

Marco Rogna
Carla J. Vogt

Accounting for Inequality Aversion Can Justify the 2° C Goal



#925

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer

RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger

Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen

University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Ronald Bachmann, Prof. Dr. Manuel Frondel, Prof. Dr. Torsten Schmidt,
Prof. Dr. Ansgar Wübker

RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler

RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #925

Responsible Editor: Manuel Frondel

All rights reserved. Essen, Germany, 2021

ISSN 1864-4872 (online) – ISBN 978-3-96973-082-9

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #925

Marco Rogna and Carla J. Vogt

**Accounting for Inequality Aversion
Can Justify the 2° C Goal**



Bibliografische Informationen der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

<http://dx.doi.org/10.4419/96973082>

ISSN 1864-4872 (online)

ISBN 978-3-96973-082-9

Marco Rogna and Carla J. Vogt¹

Accounting for Inequality Aversion Can Justify the 2° C Goal

Abstract

Impact assessment models are a tool largely used to investigate the benefit of reducing polluting emissions and limiting the anthropogenic mean temperature rise. However, they have been often criticised for suggesting low levels of abatement. Countries and regions, that are generally the actors in these models, are usually depicted as having standard concave utility functions in consumption. This, however, disregards a potentially important aspect of environmental negotiations, namely its distributive implications. The present paper tries to fill this gap assuming that countries\regions have Fehr and Schmidt (1999) (F&S) utility functions, specifically tailored for including inequality aversion. Thereby, we propose a new method for the empirical estimation of the inequality aversion parameters by establishing a link between the well known concept of elasticity of marginal utility of consumption and the F&S utility functions, accounting for heterogeneity of countries/regions. By adopting the RICE model, we compare its standard results with the ones obtained introducing F&S utility functions, showing that, under optimal cooperation, the level of temperature rise is significantly lower in the last scenario. In particular, in the last year of the simulation, the optimal temperature rise is 2.1° C. Furthermore, it is shown that stable coalitions are easier to be achieved when F&S preferences are assumed, even if the advantageous inequality aversion parameter (altruism) is assumed to have a very low value. However, self-sustaining coalitions are far from reaching the environmental target of limiting the mean temperature rise below 2° C despite the adoption of F&S utility functions.

JEL-Code: C72, D63, Q54

Keywords: Abatement; climate policy; inequality aversion; Paris agreement; RICE model

October 2021

¹ Marco Rogna, Hochschule Bochum; Carla J. Vogt, Hochschule Bochum and RWI Research Network. - All correspondence to: Carla J. Vogt, Bochum University of Applied Sciences, Am Hochschulcampus 1, 44801 Bochum, Germany, e-mail: carla.vogt@hs-bochum.de

1 Introduction

In the current debate about climate change, there are two numbers that have become very prominent: 1.5 and 2. These are the two thresholds, expressed in degrees Celsius, representing the increase in mean temperature above pre-industrial era, that should not be passed. The former threshold is more environmentally conservative and generally preferred by ecologists (Knutti et al., 2016), whereas the latter is sometimes judged as more realistic given the prompt and firm actions, and the related costs, required to meet the 1.5° target (Jewell and Cherp, 2020). Despite the debate, sometimes very fierce, on which of the two should be the target threshold over which to shape environmental policies, it is now almost given for granted that no other target levels of mean temperature increase will be taken into serious consideration (Knopf et al., 2012).

Therefore, it is not surprising that a large portion of the economic investigation related to climate change is nowadays dedicated to analyse the possibility to reach the mentioned thresholds, with a particular emphasis on the 1.5° C limit (Grubler et al., 2018). However, before the widespread convergence on these two numbers, there has been a long debate on which was the optimal limit below which to contain the mean global temperature rise. Economic-environmental models, either in the form of impact assessment models (IAMs) or computable general equilibrium models (CGEs) have been largely used to investigate the topic. Among the formers, the DICE\RICE¹ model of Nordhaus (1992) has played a central role having been one of the first to include detailed equations to depict the relations between economic activities, emissions, temperature rise and climate damages.

According to this model, even when considering the fully cooperative scenario, the derived optimal mean temperature increase for the end of the 21st century is well above 2.5° C (Nordhaus and Yang, 1996). In the subsequent revisions of the model, from one side, the inclusion of more serious environmental damages induced a decrease of optimal mean temperature rise, but on the other, a more sensitive relation between emissions and climate

¹The DICE model (Dynamic Integrated model of Climate and the Economy), firstly presented in Nordhaus (1992), considers the world as a whole, whereas the RICE model (Regional Integrated model of Climate and the Economy), firstly presented in Nordhaus and Yang (1996), decomposes it into countries\regions.

change pushed in the opposite direction, obtaining a substantial parity. In the 2010 version of RICE, in fact, the optimal mean temperature rise under full cooperation is slightly below 3° C (Nordhaus, 2010).

Several reasons have been adduced for this substantial difference between the DICE\RICE derived prescriptions and the one provided by environmentalists and climatologists. Many critiques have been directed to the damage function that, even after the updates, has been judged to underestimate catastrophic and immaterial damages (Weitzman, 2012; Howard and Sterner, 2017) and not to account for uncertainty (Roughgarden and Schneider, 1999; Diaz and Moore, 2017). Another crucial aspect that plays a central role in determining optimal emissions is the inter-temporal distribution. A high value of the discount factor, as many judge the one adopted in DICE\RICE (Dietz et al., 2018), naturally leads to a low evaluation of future climate damages, thus calling for lower levels of abatement. Alternative configurations of inter-temporal distribution in the DICE model have proven to be determinant in sensibly decreasing the derived optimal level of temperature rise: e.g. Botzen et al. (2018); Hänsel and Quaas (2018).

With climate damages and inter-temporal discounting having been the two most controversial aspects over which the academic debate has flourished, there is another important theme that has been partially shadowed by the mentioned debate: equity. Besides inter-generational equity represented by discounting, the sharp difference in economic attainment and emissions level of the 12 (originally 6) countries\regions in the RICE model calls for a thorough consideration of this aspect. The RICE model attempts to consider this aspect by including the elasticity of marginal utility (EMU) of consumption into the countries\regions utility function. Actors with larger levels of per-capita consumption will enjoy lower levels of utility increase for additional units of consumption. However, we argue that this standard formulation of a concave utility function in per-capita consumption may not be adequate to fully capture the disutility caused by inequality.

In particular, in the present paper we propose a more systematic inclusion of equity concerns based on insights from behavioural economics in the form of Fehr and Schmidt

(1999) (F&S) utility functions. In fact, these capture the phenomenon that people compare themselves to others and possibly derive dis-utility if their payoff is below or above other players' payoffs. This utility function is in line with numerous observations made in experimental economics and it has proven to be successful in explaining observed behaviours in bargaining and cooperation games (see Fehr and Schmidt (2006) for a review).

