



Delft University of Technology

Document Version

Final published version

Licence

CC BY

Citation (APA)

Biswas, P., Zatarain Salazar, J., Kwakkel, J. H., & Marangoni, G. (2026). Prioritizing the welfare of vulnerable nations promotes equitable achievement of Paris target. *npj Climate Action*, 5, Article 57. <https://doi.org/10.1038/s44168-026-00385-z>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

<https://doi.org/10.1038/s44168-026-00385-z>

Prioritizing the welfare of vulnerable nations promotes equitable achievement of Paris target

Palok Biswas¹✉, Jazmin Zatarain Salazar¹, Jan H. Kwakkel¹ & Giacomo Marangoni^{1,2}

Equitable mitigation to limit warming to 2 °C has been central to international climate negotiations and global action. Approximately half of the mitigation pathways in the Intergovernmental Panel on Climate Change report use Cost-Benefit Integrated Assessment Models (CB-IAMs). These pathways are sensitive to normative assumptions ingrained in CB-IAMs and have been criticized for unfairly burdening developing nations. Here, we show that making normative assumptions explicit, separating conflicting policy objectives, and employing adaptive policies can substantially alter the prescribed global mitigation pathways and the underlying distribution of mitigation burdens. Our findings demonstrate that justice framing is pivotal. A Utilitarian framing (maximizing global welfare) results in less ambitious mitigation, exceeding 2 °C warming and unjustly distributing responsibilities. In contrast, a Prioritarian framing (prioritizing the worst-off) prescribes pathways that meet the Paris Agreement goals and ensure equitable distribution of mitigation responsibility. Our approach integrates diverse ethical perspectives, facilitating informed deliberation and enabling stakeholders to reach a negotiated compromise.

What awaits us beyond the 2 °C Paris Target? Exceeding 2 °C magnifies the risk of interconnected climate hazards that cascade and compound, intensifying overall damage. While this threat is undeniable, deep uncertainties in both economic and climate systems—such as the potential collapse of the Atlantic Meridional Overturning Circulation or the disintegration of the Antarctic Ice Sheet¹—hinder our ability to accurately predict the resulting socioeconomic repercussions^{2–5}. These irreversible hazards could compromise ecosystem stability and human well-being, underscoring the urgency to keep warming well below 2 °C. The economic consequences will be asymmetric, with vulnerable and developing countries facing the greatest impacts on their well-being compared to developed nations^{5,6}. This disparity highlights the critical importance of equity in global climate action, a foundational principle of the Paris Agreement⁷. Although the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) acknowledges that nations contributing the least to global warming face disproportionate adverse impacts⁵, critics argue that the mitigation pathways in the Working Group III (WGIII) report do not reflect this disparity and thus perpetuate inter-regional inequalities. In particular, these pathways constrain the growth of developing countries while projecting higher per capita GDP and consumption growth for developed nations, resulting in an unfair allocation of emission reductions and global carbon budgets^{8–10}.

Integrated Assessment Models (IAMs) underpin the IPCC's WGIII analyses by generating long-term mitigation pathways and informing global climate policy^{11,12}. In particular, Cost-Benefit IAMs (CB-IAMs) identify near-term emission reduction strategies and evaluate macroeconomic outcomes by balancing immediate costs against long-term benefits. Nearly half of the modelled global mitigation pathways in the IPCC AR6 use Cost-Benefit Analysis (CBA)^{5,13}. This is problematic as CB-IAM policy prescriptions are highly sensitive to modelling assumptions¹⁴ and have been criticized for overlooking global equity and undermining near-term rapid mitigation benefits^{15,16}. Since CB-IAMs are rooted in CBA, they inherit its well-documented shortcomings, raising serious ethical dilemmas in climate policy recommendations^{17,18}. Specifically, these issues include (1) reducing incommensurate objectives to a single monetary metric, (2) making implicit normative assumptions, and (3) inadequately treating deep uncertainties in climate projections, where the probabilities, relationships, or outcomes are unknown.

First, CBA simplifies climate policymaking by converting diverse, incommensurate objectives into a single monetary metric. This conversion requires monetizing unquantifiable values like planetary health and human lives before aggregation¹⁷, which tends to favour specific stakeholders unpredictably and devalues non-economic factors^{17,19}. The challenge intensifies when quantifying future climate damages and abatement costs

¹Delft University of Technology, Technology, Policy and Management, Delft, The Netherlands. ²European Institute on Economics and the Environment, Milan, Italy.

✉ e-mail: p.biswas.nl@gmail.com

under deep uncertainty, reflected in the high variability of damage³ and abatement cost estimates^{20,21} driven by uncertain factors such as future climate impacts, technological progress, and societal and political dynamics. A further contentious assumption is that irreversible damages and loss of non-renewable resources can be compensated by economic gains, implying that welfare losses from climate damages can be offset by an equal increase in welfare from consumption gains²². This equivalence presents two major issues. One is the limited substitutability between different objectives. For example, damages to food supply cannot be adequately compensated by growth in other economic objectives²³. Another is evidence from prospect theory, which shows that losses weigh more heavily than equivalent gains^{24–26}. This asymmetry is observed not just individually but also at the macro scale²⁷. Although aggregating incommensurate objectives is not impossible, it requires making a priori normative judgments, specifically assigning weights to objectives reflecting their relative importance. Because these objectives are incommensurate, different stakeholders (given their asymmetric responsibilities and impacts) weigh them differently. Such an implicit weighting in CBA obscures trade-offs and often breeds disagreement among stakeholders. Multi-objective optimization keeps incommensurate objectives separate and seeks to optimize policies across competing goals, revealing trade-offs and supporting transparent decision-making²⁸. By making decision-makers' preferences over objectives explicit, multi-objective approaches facilitate a posteriori navigation of conflicting interests and have recently gained attention in climate policymaking literature as a means to enable dialogue among diverse actors^{29–31}.

Second, fair climate policymaking requires accommodating the value pluralism of diverse actors to ensure the process is inclusive and transparent. This is particularly important for CB-IAMs, as these models embed implicit normative assumptions that obscure the value judgments underlying their policy recommendations and can give rise to disagreements between decision-makers. For instance, pure CBA ignores diminishing marginal utility—the idea that each additional unit consumed provides progressively less satisfaction as overall consumption grows. This omission undervalues impacts on less well-off regions and leads to unjust policy recommendations^{14,16}. In contrast, the Social Welfare Function (SWF) framework, designed to ethically evaluate policy options, incorporates diminishing marginal utility to overcome CBA's limitations. Consequently, CB-IAM literature increasingly employs the Utilitarian SWF, which aggregates the well-being of regions while discounting each region's future well-being. The Utilitarian SWF makes value judgments over the distribution of consumption more explicit¹⁸ and offers a more equitable alternative to traditional CBA methods³². However, a Utilitarian SWF overlooks the distribution of well-being³³ and, due to discounting, devalues future damage costs relative to immediate abatement costs, thus inflating the cost-benefit ratio¹⁵. These normative judgments become even more contentious under uncertainty, as debates over the appropriate discount rates illustrate^{34,35}.

Since mitigation benefits and burdens vary starkly between developed and developing nations, a single definition of justice is unsuitable because policies optimized for a single conception of justice are ineffective and result in deadlock³⁶. Without accounting for multiple values and preferences in mitigation policy design, scientific insights risk becoming politicized, potentially reinforcing dominant parties' interests in justifying climate inaction while shifting risks to vulnerable populations³⁷. Additionally, mitigation strategies that exclude diverse normative perspectives entrench structural injustices beyond those already caused by climate change, particularly through how policies are framed³⁸. Policies perceived as unfair face public resistance, erode trust, and undermine legitimacy³⁷, which ultimately becomes a barrier to effective climate action. A systematic examination of the consequences of mitigation policies optimized under different principles of justice is critical in informing policymakers about the consequences of their preferences and values, as the CB-IAM policy recommendations are strongly sensitive to these normative preferences¹⁴. Therefore, several studies advocate making normative concerns more explicit and incorporating diverse ethical viewpoints into policy design^{15,39,40}. A just and effective alternative is to adopt policy design frameworks that accommodate multiple

values and preferences of different stakeholders, fostering mutual understanding and consensus.

Third, CBAs are not equipped to support decision-making under deep uncertainty. Climate sensitivity and damage functions are major sources of deep uncertainty⁴¹, making the design of effective and just mitigation policies challenging. Most CB-IAMs rely predominantly on average outcomes, overlooking low-probability, high-impact events and deep uncertainties in climate sensitivity and damages^{15,42,43}. This simplification places CB-IAMs at odds with climate science, where expert opinion on “optimal” global warming differs dramatically between climate scientists and economists⁴⁴. Traditionally, CB-IAMs optimize mitigation policies by assuming deterministic climate sensitivity and damage functions. In addition, these mitigation strategies are decided a priori for every year (or model timestep) and remain fixed, regardless of whether the observed climate sensitivity differs from the initial assumptions^{30,45}. This approach is evident in the popular CB-IAM DICE: mitigation strategies for an entire century are optimized under a single realization of climate sensitivity⁴⁶, leaving no room for learning as climate uncertainty evolves. Such non-adaptive strategies remain common in mitigation assessments⁴⁷. Optimizing policy under deterministic assumptions about inherently uncertain parameters prevents adaptive adjustment as new information emerges. This inflexibility leads to sub-optimal outcomes and high regrets if future conditions deviate from the anticipated scenario^{30,48}, yielding overconfident and potentially misleading policy recommendations⁴⁹.

Some CB-IAM literature addresses policy optimization under uncertainty through Stochastic Dynamic Programming (SDP), which formulates the optimization problem recursively⁵⁰. The SDP approach enables decision-makers to update mitigation policies at each timestep as new information emerges, allowing them to learn about climate sensitivity from observed temperature. Although SDP makes uncertainty conceptually tractable, its application to complex IAMs with detailed climate modules remains limited, as it substantially increases computational complexity, especially when optimizing across multiple objectives. However, a more scalable solution is to optimize adaptive policy rules that link climate observations to policy actions, which function as a closed-loop control approach where mitigation decisions adjust based on observed warming at each timestep^{30,51}. These adaptive policies typically take functional forms, such as neural networks, whose parameters are optimized across multiple objectives using Multiobjective Evolutionary Algorithms (MOEAs), which can handle high-dimensional decision spaces, multiple objectives, and uncertainties⁵². To implement adaptive policies, IAMs function as simulators where decisions at each timestep affect future climate states, creating path-dependencies between current actions and long-term outcomes. The adaptive mitigation policy adjusts actions based on observed warming, enabling timely adjustments in response to climate variability. This feedback mechanism makes adaptive policies robust against deep uncertainties and allows them to accommodate diverse future states-of-the-world (SOW) and extreme events. Adaptive policies better manage trade-offs between conflicting objectives such as economic growth and climate mitigation. They perform well under uncertainty while avoiding the shifting of undue risks across vulnerable regions or to future generations, making them more effective and just³¹.

The aforementioned shortcomings of CB-IAMs have been separately addressed in the literature. Some studies employ multiobjective optimization to examine trade-offs between incommensurate objectives^{29–31} and assess the robustness of adaptive policies under climate uncertainty^{30,53,54}. However, these studies do not examine how diverse stakeholders' value preferences affect mitigation pathways and regional burden distribution. A separate strand of literature compares the justice implications of CBA versus Utilitarian Social Welfare Functions³², while another examines global cooperation versus non-cooperation scenarios using hybrid Utilitarian and Prioritarian Social Welfare Functions⁵⁵. Under cooperative scenarios, the hybrid approach combines Utilitarian time-discounting with Prioritarian inequality aversion, while non-cooperative scenarios optimize individual regional welfare with time-discounting but without inequality aversion⁵⁵.

However, both these studies optimize for global welfare without accounting for stakeholder heterogeneity in preferences over objectives and values. The global social planner optimizes under a single objective from a single ethical perspective, assuming all stakeholders have identical preferences and values, and designs non-adaptive policies optimized for a single realization of climate uncertainty. Concretely, no existing study simultaneously addresses all three aforementioned shortcomings of CBA. In this study, we introduce a multi-objective, multi-value IAM framework to guide just climate policy-making under deep uncertainty by examining how normative assumptions in IAMs influence global mitigation pathways and burden-sharing. Our contribution is twofold. We develop the JUSTICE IAM framework, a simulation-based optimization model that enables adaptive mitigation policy optimization across multiple objectives. In addition, we introduce a General Social Welfare Function integrated within JUSTICE that accommodates multiple principles of distributive justice and captures diverse value preferences relating to intra-generational and intergenerational equity, and risk aversion. These preferences are operationalized through normative parameters that aggregate welfare across regions, generations, and SOWs.

