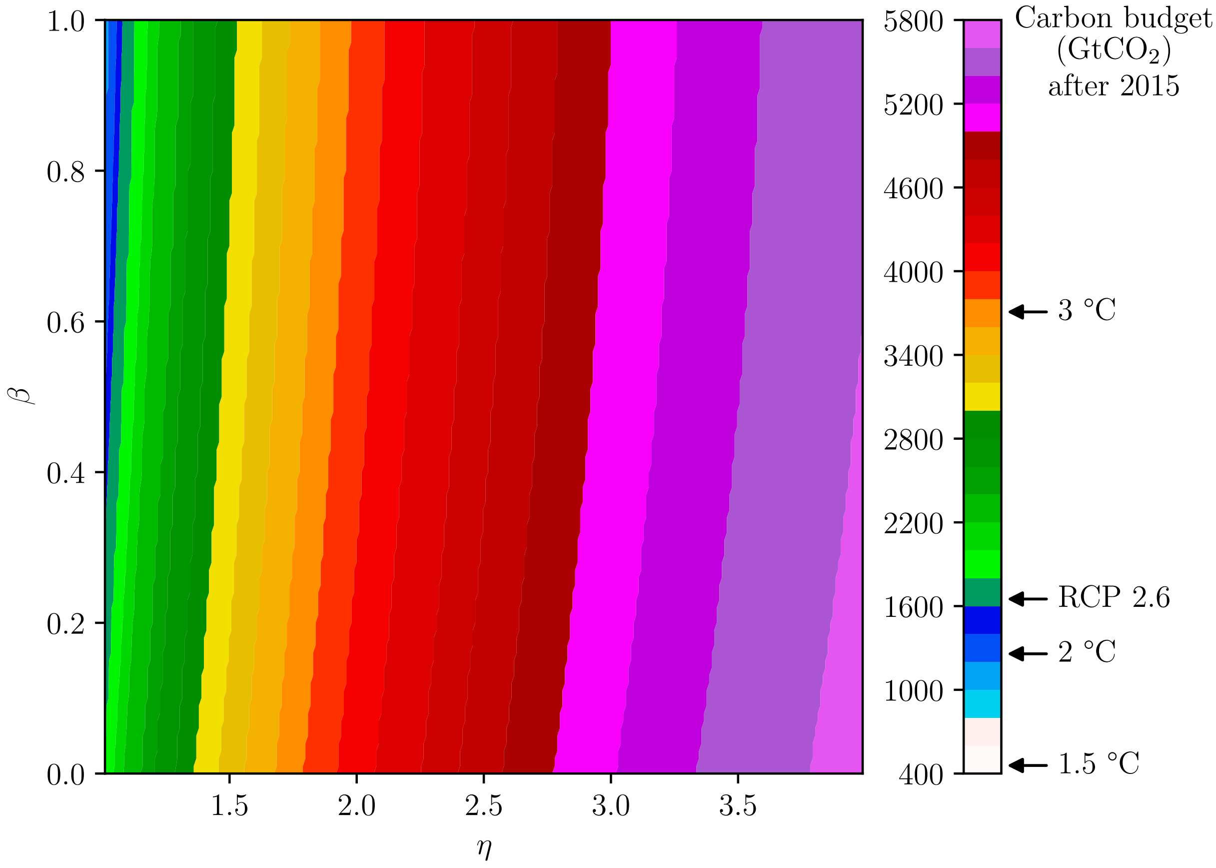


Figure 1: Optimal climate budget as a function of the social preference for large populations (β) and the inequality aversion coefficient (η) for an exogenous hazard rate ($b = 0$)