F&S utility functions have already been applied to the problem of voluntary agreements in international climate policy by Lange and Vogt (2003), Vogt (2016) and Rogna and Vogt (2020). However, the focus in these papers has primarily been to analyse the effect of inequality aversion on the prospects of voluntary cooperation via the means of coalitions. Furthermore, they all share the use of very simplified and atemporal models, typical of a game-theoretic analysis, that are not suitable to derive predictions about environmental and economic outcomes. The present paper, instead, directly includes F&S utility functions into the RICE model (version 2013 – v2013), assuming that countries\regions have such preferences, and compares the results to the ones obtained with the original model.

The inequality aversion parameters present in the F&S utility function are empirically derived by linking them to the EMU values as used in the RICE model. In particular, we decompose the total utility change that is captured by EMU in three components: the change in absolute payoff, the change in advantageous and the change in disadvantageous inequality, respectively. In this way, for a given value of EMU, the relative degrees of inequality aversion across income levels are determined.

Our results show that, even keeping the original inter-temporal discounting and climate damages as in RICE v2013, the adoption of F&S utility functions sensibly reduces the global level of pollution by the end of the 21st century in the fully cooperative case. The mean temperature increase, in fact, is approximately 2.10° C in 2100 under F&S preferences compared to 2.73° C in the standard model run. We further show that cooperation, in the form of stable coalitions, is significantly enhanced by the adoption of F&S utility functions. However, self-sustaining coalitions, even when F&S preferences are assumed, cannot grant a mean temperature increase less than 3.05° C.

2 Literature review

Impact assessment models have been one of the main instruments to investigate the costs and benefits of taking actions to counteract human-induced climate warming. The large use of these modeling tools in the Intergovernmental Panel on Climate Change (IPCC) reports well testifies this fact (Rosen, 2015; Hansson et al., 2021). Several models, with different underlying assumptions, focus and databases, have been proposed, among which MIRAGE (Easter et al., 2004), WITCH (Bosetti et al., 2006), MAgPIE (Dietrich et al., 2019) and POLES (Keramidas et al., 2017) are prominent examples. As mentioned earlier, the DICE\RICE model plays a central role in this list being among the first attempts to link the whole world economy to the earth's climate system and depicting the influence that each of the two has on the other.

From its original formulation in 1992 (Nordhaus, 1992), the DICE model has been subject to several major revisions along time (Nordhaus, 2018b). The decomposition of the world into 6 countries\regions in 1996 with the introduction of the RICE model (Nordhaus and Yang, 1996) has been one of the major changes. To this, several updates have followed in 2000 (Nordhaus and Boyer, 2000), in 2007 (Nordhaus, 2007), in 2010 (Nordhaus, 2010), in 2013 (Nordhaus, 2013) and in 2016 (Nordhaus, 2018a), increasing the number of countries \regions considered in RICE from 6 to 12, refining the economic and damage equations and changing the underlying database.

While the climatic and environmental side of the model has remained rather untouched along time, its economic side has seen major revisions (Nordhaus, 2018b). By comparing the estimates of the original DICE model (1992) for the year 2015, Nordhaus (2018b) notes how predictions have been substantially biased. In particular, output and damages have been strongly underestimated, thus leading to a value for the social cost of carbon (SCC) downward biased. The difference in the estimated SCC for 2100 between the original and the last version of the DICE model is more than six fold: 5\$ versus 31\$ per tonne of CO₂ (Nordhaus, 2017).

Despite the model updates in its economic side, the optimal emissions of CO₂, and the

associated temperature rise, has remained far above the targets of $1.5^{\circ}\text{--}2^{\circ}\text{ C}$ settled in the Paris Agreement. Note that this is not a peculiarity of DICE\RICE, since several other models provide an estimate of the SCC that is inadequate to meet the $1.5^{\circ}\text{--}2^{\circ}\text{ C}$ targets, as shown in Tol (2019) and in Ackerman and Munitz (2016). This has generated a wave of critiques directed towards IAMs. As mentioned in the introduction, the underestimation of catastrophic and immaterial damages (Weitzman, 2012; Howard and Sterner, 2017), the lack of account for uncertainty (Roughgarden and Schneider, 1999; Diaz and Moore, 2017) and a high inter-temporal discounting (Dietz et al., 2018) are the main targets of the mentioned critiques.

Several attempts have been made to overcome these perceived shortcomings. De Bruin et al. (2009) separate the mitigation and the adaptation costs in the DICE model, while Michaelis and Wirths (2020) consider the rate of temperature rise in addition to its level of increase, showing that ignoring the former aspect may substantially underestimate climate damages. Still adopting DICE, Tol (1994) proposes a different method for incorporating intangible damages, Botzen and van den Bergh (2012) assume an alternative specification of the damage function and Ackerman et al. (2010) attempt to model catastrophic damages and their distribution. Finally, Dietz and Asheim (2012) and Botzen et al. (2018) investigate different forms of inter-temporal discounting. Generally, these modifications lead to a lower level of temperature rise under optimality and earlier and tighter efforts for decarbonizing the economy.

With its disaggregation into 12 countries\regions, the RICE model has also been used to investigate the stability of international environmental agreements (IEAs), depicted as coalitions. The results, in line with other numerical models such as WITCH (Bosetti et al., 2006) and STACO (Dellink, 2011), have confirmed the grim predictions of early game-theoretic analyses – e.g. Carraro and Siniscalco (1993) and Barrett (1994) –, namely that only few and relatively small coalitions are stable (Yang et al., 2008).

On the game-theoretic side of climate change analysis, there is a number of papers that have tested the possibility of going beyond the standard assumption of pure self-interest,

embracing insights derived from behavioural and experimental economics. Lange and Vogt (2003) assume preferences á la Bolton and Ockenfels (2000), van der Pol et al. (2012) add a component of pure altruism while Vogt (2016) and Rogna and Vogt (2020) consider a utility function based on F&S preferences. A common finding of these papers is that, despite a general increase in stability once abandoning standard preferences, cooperation is not dramatically enhanced without transfers. Being game-theoretic papers, however, they all portray a very stylized and scarcely realistic representation of both the economic and the environmental side.

The present paper aims at including non-standard preferences into a dynamic and more complex model framework such as the RICE model, thus filling a current gap of the literature. The choice is for F&S preferences whose consideration of aversion for both advantageous (altruism) and disadvantageous (envy) inequality has proven to be able to capture several deviations from standard economic theory observed in laboratory experiments (Fehr and Schmidt, 2006). In particular, we are interested in observing which is the effect of this alternative specification of the utility function both on the optimal level of emissions abatement, and, therefore, on the temperature rise, and on the stability of climate coalitions.

3 The RICE model with F&S preferences

Our starting model, also used as benchmark, is RICE v2013. A synthetic description of all its variables (endogenous and exogenous) and parameters can be found in the Appendix, section A1. Its basic equations, instead, can be found in section A2, in the Appendix. Compared to the original RICE v2013, two modifications have been introduced. The first is to have reduced the number of control variables to one, namely the level of proportional abatement ($\mu_{i,t}$ in the model), whereas the original model has the saving–investment rate as an additional control variable. In particular, we treat the saving–investment rate – $\sigma_{i,t}^I$ in the model, equation (A4) – as exogenous, deriving its value from running the original model in the non-cooperative scenario. The reason for this choice is to simplify the model given that the introduction of F&S preferences adds a considerable computational burden. Fur-

thermore, leaving a single control variable, abatement, simplifies and renders more explicit the interpretation of results.