JUSTICE couples the peer-reviewed RICE50+ model⁵⁵, a successor of the CB-IAM RICE⁵⁶, with the FaIR climate model⁵⁷. The RICE50+ component represents 57 regions with calibrated marginal abatement cost curves and regional damage functions following Kalkuhl and Wenz⁵⁸, using regional Shared Socioeconomic Pathway (SSP) data⁵⁹. FaIR emulates complex Earth System Models from the Coupled Model Intercomparison Project Phase 6 (CMIP6) and captures key feedback mechanisms by accounting for nonlinearities in the carbon cycle. FaIR also enables probabilistic climate uncertainty assessments across different Representative Concentration Pathway (RCP) scenarios^{60,61}. Climate uncertainty in FaIR is quantified through Equilibrium Climate Sensitivity (ECS) and Transient Climate Response (TCR), with different combinations of these parameters captured through constrained ensemble members. These ensemble members are selected to be consistent with IPCC AR6 assessed ranges of ECS and TCR, historical climate data, and the successful reproduction of observed warming trends⁵⁷. The General SWF framework accommodates key distributive justice principles (Utilitarianism, Prioritarianism, Sufficiency, and Egalitarianism) that are central to climate justice discourse⁴⁰. This enables JUSTICE IAM to optimize adaptive mitigation policies across these four justice framings, each incorporating four distinct objectives reflecting diverse decision-maker preferences: economic welfare for purely economic actors, temperature objective for climate advocates, welfare loss from climate damages for vulnerable regions, and welfare loss from abatement costs for abatement-focused regions. Keeping deeply uncertain objectives separate is essential because the multiobjective setup allows varying aversion to these welfare losses a posteriori and clarifies trade-offs with other objectives³⁰.

For our analysis, we optimize an adaptive emission control rate policy rule (which adjusts based on temperature and rate of temperature rise feedback) under each justice framing, assuming global cooperation. We explicitly contrast these axiomatic SWFs from the perspective of a global social planner. Multiobjective optimization produces a Pareto-optimal set of solutions representing different preferences over objectives. We select a representative policy from each set to demonstrate how different preferences across objectives shape global mitigation pathways, end-of-century temperature rise, and the distribution of regional mitigation burdens.

Results

Making Values and Preferences Explicit

We introduce our General SWF framework in Equation (30) in the Methods section, capturing preferences across regions, timescales, and SOWs. The marginal utility of consumption (λ) is used as the default for all the SWFs. An integral component of the SWF framework is the 'ethical rule', which is a utility function that evaluates welfare distributions based on a normative preference¹⁶.

To reflect different principles of distributive justice, we incorporate distinct normative parameters within the ethical rule: Atkinson's priority

parameter (φ_N) for the Prioritarian SWF, the sufficiency threshold (c^{Thresh}) for the Sufficiency SWF, and the equality strictness parameter (ψ) for the Egalitarian SWF. The pure rate of time preference (φ_T), for time discounting, is applied in the Utilitarian and Sufficiency SWFs ($\varphi_T > 0$), while the Prioritarian and Egalitarian SWFs set it to zero ($\varphi_T = 0$). In addition, we include the risk aversion parameter (φ_S), which quantifies the social planner's risk preference across different SOWs. This step is vital to prevent undervaluation of catastrophic climate outcomes, especially when discounting leads to less stringent policy recommendations¹⁴. All normative parameters are listed in Table 1; different principles of distributive justice correspond to different values of these parameters. The selected parameter values allow a pure axiomatic comparison of the four distributive justice principles, ensuring adherence to the key SWF axioms (see the Supplementary Document for details). Importantly, the General SWF is flexible enough to allow different parameter choices and can accommodate hybrid versions of the justice principles.

The General SWF calculates regional welfare scores based on per capita consumption and welfare losses from damage and abatement costs under different distributive justice framings. We introduce an additional climate objective counting the number of years above the 2 °C temperature threshold. This creates a multiobjective optimization problem with four objectives: maximizing welfare and minimizing years above 2 °C, welfare loss from damage costs, and welfare loss from abatement costs. Using this setup, we optimize adaptive mitigation policies for four distributive justice framings via the General SWF. The SWF computes welfare and welfare losses (see Equations (36) and (37)) for all regions, years, and SOWs.

A Multiobjective Evolutionary Algorithm (MOEA) generates a Pareto-optimal solution set for the emission control rate policy in JUSTICE for each SWF. The Pareto set is a set of 'non-inferior' policies in which no objective can be improved without compromising at least one other. Each policy in the Pareto set corresponds to an emission control rate policy, specifying the decarbonization rate that each of the 57 regions should follow at every timestep and SOW. Figure 1 shows the full Pareto sets for each justice framing. From the top 10% of welfare-maximizing solutions, we select the policy with the best climate performance for analysis; this policy is highlighted in the figure, while others are grayed out.

For clarity, the preference directions on the parallel axes are aligned by flipping the temperature and welfare loss objectives so that all objectives have their preferred direction pointing upward. Additionally, the welfare and welfare loss axes are normalized, as their values are different across different justice framings (due to different nonlinear transformations). Unlike traditional CB-IAMs that produce a single optimal policy, our approach provides a rich set of solutions for each justice framing, revealing distinct trade-offs among the four objectives. Trade-off patterns differ markedly between the Utilitarian and Sufficiency framings versus the Prioritarian and Egalitarian ones.

While trade-offs between abatement and damage costs are well documented, their interaction with welfare and temperature objectives under different principles of justice is underexplored. Under the standard Utilitarian SWFs, trade-offs between welfare and damage or abatement objectives have been reported³⁰, and we observe the same pattern for the Utilitarian and the Sufficiency SWFs. In contrast, Prioritarian and Egalitarian framings show synergy between welfare and the temperature and damage objectives, as seen in Fig. 1b, d. All four framings exhibit trade-offs with welfare loss due to abatement costs.

Social Welfare Functions Rooted in Four Principles of Distributive Justice

The Utilitarian SWF's ethical rule aggregates the consumption per capita of all regions after adjusting for diminishing marginal utility. Consumption per capita is a standard proxy for well-being (or welfare) in the IAM literature. In contrast, the Prioritarian SWF adds a priority parameter that amplifies the gains and losses of less well-off regions. Unlike the Utilitarian SWF, the Prioritarian SWF is sensitive to the distribution of welfare and welfare losses across regions, placing greater emphasis on improving welfare and reducing

Table 1 | Normative Parameters and their values for different Social Welfare Functions

Normative Parameters	Utilitarian	Prioritarian	Sufficientarian	Egalitarian
Pure Rate of Time Preference (ρ_T)	0.015	0	0.015	0
Atkinson's Priority (ρ_N)	0	2	0	0.5
Risk Aversion (ρ_S)	0	0	0	0
Declining Marginal Utility (λ)	1.45	1.45	1.45	1.45
Sufficiency Threshold (c^{Thresh})	0	0	1.25 (USD PPP)	0
Equality Strictness (ψ)	0	0	0	1

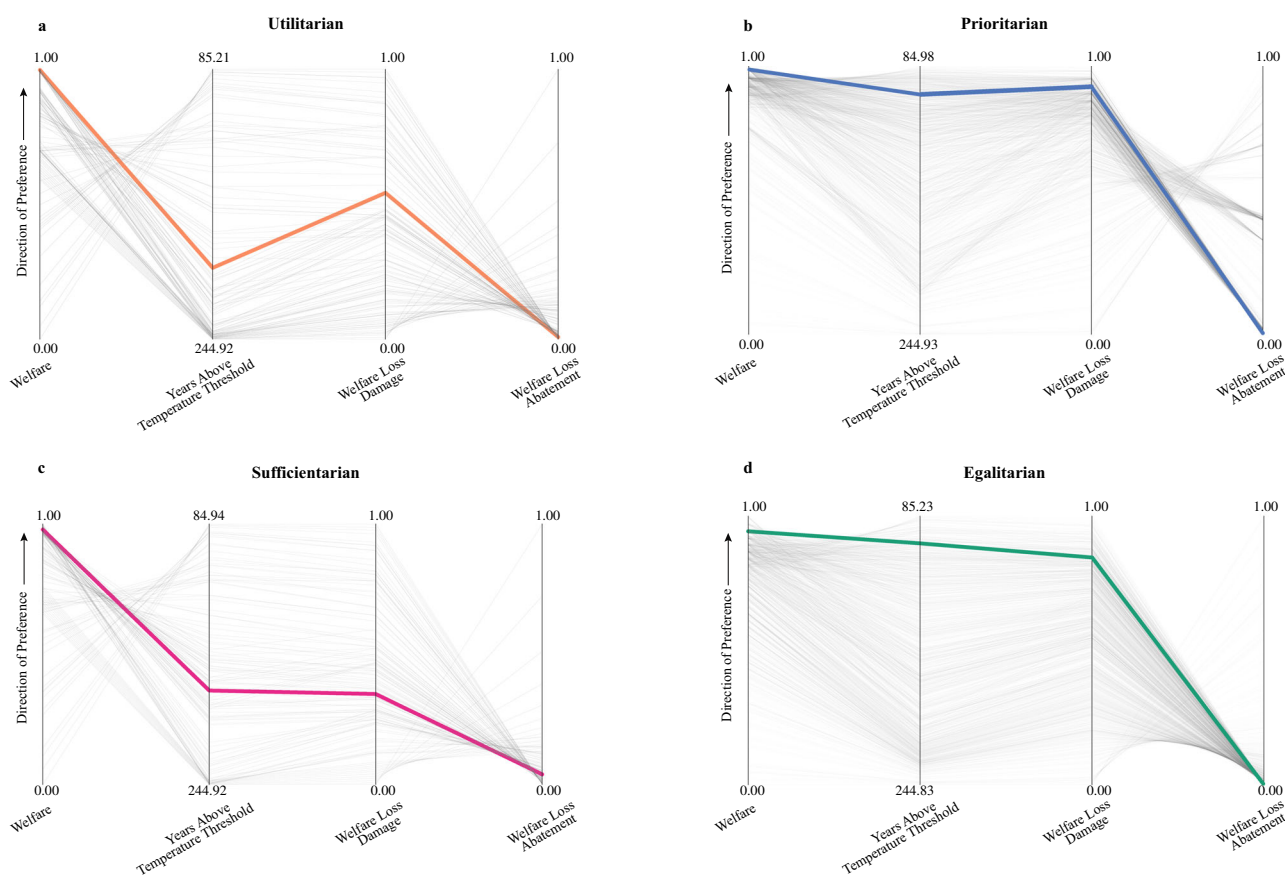


Fig. 1 | Parallel Axes Plots of the Pareto-optimal Policy Set for Four Justice Framings. a Utilitarian, **b** Prioritarian, **c** Sufficientarian, and **d** Egalitarian framings. Each panel contains four vertical axes representing the four optimization objectives: welfare, total number of years above the 2 °C temperature threshold, and welfare loss from damage and abatement costs. Each line across the vertical axes represents a Pareto-optimal emission control rate policy, in which no objective can be improved

without compromising another. All axes are oriented so that preferred directions point upward, and welfare and welfare loss axes are normalized across framings. The highlighted policy is the Pareto-optimal solution with the best climate performance among the top 10% of welfare-maximizing policies, while all other policies are grayed out.

welfare losses from climate damages and abatement costs for worse-off regions. It is well established in the literature that introducing inequality aversion tends to increase mitigation efforts⁵⁵. However, a comprehensive axiomatic comparison of the Utilitarian and Prioritarian SWFs has not been sufficiently explored. In particular, this includes how the distribution of regional mitigation burdens varies between SWFs, what unique trade-off patterns emerge for each SWF in a multi-objective setting, and how the selection of a candidate policy from the Pareto front shapes the intensity of regional mitigation. The Utilitarian and Prioritarian SWFs produce distinct burden distributions among regions, and the intensity of regional mitigation depends on how well the chosen policy performs on the welfare and temperature objectives (See the comparison between different Pareto-optimal policies from the Utilitarian and Prioritarian SWF in Supplementary Fig. 3). This added transformation in the Prioritarian SWF leads to distinct

mitigation outcomes compared to the Utilitarian SWF, as shown in Fig. 2, highlighting the distribution of average emission control rates across 57 regions. The Prioritarian approach prescribes strong mitigation till 2050, predominantly from developed nations, while Sub-Saharan Africa undertakes minimal mitigation. In contrast, the Utilitarian approach assigns more mitigation responsibility to least-developed and developing regions, such as those in Sub-Saharan Africa, than to countries like the United States. Changing the SWF from Utilitarian to Prioritarian significantly shifts the mitigation ambition and burden distribution.