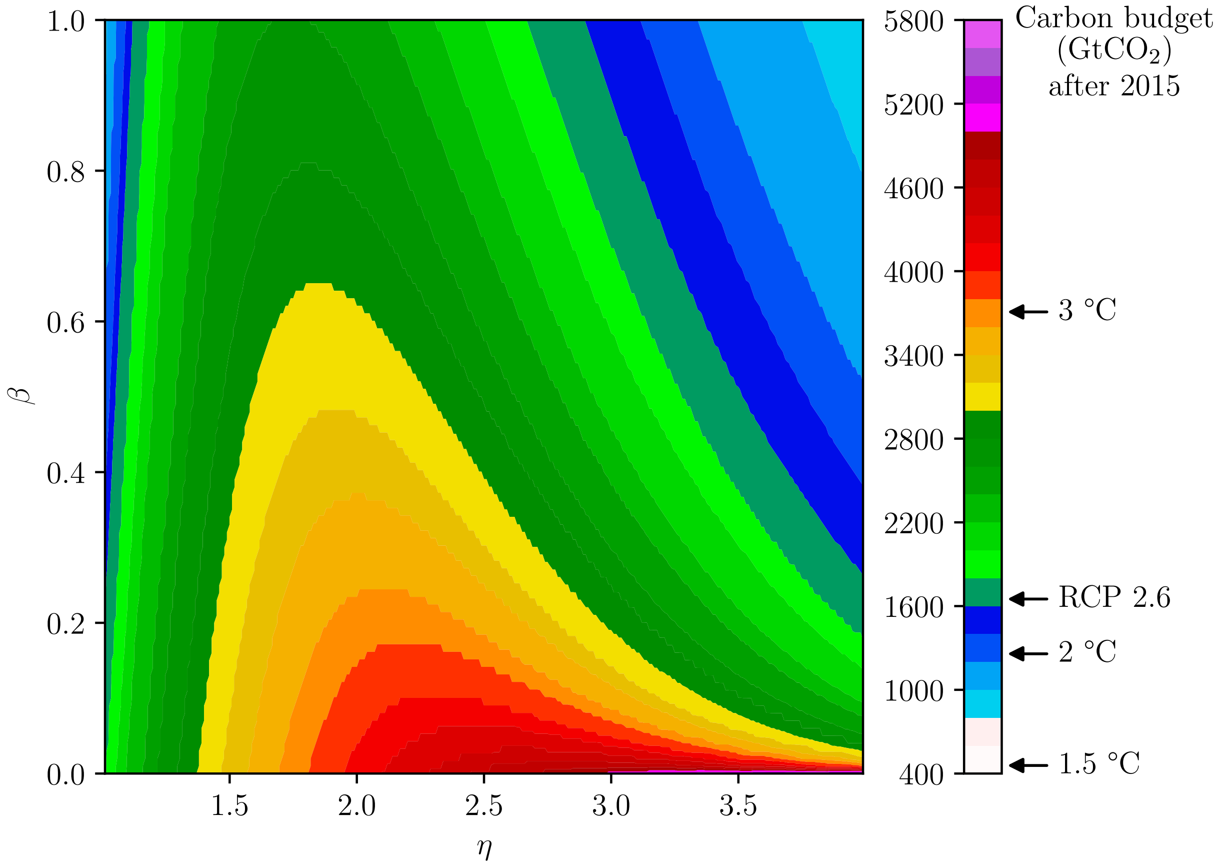
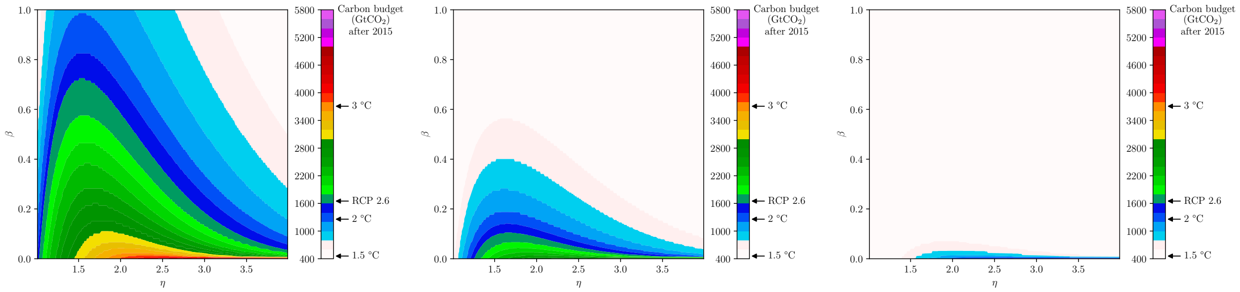
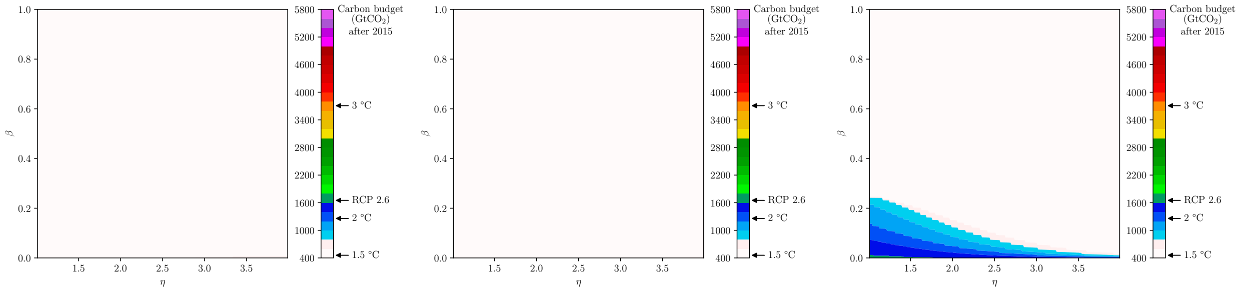