The second modification refers to the discounted utility function of countries – equation (A2) in the Appendix – whose numerator, differently from the original version, is not multiplied by the population size ($L_{i,t}$). This term will then act as a weight when summing utilities in coalitions, with more populous countries\regions gaining more importance. It may actually be reasonable and realistic to have such term since, in a bargaining process, larger countries could effectively hold more bargaining power. However, the F&S preferences only consider per-capita consumption when operating the inter-countries comparison. In order to keep as close as possible the two types of utility functions that will be compared, it seems then opportune to drop the population weight. Furthermore, this drop can be theoretically justified by the fact that countries\regions, being sovereign entities, act as individuals and the “power” granted by a larger population size is hardly quantifiable, if justifiable at all.

Except for the two modifications just explained, the set of equations in section A2 in the Appendix faithfully reproduces the original RICE model v2013. This will be used as our benchmark scenario, without adding any exogenous environmental target or any price for CO₂. As mentioned in the introduction, our main assumption is that, in order to properly capture the relational component of utility arising from comparing the own level of economic attainment (per-capita consumption) with the one of the others, the marginal utility of consumption – η_i in equation (A2) – is not sufficient. The F&S utility function, instead, is better suited for this purpose. Following is the mathematical definition of the F&S utility function:

$$U_i = \pi_i - \frac{\alpha_i}{n-1} \sum_{j \in I^+} (\pi_j - \pi_i) - \frac{\beta_i}{n-1} \sum_{k \in I^-} (\pi_i - \pi_k), \quad (1)$$

where i is a generic player of set N , whose cardinality is represented by n , π is the payoff of a player, I^+ and I^- are the sets of players having, respectively, a payoff higher and

lower than player i and, finally, α_i and β_i are the parameters representing the aversion for disadvantageous, the former, and for advantageous, the latter, inequality. The expression $\frac{\alpha_i}{n-1} \sum_{j \in I^+} (\pi_j - \pi_i)$, where the component inside the round brackets is always positive since $\pi_j > \pi_i$ by definition, represents the disutility suffered by player i for having a payoff lower than all players j (envy). Similarly, the expression following β_i , necessarily positive by definition as well, represents the disutility for advantageous inequality (altruism).

Willing to adopt the F&S utility function in the RICE v2013 model, equation (A2) in the Appendix must be substituted by the following two equations:

$$\begin{aligned} \pi_{i,t} &= \frac{C_{i,t}}{L_{i,t}} (1 + \rho_i)^{-ts \times t}, & \forall i \in N, \forall t \in T, \\ U_{i,t} &= \pi_{i,t} - \frac{\alpha_i}{n-1} \sum_{j \in I^+} (\pi_{j,t} - \pi_{i,t}) - \frac{\beta_i}{n-1} \sum_{k \in I^-} (\pi_{i,t} - \pi_{k,t}), & \forall i \in N, \forall t \in T, \end{aligned} \quad (2)$$

where $\pi_{i,t}$ is simply defined as the discounted value of per-capita consumption. While all the parameters of the model can be retrieved from the documentation of RICE, the addition of the new utility function brings the burden of estimating α and β . The next sub-section is dedicated to describe the procedure adopted for retrieving them.

The estimation of α and β

Since the values of α and β represent the intensity with which the disutility from disadvantageous and advantageous inequality is felt, they are of crucial importance in the present paper. Several works have tried to estimate them. In particular, the original work of Fehr and Schmidt (1999) provides some estimates of these two parameters, retrieved by a sort of backward induction, as to say by finding that values of α and β capable of explaining the deviations from standard theory reported in experimental economics papers. For α , Fehr and Schmidt (1999) report an interval of $[0, 4.5]$, with 0.833 as median value, whereas for β the interval is $[0, 0.6]$, with 0.288 as median. Subsequent studies, such as Dannenberg et al. (2010), Blanco et al. (2011) and Ponti and Rodriguez-Lara (2015), have tried, through ad-hoc experiments, to determine the values of α and β , roughly confirming the intervals

provided in Fehr and Schmidt (1999). In Rogna and Vogt (2020), the median values of α and β as provided in Fehr and Schmidt (1999) are used as base case and countries\regions are assumed to be homogeneous with respect to both parameters. In the present paper, the case of homogeneous values for all players, with $\alpha = 0.833$ and $\beta = 0.288$, will be used as benchmark, but an effort to obtain more realistic values is made.

One of the problematic aspects of the literature mentioned in the previous paragraph is that the estimates of α and β are almost always retrieved from laboratory experiments having individuals as subjects. If it is true that the main justification for assuming F&S preferences at country level stays in the median voter argument, as claimed in Rogna and Vogt (2020), it is likely that a simple translation of parameters from the individual to the country\regional level is potentially biased. Moreover, the assumption of homogeneity among all players with regard to both parameters seems also quite problematic. For example, when considering ethical consumerism, a choice generally regarded as subjective, Summers (2016) shows how structural and cultural contexts, mostly determined at country level, are important in shaping such choice. This causes significant differences among countries often related to their level of affluence. In a similar vein, envy and altruism may be culturally influenced, thus determining important inter-country differences.

In order to partly remedy to these shortcomings, we attempt to estimate the values of α by establishing a relation with the elasticity of marginal utility of consumption (EMU) and the level of per-capita consumption of each country\region. The possibility to estimate both α and β is precluded since there are too many unknowns for the system of equations that will be used, therefore we still need to assume a value for the β s. Furthermore, given the need of assuming such values, we have no indication on how to vary them among the countries\regions, therefore we prefer to retain the assumption of homogeneity. The choice of estimating α and not β is due to the fact that, in Rogna and Vogt (2020), the former has a much more important role in determining the stability of coalitions. Furthermore, pure altruism at country level seems less likely to exist than dissatisfaction for disadvantageous inequality, therefore its magnitude may be very small, leading to a negligible role.

It is important to remind here that F&S utility functions address the underlying causes of inequality aversion, namely interpersonal income comparisons, which are not captured by the concept of EMU. This has an important consequence: F&S utility functions capture the psychological externalities additionally to the classical, environmental external effects. With EMU, instead, only the diminishing marginal utility of consumption for higher levels of affluence is captured, leaving aside the psychological effect of inequality. This may have a significant impact on the optimal level of abatement and, consequently, on the level of temperature rise.