The normalized stacked area plot in Fig. 2 illustrates the temporal distribution of average emissions and average mitigated emissions across ten macro-regions (57 regions grouped into ten macro regions following the R10 aggregation⁶²; see Supplementary Table S1). The total area represents SSP2 baseline emissions, with lighter colors indicating mitigated emissions

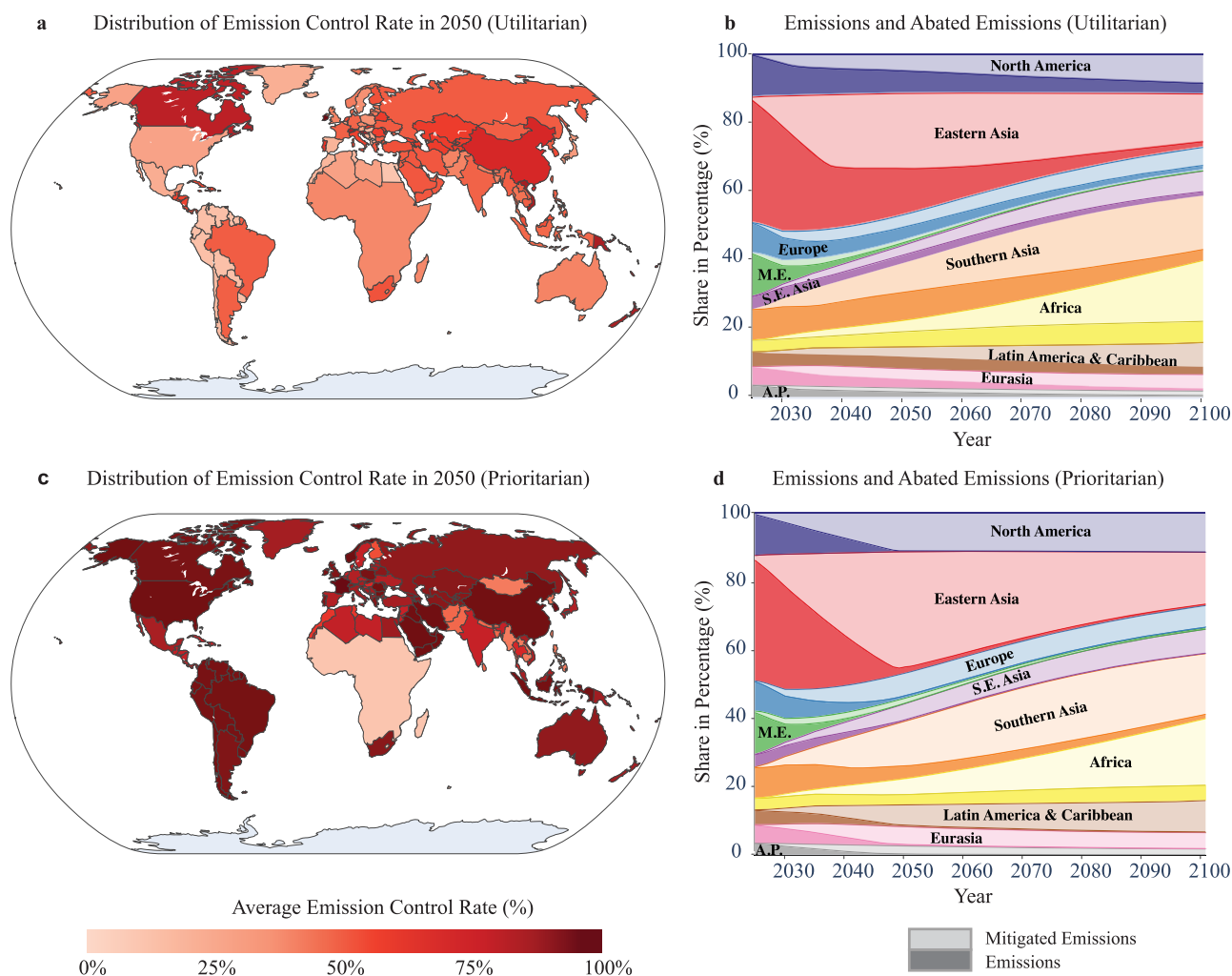


Fig. 2 | Distribution of Mitigation Burdens and Emissions. **a** Global distribution of average emission control rates in 2050 and **b** Temporal evolution of average emissions and average mitigated emissions under the Utilitarian framing, shown as a normalized stacked area plot. **c** Global distribution of average emission control rates in 2050 and **d** Temporal evolution of average emissions and average mitigated emissions under the Prioritarian framing, shown as a normalized stacked area plot.

The emission control rate is expressed as a percentage of the SSP2 baseline emissions mitigated. In the stacked plots, the lighter area represents mitigated emissions, and the darker area represents total emissions, with values normalized relative to the SSP2 baseline. The 57 regions are grouped into ten macro-regions following the R10 aggregation. Middle East and Asia Pacific regions are abbreviated as M.E. and A.P. respectively.

and darker colors indicating CO₂ emissions from fossil fuels and industries. Under the Prioritarian SWF, net zero emissions are achieved on average across all SOWs for well-off regions such as North America (United States and Canada), Eastern Asia (China, South Korea, and other East and Southeast Asian countries), Europe, the Middle East, Southeast Asia, Latin America and the Caribbean, Eurasia, and Asia Pacific (Australia and Japan), while developing regions such as Southern Asia (India and other South Asian countries) and Africa maintain limited emissions until 2100. In contrast, the Utilitarian approach allows continued emissions in North America and Eastern Asia until 2100. The Prioritarian policy emphasizes the welfare and welfare loss of Least Developed Countries (LDCs), Small Island Developing States (SIDS), and other developing nations, recognizing their disproportionate vulnerability, in alignment with the IPCC AR6¹³. While all regions must mitigate to achieve a sustainable climate future, the Prioritarian approach places more near-term responsibility on developed nations, prohibiting their emissions after 2050. It requires China to achieve net zero by 2050, which is earlier than its 2060 target and contrasts with its current coalition stance in climate negotiations. This aligns with Robiou du Pont et al.⁷, who found China's nationally determined contributions to be less ambitious than equitable burden-sharing suggests. Findings from the Prioritarian framing are particularly relevant to discussions of equity and

fairness, as highlighted in COP meetings regarding Loss and Damage for LDCs, SIDS, and other developing nations. We retain the Utilitarian SWF for comparative purposes with other IAM studies.

The ethical rule in the General SWF in Equation (30) can incorporate various principles of distributive justice beyond Utilitarian and Prioritarian views. Here, we use the Sufficiency and Egalitarian SWFs to calculate welfare and use these SWFs for the multiobjective optimization. Sufficiency extends the Utilitarian SWF by introducing a sufficiency threshold corresponding to the international poverty line of \$1.25 United States Dollars (adjusted to 2005 Purchasing Power Parity) stipulated by the World Bank. The sufficiency threshold in the General SWF is a vector of 57 values corresponding to the 57 regions, which allows different sufficiency thresholds based on regional economic conditions. For this study, we use \$1.25 for all the regions. The Egalitarian approach extends the Prioritarian framing by integrating a measure of relative inequality using the GINI index that scales the welfare score. The impact of this GINI scaling is controlled using the equality strictness parameter (ψ), which adjusts how much weight is given to relative inequality in the final welfare score. The Egalitarian SWF extends the Prioritarian SWF to satisfy the Fundamental Axioms of SWF, essential for the intercomparison between the four SWFs (see Supplementary Document for detailed explanation). We focus on pure Egalitarian

Fig. 3 | Comparison of Median Emission Pathways and Median Global Temperature Rise for the Four Ethical Framings. **a** Global Emission Pathways; **b** Global Temperature Rise Projections for Utilitarian, Prioritarian, Sufficitarian, and Egalitarian Framings. Each pathway is the median across all the states-of-the-world from FaIR ensemble.

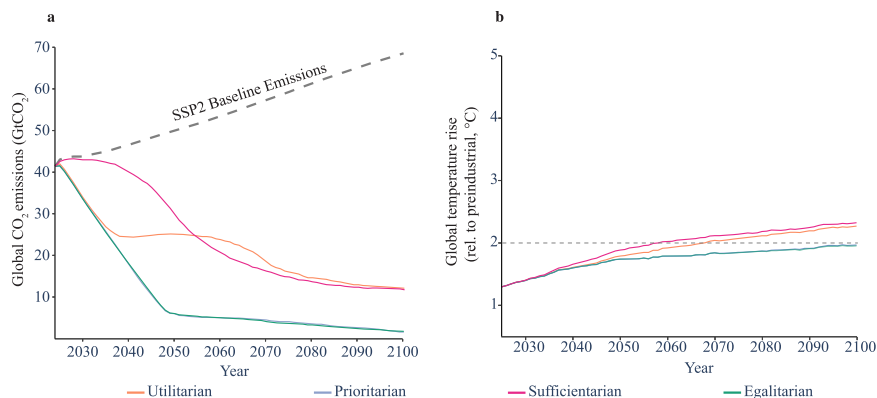
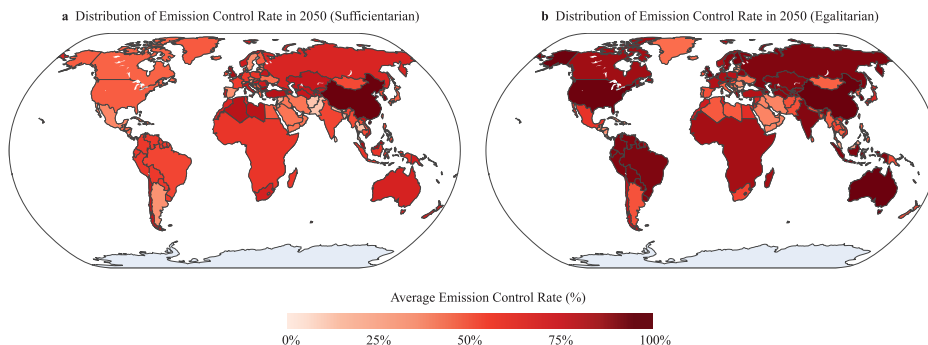


Fig. 4 | Distribution of Mitigation Burdens for Sufficitarian and Egalitarian Framings. **a** Global distribution of average emission control rates in 2050 under the Sufficitarian framing. **b** Global distribution of average emission control rates in 2050 under the Egalitarian framing. The emission control rate is expressed as a percentage of the SSP2 baseline emissions mitigated.



preference and thus set the equality strictness parameter to ($\psi = 1$), leaving exploration of heterogeneous sufficiency thresholds and varying equality strictness for future work.

Figure 3 compares the global median emission and median temperature pathways for all the four SWFs (median calculated across the SOW). In addition, Fig. 4 shows the average emission control rate in different regions under the Sufficitarian and Egalitarian SWFs. Unlike the Utilitarian approach, the Sufficitarian allows for economic growth in developing nations and does not prescribe rapid near-term mitigation (Fig. 3a). However, the Sufficitarian policy intensifies mitigation around 2050, surpassing the Utilitarian levels while maintaining a similar regional distribution of the mitigation burden as seen in Fig. 4a. The Egalitarian policy suggests a mitigation pathway similar to the Prioritarian approach, recommending significant reductions leading up to 2050. Yet, due to its emphasis on the relative positions of regions (i.e. how equally welfare and welfare losses are distributed), the mitigation burden is more evenly distributed (Fig. 4b). This approach has a levelling-down effect, meaning it makes everyone worse off in an attempt to equalize welfare and welfare losses, which goes against the Common but Differentiated Responsibilities and Respective Capabilities principle adopted by the United Nations Framework Convention on Climate Change (UNFCCC).

The Sufficitarian mitigation pathway results in the highest temperature rise by the end of the century because it does not recommend rapid near-term mitigation, as evidenced by its emission pathway and temperature trajectory (See Fig. 3). Like the Prioritarian policy, the Egalitarian policy keeps the median warming below 2 °C by recommending near-term rapid cuts in emissions. This rapid mitigation underscores the urgent need for near-term rapid and sustained mitigation actions on a global scale, as highlighted by the IPCC AR6, to maintain warming below 2 °C^{5,13}.

This comparison in Fig. 3 highlights how justice framings substantially influence optimal mitigation pathways, temperature goals, and the regional distribution of mitigation efforts. In particular, although stakeholders may agree on high-level goals such as limiting warming to below 2 °C and

achieving net-zero emissions by 2050, they often differ on how mitigation burdens should be apportioned across regions, as illustrated by the similar median emission and temperature pathways under the Prioritarian and Egalitarian SWFs (Fig. 3a) and their divergent underlying burden distributions (Figs. 2c and 4b). In the following section, we investigate how climate uncertainty reshapes these mitigation pathways and how alternative policy selections from the Pareto front affect the regional distribution of mitigation efforts.

Adaptive Mitigation Pathways under Climate Uncertainty

CB-IAMs produce a single optimal path, a problematic assumption given the uncertainty in key climate-economy interactions⁴⁴. Most IAMs use average estimates¹⁵ and assume perfect foresight¹² to optimize non-adaptive policies. This rigidity leads to poor outcomes, particularly when climate responses deviate from expectations⁶³. Non-adaptive policies thus risk high regrets if future conditions deviate from those assumed during optimization. In contrast, adaptive policies allow dynamic adjustments, balancing socioeconomic and climate goals³⁰, making them robust under climate uncertainty⁵⁴ and more just³¹.