Figure 2: Optimal climate budget as a function of the social preference for large populations (β) and the inequality aversion coefficient (η) for a hazard rate that depends on temperature ($b = 10^{-7}$ per °C)



(a) $b = 10^{-6}$ per annum per °C **(b)** $b = 10^{-5}$ per annum per °C **(c)** $b = 10^{-4}$ per annum per °C



(d) $b = 10^{-3}$ per annum per °C **(e)** $b = 10^{-2}$ per annum per °C **(f)** $b = 10^{-1}$ per annum per °C

Figure 3: Optimal climate budget as a function of the social preference for large populations (β) and the inequality aversion coefficient (η) for a range of values of the marginal hazard rate (b)

1. McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. & White, K. S. (eds.) *Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of the Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2001).
2. Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J. & Mastrandrea, M. D. (eds.) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of the Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2014).
3. Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences* **105**, 1786–1793 (2008).
4. Reuveny, R. Climate change-induced migration and violent conflict. *Political Geography* **26**, 656–673 (2007).
5. Hsiang, S. M., Burke, M. & Miguel, E. Quantifying the Influence of Climate on Human Conflict. *Science* **341**, 1235367–1235367 (2013).
6. UN. Fragile countries risk being stuck in a cycle of conflict and climate disaster, Security Council told (2018). URL `https://news.un.org/en/story/2018/07/1014411`.
7. Dowlatabadi, H. Climate Change Thresholds and Guardrails for Emissions. *Climatic Change* **41**, 297–301 (1999).
8. Tol, R. S. J. Is the Uncertainty about Climate Change too Large for Expected Cost-Benefit Analysis? *Climatic Change* **56**, 265–289 (2003).