The reasoning behind our estimation procedure is as follow. Consider the definition of elasticity of marginal utility of consumption:

$$EMU(c) = \frac{dMU}{dc} \frac{c}{MU},$$

where c is consumption. In our definition of F&S utility we have defined $\pi_{i,t}$ as the discounted value of per-capita consumption. Let us drop, for mere convenience, the temporal dimension and equate π_i to c_i . Given the F&S utility function as in equation (1), the marginal utility of consumption (for the sake of brevity we avoid to repeat per-capita consumption) is simply given by:

$$MU_i(c_i) = 1 + \frac{\alpha_i}{n-1}|I^+| - \frac{\beta_i}{n-1}|I^-|, \quad \forall i \in N; \quad (3)$$

where $|I^+|$ and $|I^-|$ indicate the cardinality of the sets of players with a level of consumption higher and lower than player i , respectively. Now, if we consider a discrete increase in the consumption of player i such that she switches of one position in the consumption rank – e.g. $|I^+|$ will be decreased by one unit and, consequently, $|I^-|$ will be increased by the same amount –, we can define $\frac{dMU_i}{dc_i}$ as:

$$\frac{dMU_i}{dc_i} = \frac{MU_i(c_i^2) - MU_i(c_i^1)}{c_i^2 - c_i^1} =$$

$$= \frac{1 + \frac{\alpha_{i+1}}{n-1}(|I^+| - 1) - \frac{\beta_{i+1}}{n-1}(|I^-| + 1) - \left[1 + \frac{\alpha_i}{n-1}|I^+| - \frac{\beta_i}{n-1}|I^-|\right]}{\Delta_i}, \quad (4)$$

where, for sake of brevity, $c_i^2 - c_i^1 = \Delta_i$. Note that we have α_{i+1} and β_{i+1} since, at level of consumption c_i^2 , player i shifted up of one position in the consumption ranking. By dividing c_i for the RHS of equation (3) and multiplying the result to the RHS of equation (4), we get EMU_i :

$$EMU_i = \frac{c_i(n-1)}{n-1+\alpha_i|I^+|-\beta_i|I^-|} \frac{n-1+\alpha_{i+1}(|I^+|-1)-\beta_{i+1}(|I^-|+1)-\left(n-1+\frac{\alpha_i}{n-1}|I^+|-\frac{\beta_i}{n-1}|I^-|\right)}{\Delta_i}, \quad \forall i \in N. \quad (5)$$

By using the values for the elasticity of marginal consumption (EMU_i) – equal to 1.5 for all countries\regions – and for the per-capita consumption (c_i) – η_i and $\frac{C_{i,2015}}{L_{i,2015}}$, respectively – provided in the RICE v2013 model, we are left only with the α s and β s as unknown, since n , I^+ , I^- and Δ_i are also given or can be easily computed from the former values. We have then a system of n equations with $2 \times n$ unknowns, that can be solved for the α s provided that the β values are fixed by assumption. Specifically, a value of 0.1, lower than the median value of 0.288 proposed in Fehr and Schmidt (1999) given that we assume countries to be less altruistic than individuals, is adopted. However, a further assumption is required. Due to the fact that $I^+ = \emptyset$ for the most affluent country (US in the RICE database), we are basically left in shortage of one equation since $\alpha_i|I^+| = 0$ for this country. Therefore, also this value of α must be assumed and we have set $\alpha_{US} = 0.11$. In Table 1, the estimated values of the α s for all countries\regions are reported.

Table 1: Estimated α values for all countries\regions

Country\region	Country Code	α value	Country\region	Country Code	α value
U.S.	US	0.1100	India	IND	3.5533
E.U.	EU	0.1490	Middle East	MEST	0.7112
Japan	JAP	0.1183	Africa	AFR	5.0128
Russia	RUS	0.4592	Latin America	LAM	0.7500
East Europe	EUR	1.2507	Other industrialized countries	OHI	0.1225
China	CHI	1.6860	Other South-East Asian countries	OTH	2.6116

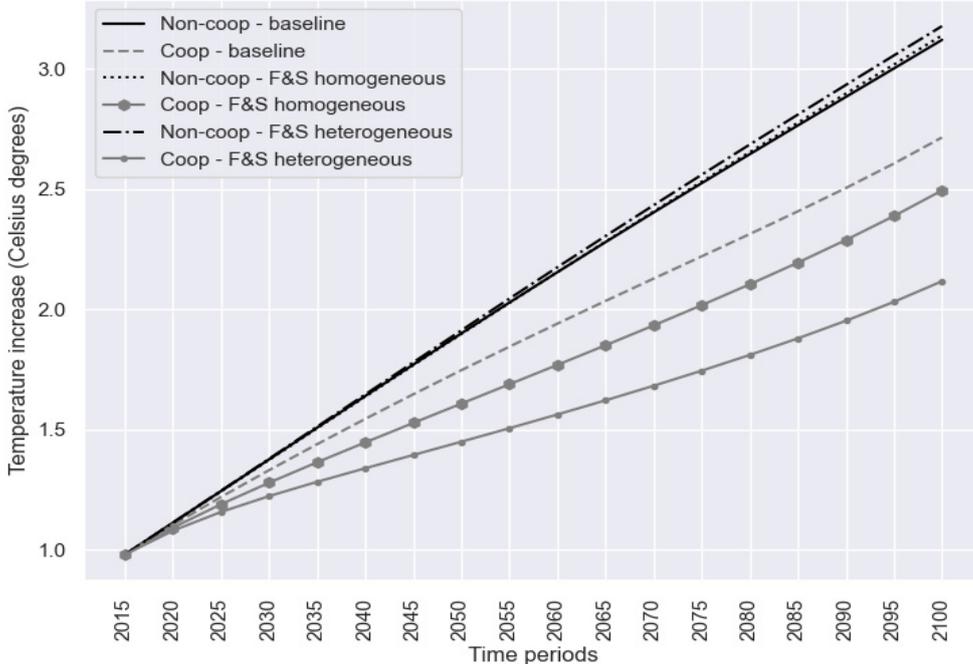
Note that all the estimated α values respect the Fehr and Schmidt (1999) assumption of $\alpha > \beta$ and that, except for Africa, they are all included in the $[0, 4.5]$ interval. Furthermore, the magnitude of α is increasing in the level of per-capita consumption.

4 Simulations and results

In our basic simulations, three scenarios are benchmarked, the original RICE v2013 model, the modified version with F&S preferences and homogeneous α and β parameters – with values, respectively, of 0.833 and of 0.288 – and the version with F&S preferences, homogeneous β – equal to 0.1 – and heterogeneous α s, with values given in Table 1. The simulation is run from 2015 to 2100², with a 5 years time step. In all three cases, all possible coalitions have been examined, meaning 4095 coalitions since the model features 12 countries\regions. The way of solving the model follows the original algorithm described in Nordhaus and Yang (1996), while [Pyomo](#), a Python package for modeling optimization problems, has been used for the computation.

²Actually, we report the results till the year 2100, but we have run the simulation till 2110 to reduce the last periods drop in abatement consequent to a “no future” scenario.