In our optimization setup, the emission control rate policy is adaptive, using both temperature rise and its rate as feedback (see Equation (41)). This adaptive policy addresses climate uncertainty by avoiding optimization for a single realization of the climate uncertainty. Adaptivity enables dynamic adjustment of regional mitigation ambitions based on different SOWs, represented by the FaIR climate ensemble. Specifically, mitigation rates increase if a SOW indicates greater warming and decrease otherwise. Within the JUSTICE IAM, the emission control rate lever is parameterized via a Radial Basis Function (RBF)-based policy rule that acts as a global approximator. This policy takes temperature rise and rate feedback as inputs and outputs regional emission control rates per year for all SOWs (see Adaptive Policy Section in Methods and Supplementary Document). At each timestep, the RBF network processes the feedback from all FaIR ensemble members, generating heterogeneous regional mitigation actions

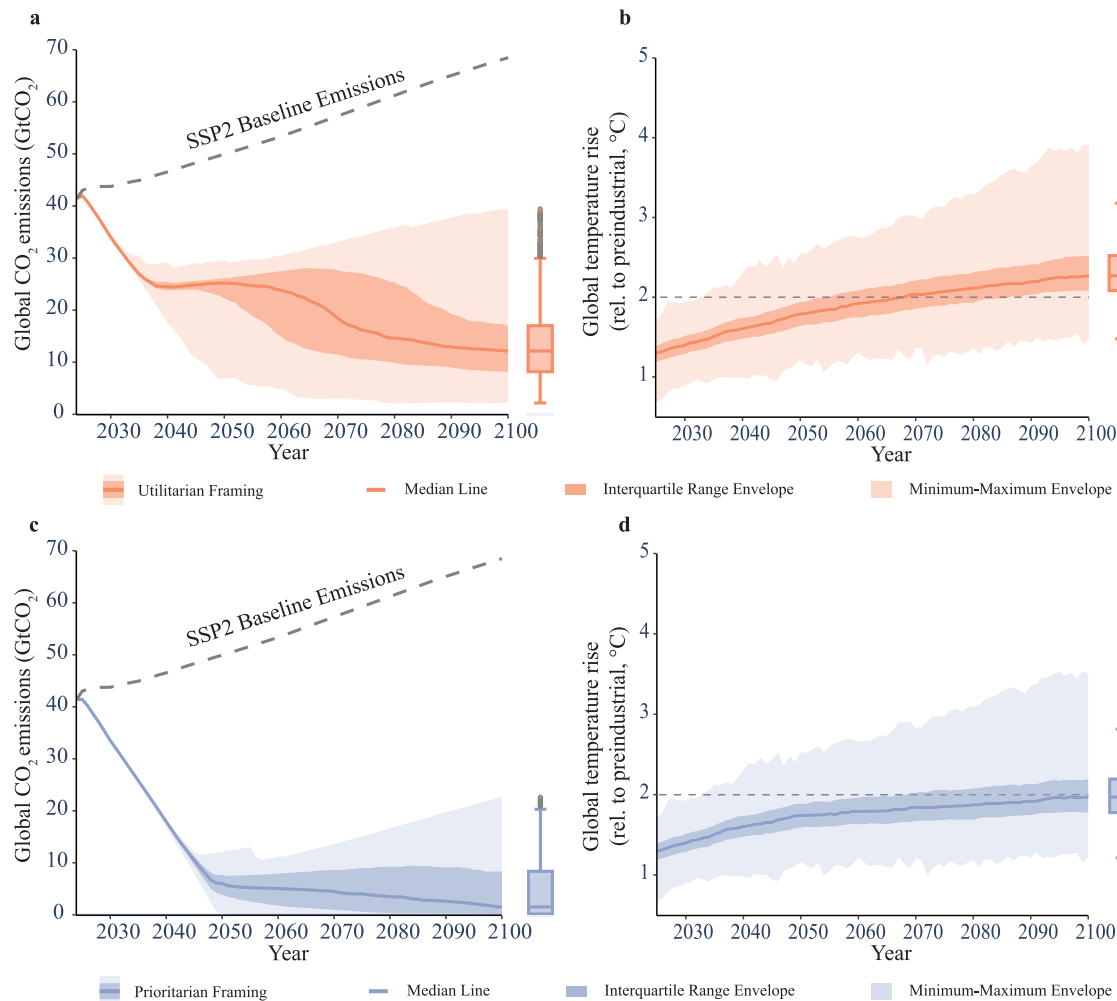


Fig. 5 | Comparison of Adaptive Global Emission Pathways and Global Temperature Projections for Utilitarian and Prioritarian Framings. **a** Global emission pathway and **b** Global temperature rise under the Utilitarian framing. **c** Global emission pathway and **d** Global temperature rise under the Prioritarian framing.

Pathways are shown for 2025–2100. The solid line indicates the median, the dark envelope represents the interquartile range (25th to 75th percentile), and the lighter envelope shows the full range (minimum to maximum) across all the states-of-the-world from the FaIR ensemble.

for all 57 regions and SOWs in parallel. This feedback mechanism enables mitigation efforts to adjust dynamically in response to the climate sensitivity of each SOW, as captured by the FaIR ensemble members. The sequential process of taking temperature rise and its rate feedback from the current timestep to calculate emission control rates for the next timestep reflects sequential decision-making. This approach mirrors real-world climate decisions, which are made sequentially to enable learning about climate sensitivity and revise policies accordingly, helping to hedge against extreme climate outcomes³⁰. Figure 5 illustrates how global emission pathways adapt based on temperature projections from FaIR for Utilitarian and Prioritarian framings. The plot's envelope covers all ensemble members representing different SOWs, with darker bands indicating the interquartile range and a median line at the centre, calculated across all the ensemble members. Emission pathways increase mitigation when observed sensitivity exceeds expectations, as shown in Fig. 5.

Both Utilitarian and Prioritarian global mitigation pathways agree on the need for rapid near-term emission cuts until at least 2040 (for all the SOWs, as there is no variation in the emission pathways), underscoring the urgent need to reduce current emissions to ensure a safe and sustainable climate. While Utilitarian emission pathways show a moderate decline in emissions until 2050, Prioritarian pathways advocate for more rapid reductions, as expected. As a result, over half of FaIR's ensemble members project a global temperature rise below 2 °C by 2100 for the Prioritarian SWF. The Prioritarian pathway also shows a plateauing trend, indicating

minimal temperature overshoot in the next century. In contrast to the Utilitarian approach, the median temperature rise surpasses 2 °C just before 2070 and continues to grow into 2100. This overshoot occurs because global emissions are sustained despite some reduction compared to the SSP2 baseline. The slower emission reduction under Utilitarian pathways results from the discount rate used in the Utilitarian SWF, which diminishes the perceived future damages, leading to a lack of urgency.

In contrast, the Prioritarian approach advocates for rapid, deep, and sustained emission reductions due to the severity of future damages in high-emission scenarios and SOWs with high climate sensitivity. This accelerated mitigation stems from the priority parameter and the absence of time discounting. Prioritarians reject time discounting because it violates the anonymity axiom¹⁶, which holds that everyone's welfare is equally important regardless of their identity, time, or location (see Supplementary Document for details). Since climate damages disproportionately affect LDCs, SIDS, and other developing nations—who have contributed least to climate change and possess limited capacity to mitigate its effects—the Prioritarian framework places greater emphasis on improving their welfare and reducing their welfare loss. This results in drastic emission reductions, reaching net zero in some SOWs with high climate sensitivity by 2050, as shown by the envelope coinciding with 2050 on the x-axis in Fig. 5.

Recall that the risk aversion parameter is set to zero in all the SWFs, implying a risk-neutral global social planner who assigns equal weights to all SOWs. The final welfare score calculated thus takes an average over the

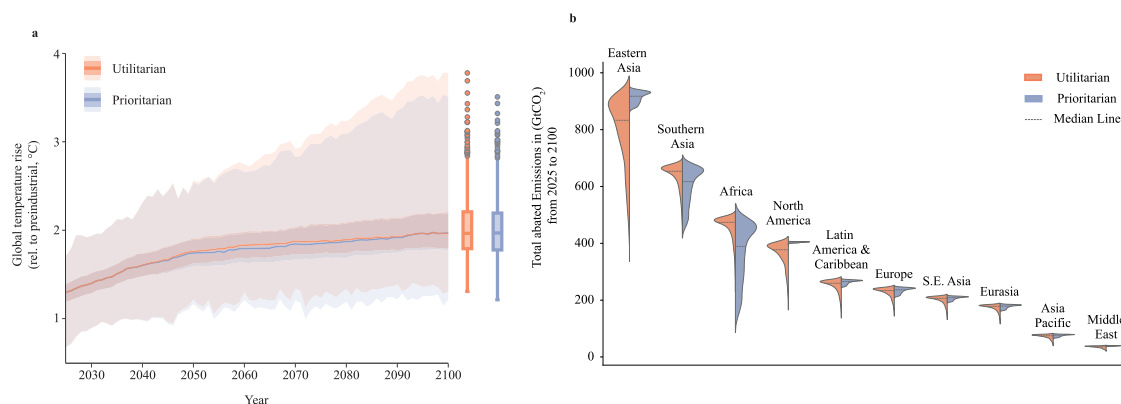


Fig. 6 | Comparison of Utilitarian and Prioritarian Pareto-Optimal Policies with Similar Climate Outcomes. (a) Comparison of global temperature rise projections between Utilitarian and Prioritarian Pareto-optimal policies. The solid line indicates the median, the dark envelope represents the interquartile range (25th to 75th percentile), and the lighter envelope shows the full range (minimum to maximum)

across all the states-of-the-world from the FaIR ensemble. **(b)** Violin plot of total abated emissions by R10 macro-region (relative to the SSP2 baseline) comparing Utilitarian and Prioritarian framings. The distribution in the violin illustrates the variability of total abated emissions across different states-of-the-world.

SOWs. However, since we implement an adaptive policy with sequential decision-making across all SOWs, the planner remains sensitive to extreme climate outcomes and adjusts mitigation rates accordingly. Risk neutrality only affects the final welfare score calculated at the end of the model run. Incorporating risk aversion would heighten sensitivity to extreme climate scenarios; however, as discussed in the Supplementary Document, doing so would prevent a clean axiomatic comparison among SWFs grounded in distinct distributive justice principles. Thus, we choose risk neutrality to facilitate a clear comparison between the different SWFs. From the planner's perspective, she is fully aware of all SOWs and adapts mitigation dynamically, even under risk neutrality.

Disaggregation of Incommensurate Objectives and the Pareto Front

In the previous sections, we illustrate how different preferences for distributive justice shape mitigation ambition and regional burden distribution for policies selected based on balancing trade-offs between welfare and temperature objectives. All policies in the Pareto front for the four justice framings capture different trade-offs across the four objectives, reflecting varying stakeholder preferences. In contrast, single-objective optimization implicitly assigns relative importance to different objectives before aggregating them into a single objective, which is then optimized to produce a single solution⁶⁴. Aggregating conflicting objectives into a single objective can cause some objectives to dominate over others in unpredictable ways^{17,19,65}. This aggregation also biases policy recommendations toward specific groups; hence, the literature recommends using methods that explicate preferences across different climate and socioeconomic objectives, which clarify trade-offs and stakeholder priorities for fair and just climate policy design^{66,67}.

When combined with different justice framings, multi-objective optimization makes explicit both diverse stakeholder preferences across objectives and values rooted in different principles of justice. Unlike single-objective optimization, multi-objective approaches reveal rather than obscure trade-offs between competing objectives, enabling informed decision-making. In contrast, relying on a single aggregated objective excludes diverse perspectives and risks reinforcing status-quo interests by obscuring normative assumptions in policy recommendations. This creates opportunities for interest groups to politicize scientific insights for national or sectoral advantage³⁷. Such omission erodes trust and undermines legitimacy, as policies perceived as unfair face public resistance and ultimately become a barrier to climate action. By generating a Pareto set of solutions for each justice framing, this approach engages multiple perspectives and ensures that mitigation policies reflect varied priorities and perspectives, allowing stakeholders to deliberate.

Persistent Inequitable Mitigation Burden in Utilitarian for Similar Temperature Goals

To illustrate the interaction between preferences over objectives and values rooted in different distributive justice principles, we select a Utilitarian policy from the Pareto front that achieves a temperature objective comparable to that of the Prioritarian policy from the previous section. This new Utilitarian solution keeps the median temperature rise below 2 °C in 2100, aligning with the performance of the Prioritarian solution shown in Fig. 6a. However, it comes with a cost: this Utilitarian policy scores poorly on Utilitarian welfare. This disparity illustrates the fundamental trade-off between Utilitarian and Prioritarian framings. Despite achieving similar temperature outcomes, the two solutions differ significantly in regional cumulative abated emissions, as illustrated by the violin plot in Fig. 6b. From Fig. 6b, it is clear that the Utilitarian solution provides greater mitigation flexibility across all SOWs to Eastern Asia and North America while constraining Africa, whereas the Prioritarian solution reverses these patterns, offering Africa increased flexibility under climate uncertainty. The spread in cumulative abated emissions across SOWs reflects how adaptive mitigation policies respond differently to the different SOWs from the climate model. Greater spread indicates higher flexibility, as total cumulative abatement varies with the SOW; conversely, low spread indicates that a region must meet similar mitigation requirements regardless of the SOW, leaving little room for adaptive adjustment.

Examining regional emission pathways in Fig. 7 reveals further differences in mitigation flexibility. The Utilitarian solution grants greater emissions flexibility to China and the United States while constraining Sub-Saharan Africa across SOWs. In contrast, the Prioritarian solution gives Sub-Saharan Africa increased flexibility under climate uncertainty, requires China and the United States to reach net zero by 2050 in nearly all SOWs (Fig. 7a, b), and allows India minor flexibility to emit if future conditions are favorable. As shown in Supplementary Fig. 3, the distribution pattern between Utilitarian and Prioritarian policies differs considerably. Utilitarian policies distribute emission control rates relatively evenly between developed and developing countries, concentrating the heaviest burdens on least-developed nations in Sub-Saharan Africa and South Asia while imposing minimal mitigation on Russia. Prioritarian policies reverse this pattern by shifting burdens to developed nations such as the United States, Canada, Russia, and Europe, whereas inequities intensify under Utilitarian framing when policies are optimized for the temperature objective. This extended comparison further substantiates our finding that inequitable mitigation burdens under Utilitarian framing are persistent (see Supplementary Document for an Extended Analysis).