9. Azar, C. & Lindgren, K. Catastrophic Events and Stochastic Cost-benefit Analysis of Climate Change. *Climatic Change* **56**, 245–255 (2003).
10. Cline, W. R. *The economics of Global Warming* (Institute for International Economics, Washington, 1992).
11. Nordhaus, W. D. *Managing the global commons: the economics of climate change* (MIT Press, Cambridge, Mass., 1994).
12. Cropper, M. Regulating activities with catastrophic environmental effects. *Journal of Environmental Economics and Management* **3**, 1–15 (1976).
13. Clarke, H. R. & Reed, W. J. Consumption/pollution tradeoffs in an environment vulnerable to pollution-related catastrophic collapse. *Journal of Economic Dynamics and Control* **18**, 991–1010 (1994).
14. Weitzman, M. L. On Modeling and Interpreting the Economics of Catastrophic Climate Change. *Review of Economics and Statistics* **91**, 1–19 (2009).
15. Kolstad, C. *et al.* Social, Economic, and Ethical Concepts and Methods. In Edenhofer, O. *et al.* (eds.) *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014).
16. Golosov, M., Hassler, J., Krusell, P. & Tsyvinski, A. Optimal Taxes on Fossil Fuel in General Equilibrium. *Econometrica* **82**, 41–88 (2014).

17. The World Bank. Gross savings (% of GDP). World Bank national accounts data, and OECD National Accounts data files., The World Bank (2017).

18. Dennig, F., Budolfson, M. B., Fleurbaey, M., Siebert, A. & Socolow, R. H. Inequality, climate impacts on the future poor, and carbon prices. *Proceedings of the National Academy of Sciences* **112**, 15827–15832 (2015).

19. Dasgupta, P. Discounting climate change. *Journal of risk and uncertainty* **37**, 141–169 (2008).

20. Perrissin Fabert, B., Espagne, E., Pottier, A. & Dumas, P. The Doomsday Effect in Climate Policies. Why is the Present Decade so Crucial to Tackling the Climate Challenge? Working Paper 2012.62, Fondazione Eni Enrico Mattei (2012).

21. Ambrosi, P. *et al.* Optimal Control Models and Elicitation of Attitudes towards Climate Damages. *Environmental Modeling & Assessment* **8**, 133–147 (2003).

22. Dumas, P., Espagne, E., Perrissin-Fabert, B. & Pottier, A. Comprehensive Description of RESPONSE. CIRED Working Paper 41-2012, CIRED (2012).

23. Pottier, A., Espagne, E., Fabert, B. P. & Dumas, P. The Comparative Impact of Integrated Assessment Models Structures on Optimal Mitigation Policies. *Environmental Modeling & Assessment* **20**, 453–473 (2015).

24. Hope, C., Anderson, J. & Wenman, P. Policy analysis of the greenhouse effect - An application of the PAGE model. *Energy Policy* **21**, 327–338 (1993).

25. Tol, R. S. J. On the optimal control of carbon dioxide emissions: an application of FUND. *Environmental Modeling & Assessment* **2**, 151–163 (1997).

26. Stern, N. H. *The economics of climate change: the Stern review* (Cambridge University Press, Cambridge, UK ; New York, 2007).

27. Human Health: Impacts, Adaptation, and Co-Benefits. In Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J. & Mastrandrea, M. D. (eds.) *Climate Change 2014 Impacts, Adaptation, and Vulnerability*, 709–754 (Cambridge University Press, Cambridge, 2014).

28. Sherwood, S. C. & Huber, M. An adaptability limit to climate change due to heat stress. *Proceedings of the National Academy of Sciences* **107**, 9552–9555 (2010).

29. Ramsey, F. A mathematical theory of saving. *The Economic Journal* **38**, 543–559 (1928).

Acknowledgements This research has been supported by the Chair on Welfare Economics and Social Justice at the Institute for Global Studies (FMSH - Paris), the Franco-Swedish Program on Economics and Philosophy (FMSH - Risksbankens Jubileumsfond), the ANR project EquiRisk, Grant # ANR-12-INEG-0006-01, the ANR project FairClimpop, Grant # ANR-16-CE03-0001-01 and Princeton University's Program in Science, Technology and Environmental Policy (STEP).

Conflicts of interest The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to Aurélie Méjean. (email: mejean@centre-cired.fr).