Figure 1: Temperature rise under Different Scenarios

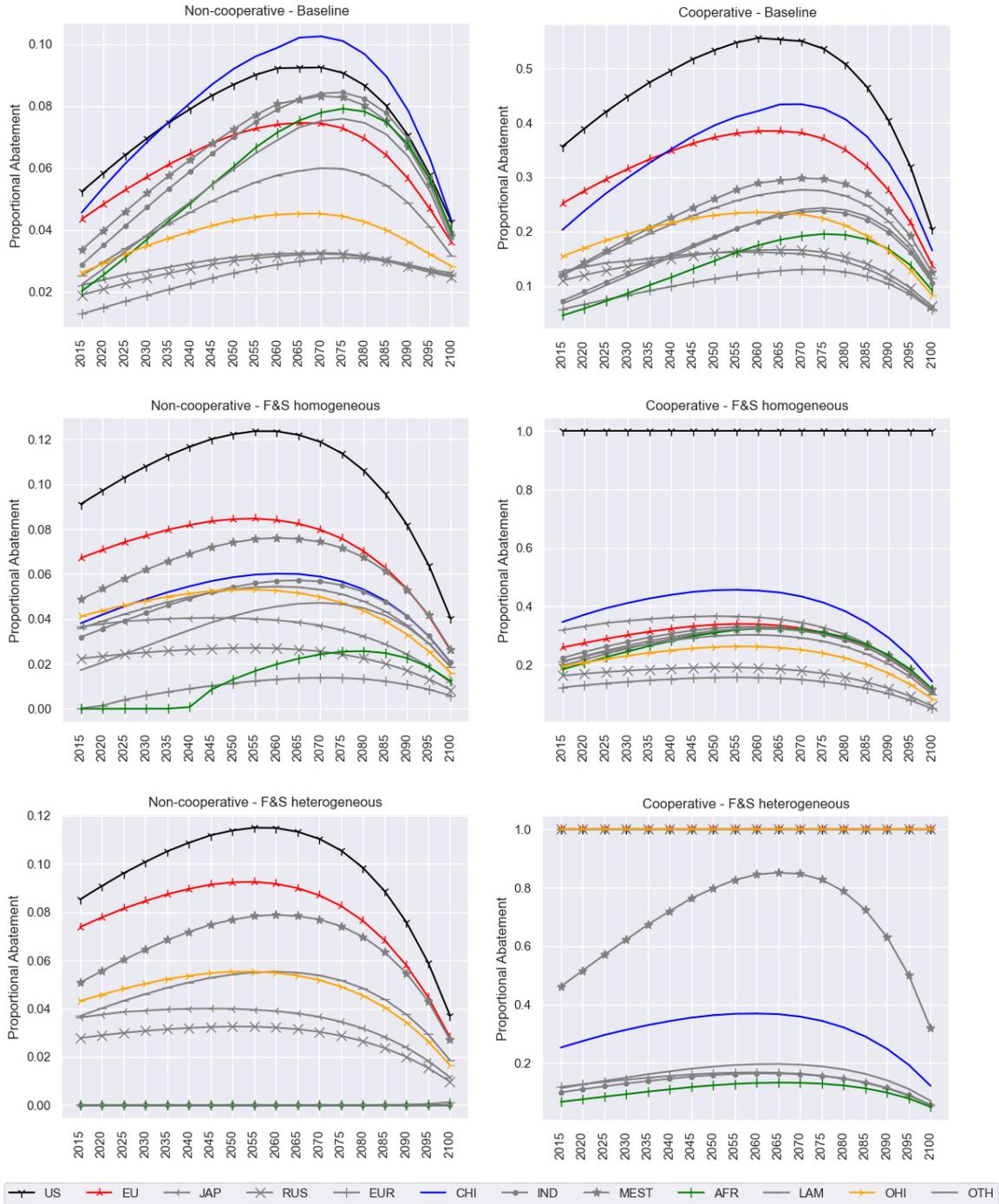


The first thing to be examined is the temperature increase above the pre-industrial level obtained in the different scenarios. Each of the three cases mentioned above is further subdivided into two: the cooperative case (grand coalition in game-theoretic jargon), and the non-cooperative case, where no multi-countries coalition is formed. From Figure 1, it is possible to observe significant differences among the various scenarios. Clearly, the three cooperative cases lead to a lower level of final temperature rise than their non-cooperative counterparts. However, it is interesting to note as, with the standard utility functions of RICE, even under full cooperation, the rise in temperature in 2100 is of 2.73° C circa, whereas with F&S preferences and heterogeneous α s the rise is far more modest, approximately 2.10° C. Despite being still far from the 1.5° target mentioned by the Paris Agreement, it is very close to the 2° threshold. Further note that the cooperative case with identical α s and β s is in between, with an end of periods temperature rise equal to 2.60°

circa. This intermediate outcome is found also for the non-cooperative scenario, where F&S preferences with homogeneous inequality aversion parameters lead to a temperature rise of 3.26° . However, the other two cases are switched, with F&S preferences and heterogeneous α s leading to the highest temperature rise, 3.30° , and the standard RICE leading to a temperature rise of 3.24° C in 2100. Taken together, the introduction of F&S preferences does not lead to significant differences in the non-cooperative case, while results in the full cooperation scenario differ widely. Accounting for inequality aversion in an F&S framework is capable of providing support for the two degrees goal, even if all other critical determinants are kept constant.

With this being the level of temperature rise, it is also interesting to see how much abatement is undertaken by each country\region in each scenario. This is represented in Figure 2, where all six scenarios are reported. Abatement is represented in proportional terms, so that 1 means a total abatement of polluting emissions. It is interesting to notice that, under F&S preferences with heterogeneous α s, all the highly industrialised countries\regions (US, EU, Japan (JPN), other industrialised countries (OHI)), plus Russia (RUS) and Latin America (LAM) totally abate their emissions in almost all periods in the cooperative scenario. With homogeneous inequality aversion parameters, instead, only US abates in all periods, whereas all other countries\regions drastically reduce their level of abatement. Such reduction is even more pronounced in the baseline scenario, where US is the only country to cut more than 50% of its emissions and just for a limited period of time. In the non-cooperative scenarios, the level of abatement is dramatically lower, hardly reaching the 12%. It is interesting to note that, with F&S preferences and heterogeneous α s, the most affluent players undertake most of the abatement efforts, in both the cooperative and non-cooperative scenario, whereas this is far less evident in the other cases. If this difference was expected in the comparison with the original RICE model, it is less obvious when the comparison is made with the homogeneous α s and β s case. In particular, since the level of altruism (β) of all players is higher in this second scenario, the reverse might be expected. However, as it can be seen in the non-cooperative case with F&S preferences

Figure 2: Abatement under Different Scenarios



and heterogeneous α s, the high level of “envy” of less affluent countries\regions leads them to totally avoid abatement, obliging the richest players to support the whole environmental burden.