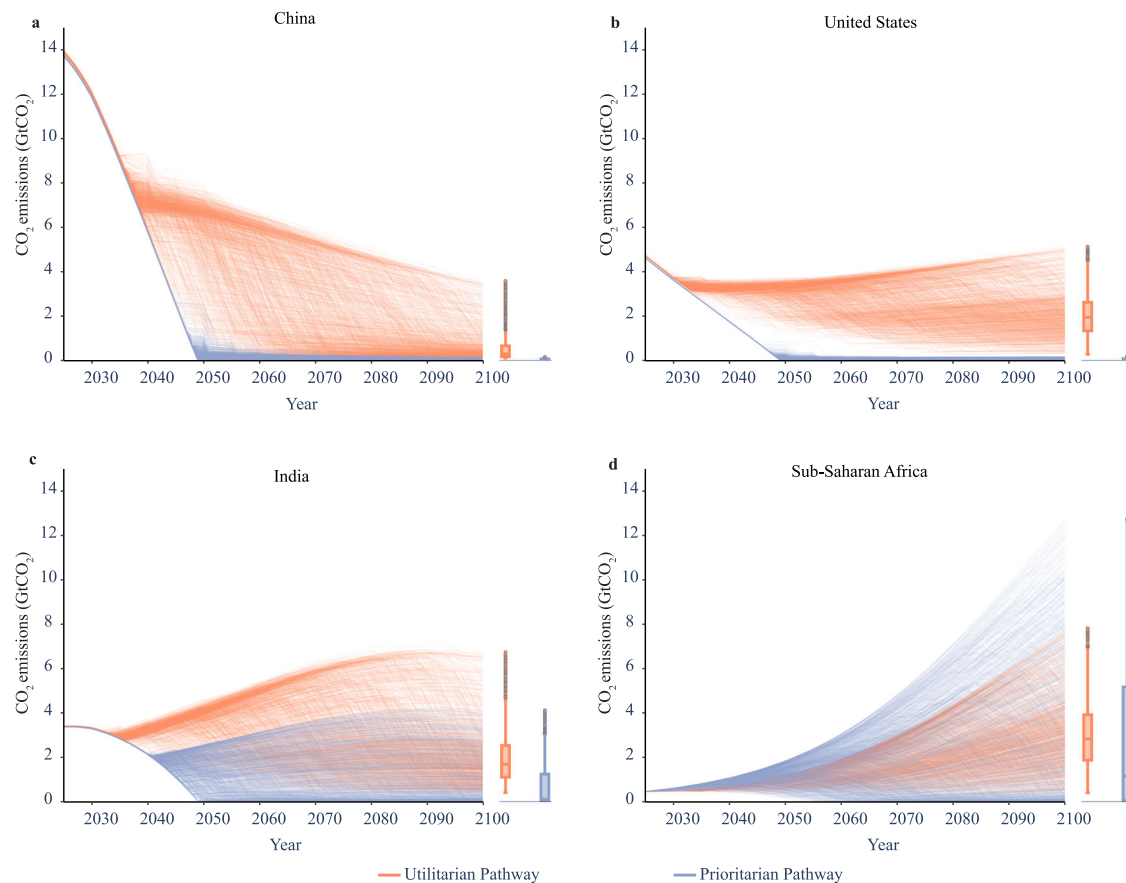


Fig. 7 | Comparison of Regional Emission Pathways for Utilitarian and Prioritarian Framings. **a** China, **b** United States, **c** India, **d** Sub-Saharan Africa. Emission trajectories from 2025 to 2100 are shown, with each pathway representing a state of the world from the FaIR climate ensemble.

Discussion

Our results show that the notion of optimality in CB-IAMs is not a neutral modelling choice but a normative one. One must always ask: optimal for whom, when, where, and under which scenario? While CB-IAMs are crucial for evaluating mitigation policies and identifying key uncertainties¹⁴, optimizing solely based on a single objective is inadequate for fair and just climate policy design. Prematurely aggregating diverse stakeholder preferences into a single metric can substantially distort policy recommendations and exacerbate existing injustices. In our multi-objective, multi-value framework, making normative assumptions explicit and separating incommensurate objectives clarifies trade-offs and reveals distributional consequences otherwise obscured under single objective optimization. We demonstrate how different justice framings shape both the urgency and distribution of mitigation burdens. Although we do not promote a specific justice framing, our findings reveal multiple pathways consistent with a 2 °C target, each with markedly different regional burden distributions. Notably, the Prioritarian approach proposes equitable pathways that align with the UNFCCC's principle of Common but Differentiated Responsibilities and Respective Capabilities, thereby upholding both intra- and intergenerational justice.

CB-IAMs also optimize non-adaptive mitigation policies under a single objective. Non-adaptive optimal policies in CB-IAMs are fixed a priori, optimized for a single assumed climate sensitivity, and remain unchanged regardless of actual climate evolution^{30,47}. This rigidity leads to poor performance and high regret if climate sensitivity differs from initial assumptions. In contrast, our framework implements adaptive policies where temperature feedback at each timestep revises mitigation actions, making policies responsive to stochastic climate realizations. This iterative, evidence-based approach helps overcome policy deadlocks and fosters collaborative convergence between diverse stakeholders⁶⁸. Adaptive

strategies improve economic and environmental outcomes, avoid shifting undue risks to vulnerable communities and future generations, and thus remain fair and robust under deep uncertainty^{30,48}. From a governance perspective, this mirrors real-world policymaking, where policies are revised based on performance, such as mitigation plans negotiated at Conference of Parties (COP) meetings or IPCC assessments that update recommendations based on new climate and economic data. By prioritizing adaptability over optimality, adaptive policies recognize that real-world climate policymaking must remain flexible in the face of deep uncertainty.

Since CB-IAMs are highly aggregated, their modelling assumptions define what is considered “optimal”, necessitating scrutiny of these assumptions and their policy implications. The damage function we used assumes a functional form calibrated on historical data⁵⁸. Extrapolating past data with a smooth functional form inevitably misses catastrophic damages and omits damage to capital, total factor productivity, non-market damages, and sea-level rise impacts. We also did not model carbon markets, redistributive transfers between regions, or technology sharing to enable developing economies to leapfrog high-emission growth. A key problem with CB-IAMs is their failure to consider increasing returns to scale for clean energy technologies⁴⁴. As a result, most CB-IAMs treat mitigation as a net societal cost, neglecting its benefits²⁰, thereby inflating abatement costs. This assumption explains the trade-offs between welfare and welfare loss from abatement costs observed across all justice framings in our analysis. Finally, our General SWF framework is forward-looking and does not explicitly account for historical emission disparities. However, the framework partially captures the effects of such disparities through regional differences in economic capabilities, which correlate with past emissions.

Another limitation is our exclusion of Negative Emission Technologies (NETs) before 2100. While NETs are expected to play a crucial role in achieving long-term climate stabilization, particularly if near-term

mitigation is delayed, incorporating NETs would significantly increase model complexity, requiring detailed representation of land-use allocation among other competing objectives such as food security and biodiversity, along with their justice dimensions⁶⁹. Bioenergy with carbon capture and storage, for example, competes with agriculture and conservation for land, creating complex trade-offs among climate, biodiversity, and economic objectives. Additionally, NETs remain deeply uncertain in terms of scalability, costs, and storage capacity^{70,71}, and large-scale deployment faces technical and ethical challenges. Excessive reliance on NETs risks delaying near-term mitigation, shifting costs and risks to future generations⁷². Our current setup follows DICE and RICE50+ assumptions where abatement represents energy system decarbonization, not NET deployment^{30,55}. We demonstrate that different framings of justice achieve Paris-aligned goals in certain SOWs without explicitly relying on NETs.

We also inherit another limitation of CB-IAMs in representing human well-being. Our analysis represents welfare primarily through economic consumption, a standard proxy in the IAM literature. However, human well-being is multidimensional, encompassing economic well-being, health, nutrition, energy access, land-use patterns, biodiversity integrity, and ecosystem services. Aggregating these into a single metric requires arbitrary weighting and assumptions about substitutability across incommensurable dimensions, reflecting subjective normative views²³. We also did not model within-country inequality in our regional aggregation. Incorporating within-country inequality by disaggregating regions into income groups would substantially increase estimates of global inequality and strengthen the case for more aggressive mitigation and burden-shifting toward developed nations⁷³.

Notwithstanding these limitations, our results demonstrate that making normative assumptions explicit, disaggregating conflicting policy objectives, and employing adaptive policies are crucial for designing just mitigation strategies. We find that the normative assumptions ingrained in CB-IAMs—both preferences over objectives and underlying value judgments—fundamentally shape the ambition and fairness of the resulting climate pathways. Utilitarian framing results in less ambitious action and uniform burden-sharing, whereas Prioritarian framing yields outcomes aligned with the Paris Agreement and the UNFCCC's principle of Common but Differentiated Responsibilities. The distribution of the mitigation burden is determined by the SWF and remains consistent across the Pareto front, while individual policies on the front vary in their overall mitigation levels. Our framework demonstrates that multi-objective, value-pluralistic, and adaptive policy design is computationally feasible and reveals insights that single-objective optimization fails to illuminate. Rather than presenting a single “optimal” policy, we provide decision-makers with defensible alternatives paired with its distributional consequences, supporting more inclusive and transparent policy dialogues. Our results empirically validate long-standing critiques of cost-benefit approaches: they systematically disadvantage developing nations and vulnerable populations, not inevitably but as a consequence of specific normative choices embedded in them. We also demonstrate that equity and net-zero ambitions are mutually reinforcing, and justice-focused approaches are not constraints on climate ambition but drivers of it. This finding has important implications for climate negotiations: the choice of justice framing is not trivial but a fundamental determinant of policy outcomes. As nations continue negotiating nationally determined contributions and long-term climate strategies, engaging with the diversity of justice principles can facilitate more legitimate, transparent, and effective collective action. By making the normative assumptions of climate policy explicit for deliberation, we contribute to more just policy design that is inclusive of diverse stakeholder preferences and values—an important aspect of procedural justice in climate policymaking⁴⁰. Future research should extend this framework to multidimensional welfare metrics, incorporate within-country inequality, model NETs and their land-use implications, and test the deliberative potential of Pareto-front decision support in real policy contexts. Additional work should explore the sensitivity of results to different normative parameters and conduct robustness analysis of Pareto-optimal policies across alternative SSP scenarios.

Methods

JUSTICE

The JUSTICE Integrated Assessment Model (IAM) is a modular framework that builds on the RICE50+⁵⁵ model's economy, abatement, and damage modules. JUSTICE comprises 57 independent regions and utilizes a yearly temporal resolution, unlike the five-year intervals of RICE50+. JUSTICE employs the Shared Socioeconomic Pathway (SSP) dataset⁵⁹ for exogenous regional economic growth, carbon intensity and population projections and the Representative Concentration Pathway (RCP) dataset⁷⁴ for the exogenous emissions trajectories for other greenhouse gases and aerosols. The SSPs comprise a series of alternative socioeconomic development trajectories based on five socioeconomic narratives, while RCP scenarios represent various future greenhouse gas concentrations and climate forcings. The SSP-RCP offers a consistent framework for synthesizing socioeconomic and climate research. The exogenous SSP data are represented by the superscript (') in the equations. The economy, damage and abatement modules are coupled with the FaIR climate model that allows probabilistic assessments of climate policies under different climate sensitivity scenarios through the constrained calibrated ensemble members. Therefore, calculations in JUSTICE take place over annual time steps (T), the regions (N), and the SOW (S) determined by the FaIR climate ensemble. The set (T) is an ordered set containing discrete time steps (Equation (1)), set (N) representing the individual regions (Equation (2)), and set (S) contains the ensemble members representing the SOWs (Equation (3)). The computation of data in JUSTICE is represented by a hypermatrix with dimensions TNS (e.g., A_{TNS}), and all subsequent operations defined in the equations are element-wise.

$$T = \{t_0, t_1, \dots, t_i\} \text{ where } t_0 < t_1 < \dots < t_i \quad (1)$$

$$N = \{n_1, n_2, \dots, n_j\} \quad (2)$$

$$S = \{s_1, s_2, \dots, s_k\} \quad (3)$$

Economy

The Economy module employs the neoclassical Cobb-Douglas production function (Equation (4)) from the established DICE and RICE IAM⁷⁵. In Equation (4), Y_{TNS}^{GROSS} represents gross economic output, TFP_{TNS}^* denotes exogenous Total Factor Productivity, and I_{TNS}^* indicates exogenous labour projections derived from the SSP dataset. Capital K_{TNS} is raised to the power of α (element-wise power denoted by $\circ$). In contrast, labour is raised to the power of $1 - \alpha$, with α being the capital elasticity of the production function. Capital dynamics at each time step, t_i , is calculated using the capital from the previous time step, t_{i-1} , adjusted by the annual capital depreciation rate δ , and adding the investment I_{t_iNS} from the current time step, as shown in Equation (5).

$$Y_{TNS}^{\text{GROSS}} = \text{TFP}_{TNS}^* \cdot K_{TNS}^{\circ\alpha} \cdot L_{TNS}^{* \circ 1-\alpha} \quad (4)$$

$$K_{t_iNS} = (1 - \delta) \cdot K_{t_{i-1}NS} + I_{t_iNS}, \forall t_i \in T \quad (5)$$