4.1 Stability of coalitions

After having examined the temperature and abatement trajectories in the two extreme cases, namely no cooperation and grand coalition, the focus will now be placed on the stability of coalitions. We will therefore consider all the intermediate possibilities of aggregation of countries\regions and look at their stability using the well known criterion of d’Aspremont et al. (1983): a coalition is stable when both internal – no member of the coalition has an incentive to leave – and external – no outsider has an incentive to join the coalition – stability are satisfied. The condition of potential internal stability (PIS) is further considered, being it a weaker version of internal stability according to which a coalition satisfies PIS if it can be internally stabilised via transfers among its members.

Table 2 shows the number of stable coalitions for each scenario divided into coalition size and for the stability condition considered. The grand coalition is excluded from the table but it is neither internally stable nor potentially internally stable in any of the three considered scenarios. As largely shown in the game-theoretic literature on environmental coalitions formation (e.g. Carraro and Siniscalco (1993) and Barrett (1994)), the number of stable coalitions is very low, particularly due to the difficulty of reaching internal stability. Furthermore, stable coalitions have a very modest size. However, it is interesting to note the difference existing between the various scenarios, with the baseline case having no stable coalitions at all, whereas when F&S preferences and heterogeneous α s are considered the number increases to 14, with some four players coalitions being stable. The case of F&S preferences with homogeneous inequality aversion parameters is very similar to the baseline scenario, with only a two players coalition being stable and a very modest number of internally stable coalitions. If the possibility to obtain significant environmental achievements thanks to the formation of climate coalitions seems very scarce, once considering the possi-

bility of internal transfers, the picture becomes less grim. Once again, it is the case of F&S preferences and heterogeneous α s to have the highest number of coalitions satisfying PIS, with 11 of them being 8 players coalitions. When homogeneous α s and β s are considered, the number drops to less than one third and the maximal size of potentially internally stable coalitions is of 6 players, further decreasing to 4 in the baseline scenario.

Table 2: Number of Stable Coalitions under Different Scenarios

Baseline Scenario (Original RICE Model)											
<i>Coalition Size</i>	2	3	4	5	6	7	8	9	10	11	<i>Total</i>
<i>Fully Stable</i>	0	0	0	0	0	0	0	0	0	0	0
<i>Internally stable</i>	5	3	0	0	0	0	0	0	0	0	8
<i>Externally stable</i>	4	26	49	104	163	206	170	102	45	11	880
<i>PIS</i>	51	151	16	0	0	0	0	0	0	0	218
<i>PIS + Externally st.</i>	4	24	2	0	0	0	0	0	0	0	30
F&S Preferences ($\alpha = 0.833, \beta = 0.288$)											
<i>Coalition Size</i>	2	3	4	5	6	7	8	9	10	11	<i>Total</i>
<i>Fully stable</i>	1	0	0	0	0	0	0	0	0	0	1
<i>Internally stable</i>	7	1	0	0	0	0	0	0	0	0	8
<i>Externally stable</i>	1	8	28	65	98	133	88	30	28	7	486
<i>PIS</i>	54	197	165	39	2	0	0	0	0	0	457
<i>PIS + Externally st.</i>	1	3	0	0	0	0	0	0	0	0	4
F&S Preferences with Heterogeneous α Values											
<i>Coalition Size</i>	2	3	4	5	6	7	8	9	10	11	<i>Total</i>
<i>Fully stable</i>	3	7	4	0	0	0	0	0	0	0	14
<i>Internally stable</i>	11	21	26	27	14	0	0	0	0	0	99
<i>Externally stable</i>	5	22	38	34	33	44	36	30	20	6	268
<i>PIS</i>	53	193	321	426	419	131	11	0	0	0	1554
<i>PIS + Externally st.</i>	5	15	16	15	17	6	0	0	0	0	74

The last row for each scenario (PIS + Externally st.) indicates the coalitions that are both externally and potentially internally stable, therefore they can be rendered stable via

transfers. While it is not surprising that the highest number is reached in the last scenario – 74 coalitions of which 6 with seven members –, the very low number under F&S preferences with homogeneous inequality aversion is rather astonishing. In the baseline scenario there are 30 potentially stable coalitions with maximal size of four players.

4.2 A condition for stability

Vogt (2016) provides a condition to obtain internal stability under F&S preferences. Remembering that internal stability is defined as $U_i(C) \geq U_i(C \setminus i)$, the condition reads as:

$$\frac{\alpha_i}{n-1} \left(\sum_{j \in I^+} [\pi_j(C \setminus i) - \pi_i(C \setminus i)] - \sum_{j \in I^+} [\pi_j(C) - \pi_i(C)] \right) \geq \pi_i(C \setminus i) - \pi_i(C) +$$

$$\frac{\beta}{n-1} \left(\sum_{k \in I^-} [\pi_i(C) - \pi_k(C)] - \sum_{k \in I^-} [\pi_i(C \setminus i) - \pi_k(C \setminus i)] \right).$$

If this inequality is satisfied for each member of coalition C , than C is internally stable. Our numerical simulation has fully confirmed this result, that, besides providing a method to check for coalition stability, is also useful to understand the reasons inducing to internal stability.

As mentioned in Vogt (2016), the bracket term on the LHS of this inequality reflects the development of disadvantageous inequality for player i after leaving coalition C , with the sign of this term being unknown a priori: disadvantageous inequality may be either decreased or increased by i 's exit. The RHS, instead, includes the material gain obtained by player i in exiting, plus the development of advantageous inequality. By leaving a coalition, and in absence of transfers, a player can generally expect to improve her absolute as well as her relative payoff. This incentive increases with increasing values of α s. However, this effect can be contrasted, and even reversed, by introducing well tailored transfer schemes as in Rogna and Vogt (2020). In particular, these lasts are effective only if they are capable of sufficiently reducing disadvantageous inequality for a certain set of players. This clearly is part of an explanation why the number of PIS coalitions is dramatically increased when

introducing (heterogeneous) inequality aversion.

4.3 Environmental targets of stable coalitions

Besides leading to a temperature rise close to the 2° C threshold in the cooperative case, that is, however, not stable, the F&S scenario with heterogeneous α s allows for more stable coalitions and for even more coalitions that can be fully stabilised through transfers. However, the simple number of stable coalitions may be considered as scarcely informative, since it does not tell much about the outcomes. In particular, it is interesting to see, in the various scenarios, which is the best and realistically achievable target. By adopting an environmental perspective, we consider as best the coalition that allows for the lowest temperature rise in the last period, and as realistically achievable a coalition that is stable or potentially stable. Clearly, in the baseline scenario, since there are no stable coalitions, we have the same outcome described earlier, with a temperature rise of 3.24° C in 2100. Even in absence of cooperation, this is a lower temperature rise than the one obtainable when F&S preferences are assumed. Under homogeneous α s and β s, in fact, the only stable coalition, between East European (EUR) and Middle East (MEST) countries, leads to a temperature rise of 3.25° C, while with heterogeneous α s it increases to 3.26° C, obtained by the coalition of USA, western Europe (EU) and the other industrialized countries (OHI). It should be noted that all the four players coalitions in this last mentioned scenario achieved a less environmental friendly outcome compared to the mentioned coalition.