Emission

Economic output from the economy model is converted to carbon dioxide emissions, EMIS_{TNS} , using exogenous carbon intensity, CI_{TNS}^* , which represents the share of fossil fuels used for energy in economic production (Equation (6)). The Agriculture, Forestry, and Other Land Uses (AFOLU) emissions are also exogenous, with both Carbon Intensity and AFOLU emissions AFOLU_{TNS}^* obtained from the SSP database. The emission control rate, ECR_{TNS} , is the model's policy lever that sets the mitigation rate at each time step, thereby reducing the baseline (SSP 2) emissions.

$$\text{EMIS}_{TNS} = \text{CI}_{TNS}^* \cdot Y_{TNS}^{\text{GROSS}} \cdot (1 - \text{ECR}_{TNS}) + \text{AFOLU}_{TNS}^* \quad (6)$$

Abatement Cost

Regions assess their abatement costs by mitigating a fraction of baseline emissions through the emission control rate ECR_{TNS} , ranging from 0 (no mitigation) to 1 (full mitigation). These regional abatement costs are calculated using marginal abatement cost curves (MACC) calibrated using Enerdata-EnerFuture data, based on the detailed process-based model POLES⁷⁶, which provides carbon prices for industrial carbon dioxide abatement levels ranging from 2025 to 2040. From 2040 to 2100, carbon price projections from the SSP database are used to derive additional MACCs. These MACCs are used to calibrate the region-specific abatement coefficients AC1_{TNS} and AC2_{TNS} , as well as a correction factor V_{TNS} (See Gazzotti et al.⁵⁵). Finally, the abatement cost is linked to the ECR_{TNS} policy lever to calculate the mitigation cost per time step, region, and SOW, as shown in Equation (7).

$$\text{ABATECOST}_{TNS} = \text{V}_{TNS} \cdot \text{EMIS}_{TNS}^* \cdot \left(\frac{\text{AC1}_{TNS}}{2} \text{ECR}_{TNS}^2 + \frac{\text{AC2}_{TNS}}{5} \text{ECR}_{TNS}^5 \right) \quad (7)$$

FaIR Climate Model

The Finite-amplitude Impulse Response (FaIR) is a simplified climate model designed to emulate the behaviour of more complex Earth System Models from the Coupled Model Intercomparison Project Phase 6 (CMIP6). FaIR employs a computationally efficient six-equation model to simulate the global mean climate response to 52 greenhouse gases (GHGs) from 1750 to 2100⁶⁰. FaIR's efficiency enables the exploration of climate uncertainty through a constrained calibrated ensemble, quantifying uncertainty using Equilibrium Climate Sensitivity (ECS) and Transient Climate Response (TCR). This constrained ensemble was derived from an initial prior of 1.5 million members, which sampled uncertainties in the carbon cycle, effective radiative forcing, and thermal response. A posterior sample of 1,001 members was then selected to simultaneously span IPCC AR6 assessed ranges of ECS and TCR, ocean heat content, global mean surface temperature, and aerosol effective radiative forcing, consistent with historical climate observations⁵⁷. These members accurately reproduce historical warming, each representing a potential "SOW" in which the social planner has perfect knowledge of the climate system's state⁵⁷. The GHG emissions used to calculate the concentration are derived from total CO₂ emissions, which include emissions from fossil fuels and industrial processes (CO₂ FFI) modelled by JUSTICE IAM, and CO₂ emissions from agriculture, forestry, and other land use (AFOLU). The AFOLU emissions, along with other GHG emissions, volcanic activity, and solar forcing data, are treated as exogenous in this study, sourced from the SSP and RCP datasets obtained via the Reduced Complexity Model Intercomparison Project (RCMIP)⁷⁷. We primarily focus on CO₂ FFI because it represents the largest share of, and is the fastest-growing source of, gross anthropogenic greenhouse gas emissions⁵. FaIR calculates the global mean temperature increase, $\text{GMT}_{t,S}$, for time step t_i across all states in S by summing the regional CO₂ FFI emissions from the previous time step t_{i-1} alongside other exogenous data (Equation (8)).

$$\text{GMT}_{t,S} = \text{FaIR} \left(\sum_{n_j \in N} \text{EMIS}_{t_{i-1}, n_j, S} \right), \quad \forall t_i \in T, \quad \forall n_j \in N \quad (8)$$

Regional Temperature Downscaling

JUSTICE uses the climate downscaler from RICE50+ to derive the regional temperature response for the 57 regions. The downscaler uses a statistical method to analyze temperature data from 244 countries, incorporating data from the CMIP to connect global and local temperature changes using historical data. It uses global temperature data from all RCPs, which are implemented by multiple global climate models for future projections. Based on this dataset, a linear regression downscaler is calibrated with

region-specific coefficients TC1_{TNS} and TC2_{TNS} to estimate the regional temperature RMT_{TNS} in region n_j at time step t_i , using the mean global temperature rise GMT_{TNS} as shown in Equation (9). JUSTICE broadcasts global mean temperature rise to all regions in set N to obtain GMT_{TNS} , as described in Equation (10).

$$\text{RMT}_{TNS} = \text{TC1}_{TNS} + \text{TC2}_{TNS} \cdot \text{GMT}_{TNS} \quad (9)$$

$$\text{GMT}_{TNS} = \{gmt_{t_i, n_j, s_k}\}, \text{ where } gmt_{t_i, n_j, s_k} = gmt_{t_i, s_k} \quad \forall t_i, n_j, s_k \quad (10)$$

Damage Function

JUSTICE adopts the damage calculation from the RICE50+ model to estimate the damage fraction denoted by DAMFRAC_{TNS} , which represents the share of economic output lost due to climate damages (Equation (11)). This damage fraction is derived from the impact factor, IMPACT_{TNS} , which is computed from the impact at the previous time step t_{i-1} and the damage function specification $\text{DAMSPEC}_{t_i, NS}$ using an approximate impact formulation (Equation (12)). We use the damage specification from Kalkuhl and Wenz⁵⁸ (Equation (13)), which analyzes data from over 1,500 regions across 77 countries, utilizing subnational GDP data (Gross Regional Product, GRP) to empirically estimate historical climate impacts across different time scales and regions. Notably, their approach excludes non-market damages, such as ecosystem and health impacts, and does not account for extreme weather events or sea level rise.

Kalkuhl and Wenz developed several variants of the annual panel model to estimate damages, and our study employs the linearized impact variant. This variant captures linear impacts on production based on lagged temperature $\text{RMT}_{t_{i-1}}$ and includes weather change effects from the previous year's temperature through an interaction term, which is the product of the change in annual regional temperature ΔRMT_{t_i} and lagged temperature $\text{RMT}_{t_{i-1}}$ (Equation (16)). The temperature-dependent annual impact on productivity is calculated using the damage coefficients DC1_{t_i} and DC2_{t_i} , which are applied to the lagged temperature and the interaction term (Equation (14) and (15)).

$$\text{DAMFRAC}_{TNS} = 1 - \frac{1}{1 + \text{IMPACT}_{TNS}} \quad (11)$$

$$\text{IMPACT}_{t_i, NS} = \frac{1 + \text{IMPACT}_{t_{i-1}, NS}}{1 + \text{DAMSPEC}_{t_{i-1}, NS}} - 1 \quad \forall t_i \in T \quad (12)$$

$$\text{DAMSPEC}_{t_i, NS} = \text{DC1}_{t_i, NS} \cdot \Delta \text{RMT}_{t_i, NS} + \text{DC2}_{t_i, NS} \cdot \text{RMT}_{t_{i-1}, NS} \cdot \Delta \text{RMT}_{t_i, NS} \quad \forall t_i \in T \quad (13)$$

$$\text{DC1}_{TNS} = \{dc1_{t_i, n_j, s_k}\}, \text{ where } dc1_{t_i, n_j, s_k} = dc1 \quad \forall t_i, n_j, s_k \quad (14)$$

$$\text{DC2}_{TNS} = \{dc2_{t_i, n_j, s_k}\}, \text{ where } dc2_{t_i, n_j, s_k} = dc2 \quad \forall t_i, n_j, s_k \quad (15)$$

$$\Delta \text{RMT}_{t_i, NS} = \text{RMT}_{t_i, NS} - \text{RMT}_{t_{i-1}, NS} \quad \forall t_i \in T \quad (16)$$

Consumption

Consumption, denoted as CONSUMPTION_{TNS} , for a region at a specific time step and a state is determined by its net economic output $\text{Y}_{TNS}^{\text{NET}}$ and investment I_{TNS} (Equation (17)). Consumption per capita C_{TNS} is calculated by dividing the consumption hypermatrix by the population or labour L_{TNS}^* from the SSP database (Equation (18)). Net economic output is the adjusted economic output obtained by subtracting the damage costs DAMCOST_{TNS} , and the abatement costs ABATECOST_{TNS} associated with mitigating CO₂ emissions (Equation (19)). Damage cost is calculated by multiplying the gross economic output $\text{Y}_{TNS}^{\text{GROSS}}$ with the damage fraction as shown in Equation (20). Investment I_{TNS} represents the portion of gross

economic output Y_{TNS}^{GROSS} that is saved and reinvested in the economy to induce economic growth. The amount saved depends on the savings rate S_{TNS}^* , which varies by region at each time step and is multiplied by net economic output to determine the investment (Equation (21)). In this study, the savings rate is exogenous, derived from regional historical data, and converges to long-term projections from DICE-2016R2⁷⁵.

$$CONSUMPTION_{TNS} = Y_{TNS}^{NET} - I_{TNS} \quad (17)$$

$$C_{TNS} = \frac{CONSUMPTION_{TNS}}{L_{TNS}^*} \quad (18)$$

$$Y_{TNS}^{NET} = Y_{TNS}^{GROSS} - DAMAGECOST_{TNS} - ABATECOST_{TNS} \quad (19)$$

$$DAMAGECOST_{TNS} = Y_{TNS}^{GROSS} \cdot DAMFRAC_{TNS} \quad (20)$$

$$I_{TNS} = S_{TNS}^* \cdot Y_{TNS}^{NET} \quad (21)$$

Social Welfare Function

The Social Welfare Function (SWF) framework is a robust mechanism for ethically evaluating climate policies by considering their multi-dimensional and heterogeneous impacts. Primarily a normative framework, the SWF assesses and ranks policy options based on ethical considerations. In this study, we apply the SWF across three dimensions: time steps (T), regions (N), and SOWs (S), to evaluate the distribution of consumption per capita across these dimensions. Following the conventions established in previous IAM studies, we use consumption per capita, C_{TNS} , as a proxy for welfare, as defined in Equation (18). The welfare score, w , is computed from the C_{TNS} hypermatrix using Equation (22). The function $\mathcal{E}^M$ represents the equity equivalent of consumption per capita in dimension M ; it signifies the level of consumption per capita that, when uniformly distributed across dimension M (where $M \in D = \{S, N, T\}$), does not change the social welfare score.

$$w = \mathcal{E}^T(\mathcal{E}^N(\mathcal{E}^S(C_{TNS}))) \quad (22)$$

The order of aggregation is non-trivial, as shown mathematically by Berger and Emmerling⁷⁸. We employ an individual risk aversion aggregation method, where the aggregation occurs in three steps: first over the SOWs (S), then over regions (N), and finally over time (T). In this approach, the social planner evaluates welfare based on the consumption risks specific to each region. Individual risk aversion differs from the global risk aversion method, which assumes the existence of global risk-sharing technologies. In contrast, individual risk aversion considers the consumption loss risks intrinsic to each region and assigns greater welfare weight to individuals with lower consumption levels during adverse states. To aggregate over states, we need to specify a risk aversion parameter, φ_S , which we set to ($\varphi_S = 0$) in this study, indicating a risk-neutral social planner. The states-aggregated consumption per capita is calculated using Equation (23). The risk aversion is applied to consumption per capita using Equation (25) and aggregated using the aggregator function over dimension S , denoted as $\mathbb{E}^S$ (Equation (24)). This aggregator function applies a weight q_{S_k} , representing the probability of a particular SOW, to the consumption per capita in that state, C_{TNS_k} . The aggregated value is then inverted using the inverse of the utility function f^{-1} to retrieve the states-aggregated consumption per capita $\tilde{C}_{TN}$.

$$\tilde{C}_{TN} = \mathcal{E}^S(C_{TNS}) = f^{-1}[(\mathbb{E}^S[f(C_{TNS_k}; \varphi_S)]; \varphi_S] \quad (23)$$

$$\mathbb{E}^S[f(C_{TNS_k})] = \sum_{S_k \in S} q_{S_k} f(C_{TNS_k}) \quad (24)$$

$$f(C_{TNS_k}; \varphi_S) = \begin{cases} \frac{C_{TNS_k}^{1-\varphi_S}}{1-\varphi_S} & \text{if } \varphi_S \neq 1 \\ \ln(C_{TNS_k}) & \text{if } \varphi_S = 1 \end{cases} \quad \forall S_k \in S \quad (25)$$

Since the distribution of consumption across regions is heterogeneous, distributive justice principles come into play during the aggregation over regions. The region-aggregated consumption per capita C_T (Equation (26)) relies on the distributive justice principle employed and how it values the distribution of welfare across different regions at any given time step. The preference parameter for the utility function is called the inequality aversion, denoted by φ_N and is applied to state-aggregated consumption per capita for region n_j , C_{Tn_j} , as described in Equation (28). The φ_N parameter is also known as Atkinson's priority parameter because it prioritizes worse-off regions when ($\varphi_N > 0$). The weights q_{n_j} in Equation (27) represent the ratio of the region's population to the total global population at a specific time step. Consumption undergoes an additional transformation to account for the declining marginal utility of consumption. This transformation ensures that increases in consumption contribute less to welfare when regions are already well-off compared to worse-off regions. This transformation is also performed using the utility function, where the preference parameter becomes the elasticity of marginal utility of consumption denoted as λ , as shown in Equation (29). Aggregating the regional consumption per capita over the region dimension results in the global consumption per capita score over the model time steps C_T .

$$\tilde{C}_T = \mathcal{E}^N(\tilde{C}_{TN}) = f^{-1}[(\mathbb{E}^N[f(\tilde{C}_{Tn_j}; \varphi_N)]; \varphi_N] \quad (26)$$

$$\mathbb{E}^N[f(\tilde{C}_{Tn_j})] = \sum_{n_j \in N} q_{n_j} f(\tilde{C}_{Tn_j}) \quad (27)$$

$$f(\tilde{C}_{Tn_j}; \varphi_N) = \begin{cases} \frac{\tilde{C}_{Tn_j}^{1-\varphi_N}}{1-\varphi_N} & \text{if } \varphi_N \neq 1 \\ \ln(\tilde{C}_{Tn_j}) & \text{if } \varphi_N = 1 \end{cases} \quad \forall n_j \in N \quad (28)$$

$$U_T = g(\tilde{C}_T; \lambda) = \begin{cases} \frac{\tilde{C}_T^{1-\lambda}}{1-\lambda} & \text{if } \lambda \neq 1 \\ \ln(\tilde{C}_T) & \text{if } \lambda = 1 \end{cases} \quad (29)$$

A general SWF for different distributive justice principles can be derived (see Supplementary Document for derivation) by adding sufficiency threshold C_{TN}^{Thresh} for Sufficiency SWF and equality strictness parameter ψ along with a GINI measure as illustrated in Equation (30). We chose the c^{thresh} to be the World Bank stipulated poverty line of US\$1.25 adjusted for 2005 Purchasing power parity (PPP). The same c^{thresh} is broadcast for all regions and time steps (Equation (31)), but the same equation can be used for heterogeneous thresholds across regions and time steps. We use the GINI metric to calculate the inequality of the distribution of consumption per capita (Equation (32)) where η is the mean consumption per capita across the regions. The GINI is based on Concept-1 variation as described by Milanovic⁷⁹, which measures international inequality by comparing the distribution of consumption per capita between regions and measures whether nations' consumption per capita levels are converging.

$$U_T^{GENERAL} = g(f^{-1}[(\mathbb{E}^N[f(\tilde{C}_{Tn_j} - C_{TN}^{Thresh}); \varphi_N]); (1 - \psi \cdot GINI_T(\tilde{C}_{TN})); \varphi_N]; \lambda), \text{ where } 0 \leq \psi \leq 1 \quad (30)$$

$$C_{TN}^{Thresh} = \left\{ c_{t_i n_j}^{thresh} \right\}, \text{ where } c_{t_i n_j}^{thresh} = c^{thresh} \quad \forall t_i, n_j \quad (31)$$

$$GINI_T(\tilde{C}_{TN}) = GINI(\tilde{C}_{Tn_j}) = \frac{1}{\eta} \cdot \frac{1}{|N|^2} \sum_{a=1}^{|N|} \sum_{b>i}^{|N|} |\tilde{C}_{Tn_b} - \tilde{C}_{Tn_a}| \quad (32)$$

The time dimension is the final dimension to aggregate where the distribution of utility, $\mathbf{U}_T$ (consumption per capita adjusted with the declining marginal utility of consumption), is transformed using Equation (35). The function $h(\mathbf{U}_t; \varphi_T)$ transforms the utility scores with the inter-generational inequity aversion parameter, commonly known as the pure rate of time preference φ_T . The weight q_{t_i} applies discounting, where future utility scores get diminished based on the value of φ_T (Equation (33) and (34)). For Utilitarian SWF, the discount rate is 1.5% ($\varphi_T = 0.015$), and for Prioritarian SWF, ($\varphi_T = 0$) and hence no discounting is applied to satisfy the fundamental axioms of SWF (See Supplementary Document).

$$w = \mathcal{E}^T(\mathbf{U}_T) = \mathbb{E}^T[h(\mathbf{U}_t; \varphi_T)] \quad (33)$$

$$\mathbb{E}^T[h(\mathbf{U}_t; \varphi_T)] = \sum_{t_i \in T} q_{t_i} h(\mathbf{U}_t; \varphi_T) \quad (34)$$

$$h(\mathbf{U}_t; \varphi_T) = \mathbf{U}_t \cdot (1 + \varphi_T)^{-t}, \forall t_i \in T \quad (35)$$

Objectives

This study utilizes four objectives to optimize the emission control rate lever, $\mathbf{ECR}_{TNS}$. One of these objectives is welfare from consumption per capita, $\mathbf{C}_{TNS}$, calculated using various principles of distributive justice, as shown in Equation (22) in the previous section. The same equation is used to calculate welfare loss from damage costs, $wloss^{\text{damage}}$, and from abatement costs, $wloss^{\text{abatement}}$, as illustrated in Equations (36) and (37), respectively. Additionally, we introduce the objective of minimizing the number of years above the temperature threshold, yat , which tracks whether the annual global mean temperature, $gmt_{t_i s_k}$, for each year and state exceeds the 2°C threshold. The yat objective aggregates the count of years exceeding 2°C and averages this across all states, as demonstrated in Equation (38). The indicator function Γ assesses whether the global mean temperature, $gmt_{t_i s_k}$, exceeds the 2°C threshold for each pair of t_i and s_k ; it returns one if the condition is satisfied (indicating that the global mean temperature is above the threshold) and zero otherwise, as outlined in Equations (38) and (40). The output of the indicator function Γ is multiplied by the model time step $\Delta t = 1$ (yearly timestep; Equation (39)). This Δt multiplier is retained for generalizability across different timesteps (e.g., 5- or 10-year intervals). The four objectives under which the adaptive policy is optimized are described in Equations (22), (38), (36) and (37). An important consideration while applying the declining marginal utility of consumption in the social welfare function SWF is that it inverts the welfare and welfare loss scores, as the exponent in the utility function (Equation (29)) becomes negative. Hence, we minimize the absolute value of welfare and maximize the absolute value of welfare loss during the optimization.

$$wloss^{\text{damage}} = \mathcal{E}^T(\mathcal{E}^N(\mathcal{E}^S(\mathbf{DAMAGECOST}_{TNS}))) \quad (36)$$

$$wloss^{\text{abatement}} = \mathcal{E}^T(\mathcal{E}^N(\mathcal{E}^S(\mathbf{ABATECOST}_{TNS}))) \quad (37)$$

$$yat = \sum_{t_i \in T} \left[\frac{1}{|S|} \left(\sum_{s_k \in S} \Gamma(gmt_{t_i s_k} > t^{\text{thresh}}) \cdot \Delta t \right) \right], \forall t_i, s_k \quad (38)$$

$$\Delta t = t_i - t_{i-1} \quad (39)$$

$$\Gamma(a) = \begin{cases} 1 & \text{if } a \text{ is true} \\ 0 & \text{if } a \text{ is false} \end{cases} \quad (40)$$

Adaptive Policy

The emission control rate, $\mathbf{ECR}_{t_i NS}$, at time step t_i is represented as the output of a Radial Basis Function (RBF) network, as shown in Equation (41).

The RBF network takes two inputs from JUSTICE at each time step, denoted as $\mathbf{INPUT}_{t_i, Q, R, S}$ (Equation (42)), and processes them using a squared exponential activation function that incorporates centres $\mathbf{CENTER}_{q, r, S}$ and radii $\mathbf{RADII}_{q, r, S}$. The transformed input is then aggregated using weights $\mathbf{WEIGHT}_{q, NS}$ associated with the edges connecting the RBF nodes. This RBF network's topology comprises 4 hidden layer nodes ($|Q| = 4$) linked to the two inputs ($|R| = 2$) and produces 57 outputs ($|N| = 57$) corresponding to 57 regions. This specific network configuration results in 8 centres and 8 radii for the 4 RBF nodes, along with 228 weights (57 outputs connected to 4 nodes). The centres, radii, and weights serve as hyperparameters that are optimized by the Multi-objective Evolutionary Algorithm (MOEA). In our setup, the inputs consist of the temperature rise $\mathbf{GMT}_{t_i QS}$ and the rate of change of temperature over 5 years, $\Delta \mathbf{GMT}_{t_i QS}$ (Equation (44)). These inputs are normalized and used as feedback to adjust the emission control rate for the regions (Equation (42)). The input hypermatrix and global mean temperature are broadcast along the Q dimension to align with the number of hidden layer nodes, as shown in Equations (43) and (45). JUSTICE employs the Evolutionary Multi-objective Direct Policy Search (EMODPS) algorithm to facilitate multi-objective optimization and adaptive policy design⁵². EMODPS has been successfully implemented in various environmental system case studies and is particularly effective in multi-objective adaptive policymaking. The EMODPS algorithm combines direct policy search with non-linear approximating networks (such as RBFs) and MOEAs to develop adaptive Pareto optimal policies.

$$\mathbf{ECR}_{t_i NS} = \sum_{q_l=1}^{|Q|} \left[\mathbf{WEIGHT}_{q_l NS} \cdot \exp \left(- \sum_{r_m=1}^{|R|} \frac{(\mathbf{INPUT}_{t_i q_l r_m S} - \mathbf{CENTER}_{q_l r_m S})^2}{(\mathbf{RADII}_{q_l r_m S})^2} \right) \right], \quad \forall t_i \in T, \forall q_l \in Q, \forall r_m \in R, |R| = 2, |Q| = 4 \quad (41)$$

$$\mathbf{INPUT}_{t_i, Q, R, S} = [\mathbf{INPUT}_{t_i Q r_1 S}, \mathbf{INPUT}_{t_i Q r_2 S}], \forall r_m \in R, \quad (42)$$

where $\mathbf{INPUT}_{t_i Q r_1 S} = \mathbf{GMT}_{t_i QS}$ and $\mathbf{INPUT}_{t_i Q r_2 S} = \Delta \mathbf{GMT}_{t_i QS}$

$$\mathbf{INPUT}_{t_i, Q, R, S} = \{\text{input}_{t_i, q_l, r_m, s_k}\}, \text{ where } \text{input}_{t_i, q_l, r_m, s_k} = \text{input}_{t_i, r_m, s_k} \forall t_i, q_l, r_m, s_k \quad (43)$$

$$\Delta \mathbf{GMT}_{t_i QS} = \mathbf{GMT}_{t_i QS} - \mathbf{GMT}_{t_{i-5} QS}, \forall t_i \in T, t_i \geq 5 \quad (44)$$

$$\mathbf{GMT}_{t_i QS} = \{gmt_{t_i, q_l, s_k}\}, \text{ where } gmt_{t_i, q_l, s_k} = gmt_{t_i, s_k} \forall t_i, q_l, s_k \quad (45)$$

We impose a maximum annual increase of 4% on the emission control rate, denoted by $abaterate^{\text{max}}$, to reflect technological inertia and practical transition constraints, consistent with findings from previous studies^{30,55,80}. The abatement cost calculation function does not factor in the increased costs associated with the speed of mitigation, and hence, we adopt these constraints that prevent unrealistic mitigation jumps. Additionally, we exclude the possibility of negative emissions before 2100 and set a cap on the maximum emission control rate at $ecr^{\text{max}} = 100\%$. Therefore, the value of $\mathbf{ECR}_{t_i NS}$ can increase from a minimum of $ecr^{\text{min}} = 0\%$. At any given time step t_i , the emissions control rate can either increase or remain unchanged, but it cannot decrease, as outlined in Equation (46) and (47).

$$\mathbf{ECR}_{t_i NS} = \begin{cases} \max(ecr^{\text{min}}, \min(\mathbf{ECR}_{t_{i-1} NS} + \min(\Delta \mathbf{ECR}_{t_i NS}, abaterate^{\text{max}}), ecr^{\text{max}})) & \text{if } \Delta \mathbf{ECR}_{t_i NS} \geq 0 \\ \mathbf{ECR}_{t_{i-1} NS} & \text{if } \Delta \mathbf{ECR}_{t_i NS} < 0 \end{cases} \quad (46)$$

$$\Delta \mathbf{ECR}_{t_i NS} = \mathbf{ECR}_{t_i NS} - \mathbf{ECR}_{t_{i-1} NS} \quad (47)$$

Data availability

Data for this paper are available via Zenodo at <https://doi.org/10.5281/zenodo.15710403>. The code for JUSTICE model is made open-source and made available via GitHub at <https://github.com/pollockDeVis/JUSTICE>.

The code and instructions to replicate the data and figure for this paper is available via Zenodo at <https://doi.org/10.5281/zenodo.15710403>.

Code availability

The code for JUSTICE model is made open-source and made available via GitHub at <https://github.com/pollockDeVis/JUSTICE>. The code and instructions to replicate the data and figure for this paper is available via Zenodo at <https://doi.org/10.5281/zenodo.15710403>.