When considering potentially stable coalitions, in the baseline scenario the lowest reachable temperature rise is of 3.17° C, with the coalition of USA, EU, Russia and China. With F&S preferences and homogeneous inequality aversion parameters we have an increase of 3.19° C thanks to the coalition of Japan (JPN), China (CHI) and Latin America (LAM). Finally, with heterogeneous α s, the seven players coalition of EU, Japan, China, India (IND), Middle East, other industrialised countries and other South-East Asian (OTH) countries limits the temperature rise to 3.05° C. This is the minimum temperature rise obtainable by allowing only for transfers among coalition members.

Some remarks are worth to be made. First of all, it should be noted the large temperature gap existing between the grand coalition and the most environmental friendly, and realistically achievable, coalition with F&S preferences and heterogeneous α s. Despite being a seven players coalition that includes several industrialized countries and a heavy polluter such as China, this last coalition leads to a temperature rise almost one degree higher than the grand coalition. F&S preferences lead to both lower temperatures rise when the grand coalition is reached and to higher degrees of cooperation, that are further conducive of more abatement. However, as shown in this last section, the possibility to meet the 1.5 or even the 2 degrees target through self-sustaining coalitions is far even when other regarding preferences are accounted for.

4.4 Sensitivity analysis

In this last section, the sensitivity of the results to variations in the α and β parameters is tested. We focus on the two extreme cases, the grand coalition and the total absence of cooperation, looking at the end of period temperature rise for different values of the homogeneous β s and for an alternative assumed α value for US. Table 3 reports the estimated α values under the two different assumptions. By comparing Table 3 with Table 1, it is possible to observe that a different value of α for US, now equal to 0.833 compared to 0.11 adopted in the previous simulations, does not cause a dramatic change in the estimated α values of the remaining players, many of which are actually identical.

It comes with no surprise, therefore, that also the environmental outcomes of the new simulation do not differ substantially from the ones obtained before. In the cooperative case, the temperature rise in 2100 is equal to 2.21° C, meaning 0.11° C higher than in the simulation with α of US equal to 0.11. With regard to the non-cooperative case, the final temperature rise is of 3.30° C, thus identical to the previous simulation with heterogeneous α s.

It is interesting to notice that, even under a different value of β for all countries (0.2 compared to 0.1 in previous simulations), that causes a significant variation in the esti-

mated α values, as can be seen from Table 3, the outcomes are still almost unchanged. The temperature rise in the cooperative case, in fact, is still equal to 2.21° C, while it is equal to 3.29° C in the non-cooperative case, a mere difference of 0.1° with previous results. The outcomes of the simulation, therefore, are quite robust, at least to relatively small perturbations in the parameters governing the aversion to advantageous and disadvantageous inequality.

Table 3: α Values for different β s and US's α

	US	EU	JAP	RUS	EUR	CHI
$\beta = 0.2$	0.1000	0.0444	0.0161	0.3489	1.1257	1.5529
$\beta = 0.1$	0.8330	0.1490	0.1327	0.4592	1.2507	1.6860
	IND	MEST	AFR	LAM	OHI	OTH
$\beta = 0.2$	3.3856	0.5962	4.8181	0.6343	0.0184	2.4614
$\beta = 0.1$	3.5533	0.7112	5.0128	0.7500	0.1228	2.6116

In a last set of simulations that completes the sensitivity analysis, we have kept the values of β at 0.1 and 0.2, but varying the ones of α US in a wider range, namely from 0.01 to 4. The calibrated α values for all other countries can be seen in Table A31 in the Appendix. In Table A32, the temperature increase in year 2100 for the cooperative and the non-cooperative case can be observed for each different calibration. From Table A32 we can see that, even for very large variations in the value of α of US, the impact on the final temperature rise is very modest, mainly because the estimated α s for the other countries are scarcely affected (see Table A31). We can notice an increasing level of temperature rise for increasing values of US α both in the cooperative and in the non-cooperative case, but this rise is so subtle – at the fifth decimal place – that can be considered as irrelevant.

5 Conclusions

The economic–environmental models that have been used to investigate the opportunity–costs of limiting the temperature rise caused by greenhouse gases have been often criticized

for suggesting low levels of CO₂ emissions reduction. A very well known example is one of the first adopted models to undertake this type of analysis, the DICE\RICE model, whose predicted optimal path leads to an increase of roughly 3° C over the pre-industrial mean temperature at the end of the century. This is double compared to the 1.5° C threshold proposed in the Paris Agreement. Several other models reach similar conclusions. The critiques have targeted the computation of climatic damages, judged too mild, the excessive discounting of future payoffs and the lack of consideration for irreversibility and catastrophic risks. The inclusion of these elements generally leads to larger abatement efforts under optimality.

Another important aspect in the challenge of limiting anthropogenic climate change is the re-distribution imbued in this process. The burden sharing, i.e. the differentiation of abatement targets across countries is one of the most prominent and controversial issues present in international climate policy from its very beginnings and evidenced, for example, in the principle of common but differentiated responsibilities laid down in the UN Framework Convention on Climate Change from 1992. The present paper tries to shed light on this aspect by assuming that the actors (countries and regions) involved in the negotiation of abatement efforts have Fehr and Schmidt (1999) preferences. Compared to the standard utility functions adopted in impact assessment models, where the concavity of the utility function described by the elasticity of marginal utility (EMU) of consumption is the only consideration paid to income differences, the F&S function has inequality aversion as its central focus.

By adopting the RICE model v2013 and leaving proportional abatement as the only control variable, the paper compares the results obtained, for the period 2015–2100, in three different scenarios: the standard, unaltered, model, the one with F&S preferences and homogeneous parameters of inequality aversion and the one with F&S preferences and heterogeneous α values, with α being the degree of aversion to disadvantageous inequality (“envy”). This last parameter has been estimated by relating it with the value of EMU as provided in RICE. As it turns out, the parameter is a decreasing function of country per

capita income, i.e. richer countries are showing relatively lower degrees of aversion towards disadvantageous inequality. The estimation procedure leads to results consistent with the range of values reported in Fehr and Schmidt (1999).

The results show that, when considering F&S utility functions in conjunction with a cooperative behaviour by the players (grand coalition), there is a significant decrease in the final level of temperature rise compared to the standard RICE model. This is much more pronounced in the case of heterogeneous α s, where the temperature rise in the final period is of 2.1° C. Thus, systematically accounting for inequality aversion lends support to the two degrees goal favoured by climatologists, even if all other critical determinants of optimal temperature increase, particularly the discount rate and the damage function, are kept constant. However, when no-cooperation is considered, F&S utility functions have an opposite role, leading to lower levels of overall abatement and, therefore, to higher temperature rise. However, the effect of F&S preferences in the non-cooperative Nash equilibrium is rather modest.