Received: 21 June 2025; Accepted: 8 May 2026;

Published online: 20 May 2026

References

- Lenton, T. M. et al. Climate tipping points—too risky to bet against. *Nature* **575**, 592–595 (2019).
- Burke, M., Davis, W. M. & Diffenbaugh, N. S. Large potential reduction in economic damages under un mitigation targets. *Nature* **557**, 549–553 (2018).
- van der Wijst, K.-I., Hof, A. F. & van Vuuren, D. P. On the optimality of 2°C targets and a decomposition of uncertainty. *Nat. Commun.* **12**, 2575 (2021).
- Dietz, S., Rising, J., Stoerk, T. & Wagner, G. Economic impacts of tipping points in the climate system. *Proc. Natl. Acad. Sci.* **118**, e2103081118 (2021).
- Lee, H. et al. Ipcc, 2023: Climate change 2023: Synthesis report, summary for policymakers. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change (2023).
- Moyer, J. D. et al. How many people will live in poverty because of climate change? a macro-level projection analysis to 2070. *Climatic Change* **176**, 137 (2023).
- Robiou du Pont, Y. et al. Equitable mitigation to achieve the paris agreement goals. *Nat. Clim. Change* **7**, 38–43 (2017).
- Hickel, J. & Slamersak, A. Existing climate mitigation scenarios perpetuate colonial inequalities. *Lancet Planet. Health* **6**, e628–e631 (2022).
- Rivadeneira, N. R. & Carton, W. (in) justice in modelled climate futures: A review of integrated assessment modelling critiques through a justice lens. *Energy Res. Soc. Sci.* **92**, 102781 (2022).
- Kanitkar, T., Mythri, A. & Jayaraman, T. Equity assessment of global mitigation pathways in the ipcc sixth assessment report. *Climate Policy* 1–20 (2024).
- Cointe, B., Cassen, C. & Nadaï, A. Organising policy-relevant knowledge for climate action: integrated assessment modelling, the ipcc, and the emergence of a collective expertise on socioeconomic emission scenarios. *Science & Technology Studies* (2019).
- Keppo, I. et al. Exploring the possibility space: Taking stock of the diverse capabilities and gaps in integrated assessment models. *Environ. Res. Lett.* **16**, 053006 (2021).
- Shukla, P. et al. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022).
- Weyant, J. Some contributions of integrated assessment models of global climate change. *Review of Environmental Economics and Policy* (2017).
- Stanton, E. A., Ackerman, F. & Kartha, S. Inside the integrated assessment models: Four issues in climate economics. *Clim. Dev.* **1**, 166–184 (2009).
- Adler, M. D. *Measuring social welfare: An introduction* (Oxford University Press, USA, 2019).
- Hansson, S. O. Philosophical problems in cost-benefit analysis. *Econ. Philos.* **23**, 163–183 (2007).
- Dennig, F. Climate change and the re-evaluation of cost-benefit analysis. *Climatic change* **151**, 43–54 (2018).
- Franssen, M. Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design. *Res. Eng. Des.* **16**, 42–56 (2005).
- Köberle, A. C. et al. The cost of mitigation revisited. *Nat. Clim. Change* **11**, 1035–1045 (2021).
- Rogelj, J., McCollum, D. L., Reisinger, A., Meinshausen, M. & Riahi, K. Probabilistic cost estimates for climate change mitigation. *Nature* **493**, 79–83 (2013).
- Kulkarni, S., Hof, A., van der Wijst, K.-I. & van Vuuren, D. Disutility of climate change damages may warrant much stricter climate targets. *Environ. Res. Commun.* **6**, 021001 (2024).
- Emmerling, J., Kornek, U. & Zuber, S. Multidimensional welfare indices and the ipcc 6th assessment report scenarios. *Ecol. Econ.* **220**, 108182 (2024).
- Tversky, A. & Kahneman, D. Loss aversion in riskless choice: A reference-dependent model. *Q. J. Econ.* **106**, 1039–1061 (1991).
- Tversky, A. & Kahneman, D. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).
- Kahneman, D. & Tversky, A. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, 99–127 (World Scientific, 2013).
- De Neve, J.-E. et al. The asymmetric experience of positive and negative economic growth: Global evidence using subjective well-being data. *Rev. Econ. Stat.* **100**, 362–375 (2018).
- Hobbs, B. F., Chankong, V., Hamadeh, W. & Stakhiv, E. Z. Does choice of multicriteria method matter? an experiment in water resources planning. *Water Resour. Res.* **28**, 1767–1779 (1992).
- Garner, G., Reed, P. & Keller, K. Climate risk management requires explicit representation of societal trade-offs. *Climatic Change* **134**, 713–723 (2016).
- Marangoni, G., Lamontagne, J. R., Quinn, J. D., Reed, P. M. & Keller, K. Adaptive mitigation strategies hedge against extreme climate futures. *Climatic Change* **166**, 37 (2021).
- Ferrari, L., Carlino, A., Gazzotti, P., Tavoni, M. & Castelletti, A. From optimal to robust climate strategies: expanding integrated assessment model ensembles to manage economic, social, and environmental objectives. *Environ. Res. Lett.* **17**, 084029 (2022).
- Budolfson, M. B. et al. Utilitarian benchmarks for emissions and pledges promote equity, climate and development. *Nat. Clim. change* **11**, 827–833 (2021).
- Adler, M. et al. Priority for the worse-off and the social cost of carbon. *Nat. Clim. Change* **7**, 443–449 (2017).
- Drupp, M. A., Freeman, M. C., Groom, B. & Nesje, F. Discounting disentangled. *Am. Economic J.: Economic Policy* **10**, 109–134 (2018).
- Heal, G. M. & Millner, A. Agreeing to disagree on climate policy. *Proc. Natl. Acad. Sci.* **111**, 3695–3698 (2014).
- Underdal, A. & Wei, T. Distributive fairness: A mutual recognition approach. *Environ. Sci. policy* **51**, 35–44 (2015).
- Pielke Jr, R. A. *The honest broker: Making sense of science in policy and politics* (Cambridge University Press, 2007).
- Reich, P. B., Grace, K., Agrawal, A. & Nagendra, H. Mitigation justice. *Proc. Natl. Acad. Sci.* **122**, e2411231122 (2025).
- Tavoni, M. & Valente, G. Uncertainty in integrated assessment modeling of climate change. *Perspect. Sci.* **30**, 321–351 (2022).
- Zimm, C. et al. Justice considerations in climate research. *Nat. Clim. Change* **14**, 22–30 (2024).
- Weitzman, M. Revisiting fat-tailed uncertainty in the economics of climate change. *Review of Environmental Economics and Policy* (2011).
- Pindyck, R. S. The use and misuse of models for climate policy. *Review of Environmental Economics and Policy* (2017).
- Rogna, M. & Vogt, C. J. Optimal climate policies under fairness preferences. *Climatic Change* **174**, 25 (2022).

44. Koomey, J., Hausker, K., Schmidt, Z. & Lashof, D. Abandon the idea of an “optimal economic path” for climate policy. *Wiley Interdiscip. Rev.: Clim. Change* **14**, e850 (2023).
45. Knutti, R., Rugenstein, M. A. & Hegerl, G. C. Beyond equilibrium climate sensitivity. *Nat. Geosci.* **10**, 727–736 (2017).
46. Nordhaus, W. D. Revisiting the social cost of carbon. *Proc. Natl. Acad. Sci.* **114**, 1518–1523 (2017).
47. Clarke, L. et al. Assessing transformation pathways (2014).
48. Taebi, B., Kwakkel, J. H. & Kermisch, C. Governing climate risks in the face of normative uncertainties. *Wiley Interdiscip. Rev.: Clim. Change* **11**, e666 (2020).
49. Pindyck, R. S. Climate change policy: what do the models tell us? *J. Economic Lit.* **51**, 860–872 (2013).
50. Jensen, S. & Traeger, C. Uncertainty in climate-economic modeling. In *Handbook of the Economics of Climate Change*, vol. 1, 351–423 (Elsevier, 2024).
51. Carlino, A., Tavoni, M. & Castelletti, A. Self-adaptive multi-objective climate policies align mitigation and adaptation strategies. *Earth's Future* **10**, e2022EF002767 (2022).
52. Giuliani, M., Castelletti, A., Pianosi, F., Mason, E. & Reed, P. M. Curses, tradeoffs, and scalable management: Advancing evolutionary multiobjective direct policy search to improve water reservoir operations. *J. Water Resour. Plan. Manag.* **142**, 04015050 (2016).
53. Lamontagne, J., Reed, P., Marangoni, G., Keller, K. & Garner, G. Robust abatement pathways to tolerable climate futures require immediate global action. *Nat. Clim. Change* **9**, 290–294 (2019).
54. Terhaar, J., Frölicher, T. L., Aschwanden, M. T., Friedlingstein, P. & Joos, F. Adaptive emission reduction approach to reach any global warming target. *Nat. Clim. change* **12**, 1136–1142 (2022).
55. Gazzotti, P. et al. Persistent inequality in economically optimal climate policies. *Nat. Commun.* **12**, 3421 (2021).
56. Nordhaus, W. D. & Yang, Z. A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review* **74**, 741–765 (1996).
57. Smith, C. J., Al Khourdajie, A., Yang, P. & Folini, D. Climate uncertainty impacts on optimal mitigation pathways and social cost of carbon. *Environ. Res. Lett.* **18**, 094024 (2023).
58. Kalkuhl, M. & Wenz, L. The impact of climate conditions on economic production: evidence from a global panel of regions. *J. Environ. Econ. Manag.* **103**, 102360 (2020).
59. Riahi, K. et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Glob. Environ. change* **42**, 153–168 (2017).
60. Leach, N. J. et al. Fairv2. 0.0: a generalized impulse response model for climate uncertainty and future scenario exploration. *Geoscientific Model Dev.* **14**, 3007–3036 (2021).
61. Meinshausen, M. et al. The rcp greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic change* **109**, 213–241 (2011).
62. Smith, C., Byers, E., Werning, M. & Hooke, D. R10 region mask based on ipcc ar6 wg3 and ismip (2023).
63. Lempert, R. J., Schlesinger, M. E. & Bankes, S. C. When we don't know the costs or the benefits: Adaptive strategies for abating climate change. *Climatic Change* **33**, 235–274 (1996).
64. Grubb, M., Wieners, C. & Yang, P. Modeling myths: On dice and dynamic realism in integrated assessment models of climate change mitigation. *Wiley Interdiscip. Rev.: Clim. Change* **12**, e698 (2021).
65. Kasprzyk, J. R., Reed, P. M. & Hadka, D. M. Battling arrow's paradox to discover robust water management alternatives. *J. Water Resour. Plan. Manag.* **142**, 04015053 (2016).
66. Okereke, C. & Coventry, P. Climate justice and the international regime: before, during, and after paris. *Wiley Interdiscip. Rev.: Clim. Change* **7**, 834–851 (2016).
67. Sterner, T. et al. Policy design for the anthropocene. *Nat. Sustainability* **2**, 14–21 (2019).
68. Kwakkel, J. H., Walker, W. E. & Haasnoot, M. Coping with the wickedness of public policy problems: Approaches for decision making under deep uncertainty (2016).
69. Salin, M., Kedward, K. & Dunz, N. Price assessing integrated assessment models for building global nature-economy scenarios. Banque de France Working papers (2024).
70. Fuss, S. et al. Betting on negative emissions. *Nat. Clim. change* **4**, 850–853 (2014).
71. Mintz-Woo, K. The need-efficiency tradeoff for negative emissions technologies. *PLoS Clim.* **1**, e0000060 (2022).
72. Shue, H. Climate dreaming: negative emissions, risk transfer, and irreversibility. *J. Hum. Rights Environ.* **8**, 203–216 (2017).
73. Dennig, F., Budolfson, M. B., Fleurbaey, M., Siebert, A. & Socolow, R. H. Inequality, climate impacts on the future poor, and carbon prices. *Proc. Natl. Acad. Sci.* **112**, 15827–15832 (2015).
74. Van Vuuren, D. P. et al. The representative concentration pathways: an overview. *Climatic change* **109**, 5–31 (2011).
75. Nordhaus, W. Projections and uncertainties about climate change in an era of minimal climate policies. *Am. economic J.: economic policy* **10**, 333–360 (2018).
76. Després, J. et al. Poles-jrc model documentation. JRC Europa (2018).
77. Nicholls, Z. R. et al. Reduced complexity model intercomparison project phase 1: introduction and evaluation of global-mean temperature response. *Geoscientific Model Dev.* **13**, 5175–5190 (2020).
78. Berger, L. & Emmerling, J. Welfare as equity equivalents. *J. Economic Surv.* **34**, 727–752 (2020).
79. Milanovic, B. *Worlds apart: Measuring international and global inequality* (Princeton University Press, 2011).
80. Hänsel, M. C. et al. Climate economics support for the un climate targets. *Nat. Clim. Change* **10**, 781–789 (2020).

Acknowledgements

This work was conducted within the HIPPO Lab and supported by the TU Delft AI Labs & Talent Programme. We thank Prof. Alexander Verbraeck for his feedback on improving the mathematical notation.

Author contributions

P.B., J.Z.S., and J.H.K. conceived the study. P.B. developed the model and the general social welfare function, ran experiments and generated the results. P.B. drafted the article with contributions from J.Z.S., J.H.K., and G.M. All authors analysed the results and reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44168-026-00385-z>.

Correspondence and requests for materials should be addressed to Palok Biswas.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026