In all three scenarios, the grand coalition is not stable and, as evidenced in the environmental game-theoretic literature, cooperation is hard to reach. However, it is important to note that under F&S preferences with heterogeneous α s, the number and size of stable coalitions is higher than in the other cases. This is quite surprising particularly if compared with the case of F&S preferences and homogeneous inequality aversion parameters, where the assumed value of β , the degree of altruism, is higher. Envy, when inversely correlated to affluence, seems to have a stabilizing capability on cooperation. Such a stabilizing effect, able to ensure a self-sustaining coalition of seven players, is not able, however, to guarantee an environmental target close to the 2° C level. With this seven players coalition, in fact, the temperature rise is of 3.05° C. By considering modification to the damage function or other corrective mechanisms leading to stricter levels of abatement in conjunction with F&S utility function, it may still be possible to find self-sustaining coalitions leading to a temperature rise close to the Paris Agreement threshold. This is left for future studies.

References

- Ackerman, F. and Munitz, C. (2016). A critique of climate damage modeling: Carbon fertilization, adaptation, and the limits of FUND. *Energy Research & Social Science*, 12:62–67.
- Ackerman, F., Stanton, E. A., and Bueno, R. (2010). Fat tails, exponents, extreme uncertainty: Simulating catastrophe in DICE. *Ecological Economics*, 69(8):1657–1665.
- Barrett, S. (1994). Self-enforcing International Environmental Agreements. *Oxford Economic Papers*, pages 878–894.
- Blanco, M., Engelmann, D., and Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, 2(72):321–338.
- Bolton, G. E. and Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.
- Bosetti, V., Carraro, C., Galeotti, M., Massetti, E., and Tavoni, M. (2006). A world induced technical change hybrid model. *The Energy Journal*, (Special Issue# 2).
- Botzen, W. W. and van den Bergh, J. C. (2012). How sensitive is Nordhaus to Weitzman? Climate policy in DICE with an alternative damage function. *Economics Letters*, 117(1):372–374.
- Botzen, W. W., Van Den Bergh, J. C., and Chichilnisky, G. (2018). Climate policy without intertemporal dictatorship: Chichilnisky criterion versus classical utilitarianism in DICE. *Climate Change Economics*, 9(02):1850002.
- Carraro, C. and Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of Public Economics*, 52(3):309–328.
- Dannenberg, A., Sturm, B., and Vogt, C. (2010). Do equity preferences matter for climate negotiators? An experimental investigation. *Environmental and Resource Economics*, 47(1):91–109.

- d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., and Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics*, pages 17–25.
- De Bruin, K. C., Dellink, R. B., and Tol, R. S. (2009). AD-DICE: An implementation of adaptation in the DICE model. *Climatic Change*, 95(1):63–81.
- Dellink, R. (2011). Drivers of stability of climate coalitions in the STACO model. *Climate Change Economics*, 2(02):105–128.
- Diaz, D. and Moore, F. (2017). Quantifying the economic risks of climate change. *Nature Climate Change*, 7(11):774–782.
- Dietrich, J. P., Bodirsky, B. L., Humpenöder, F., Weindl, I., Stevanović, M., Karstens, K., Kreidenweis, U., Wang, X., Mishra, A., Klein, D., et al. (2019). MAgPIE 4—a modular open-source framework for modeling global land systems. *Geoscientific Model Development*, 12(4):1299–1317.
- Dietz, S. and Asheim, G. B. (2012). Climate policy under sustainable discounted utilitarianism. *Journal of Environmental Economics and Management*, 63(3):321–335.
- Dietz, S., Gollier, C., and Kessler, L. (2018). The climate beta. *Journal of Environmental Economics and Management*, 87:258–274.
- Easter, R. C., Ghan, S. J., Zhang, Y., Saylor, R. D., Chapman, E. G., Laulainen, N. S., Abdul-Razzak, H., Leung, L. R., Bian, X., and Zaveri, R. A. (2004). MIRAGE: Model description and evaluation of aerosols and trace gases. *Journal of Geophysical Research: Atmospheres*, 109(D20).
- Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.
- Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism—experimental evidence and new theories. *Handbook of the Economics of Giving, Altruism and Reciprocity*, 1:615–691.

- Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D. L., Rao, N. D., Riahi, K., Rogelj, J., De Stercke, S., et al. (2018). A low energy demand scenario for meeting the 1.5° C target and sustainable development goals without negative emission technologies. *Nature Energy*, 3(6):515–527.
- Hänsel, M. C. and Quaas, M. F. (2018). Intertemporal distribution, sufficiency, and the social cost of carbon. *Ecological Economics*, 146:520–535.
- Hansson, A., Fridahl, M., Anshelm, J., Haikola, S., et al. (2021). Boundary work and interpretations in the IPCC review process of the role of bioenergy with carbon capture and storage (BECCS) in limiting global warming to 1.5° C. *Frontiers in Climate*, 3:34.
- Howard, P. H. and Sterner, T. (2017). Few and not so far between: a meta-analysis of climate damage estimates. *Environmental and Resource Economics*, 68(1):197–225.
- Jewell, J. and Cherp, A. (2020). On the political feasibility of climate change mitigation pathways: Is it too late to keep warming below 1.5° C? *Wiley Interdisciplinary Reviews: Climate Change*, 11(1):e621.
- Keramidas, K., Kitous, A., Després, J., and Schmitz, A. (2017). POLES-JRC model documentation. *Publications Office of the European Union*, 10:225347.
- Knopf, B., Kowarsch, M., Flachsland, C., and Edenhofer, O. (2012). The 2° C target reconsidered. In *Climate Change, Justice and Sustainability*, pages 121–137. Springer.
- Knutti, R., Rogelj, J., Sedláček, J., and Fischer, E. M. (2016). A scientific critique of the two-degree climate change target. *Nature Geoscience*, 9(1):13–18.
- Lange, A. and Vogt, C. (2003). Cooperation in international environmental negotiations due to a preference for equity. *Journal of Public Economics*, 87(9):2049–2067.
- Michaelis, P. and Wirths, H. (2020). DICE-RD: an implementation of rate-related damages in the DICE model. *Environmental Economics and Policy Studies*, 22(4):555–584.

- Nordhaus, W. (2007). Accompanying notes and documentation on development of dice-2007 model: Notes on dice-2007. delta. v8 as of september 21, 2007. *Miscellaneous publication, Yale University, New Haven, NE, USA*.
- Nordhaus, W. (2018a). Projections and uncertainties about climate change in an era of minimal climate policies. *American Economic Journal: Economic Policy*, 10(3):333–60.
- Nordhaus, W. (2018b). Evolution of modeling of the economics of global warming: changes in the DICE model, 1992–2017. *Climatic Change*, 148(4):623–640.
- Nordhaus, W. D. (1992). An optimal transition path for controlling greenhouse gases. *Science*, 258(5086):1315–1319.
- Nordhaus, W. D. (2010). Economic aspects of global warming in a post-Copenhagen environment. *Proceedings of the National Academy of Sciences*, 107(26):11721–11726.
- Nordhaus, W. D. (2013). *The climate casino*. Yale University Press.
- Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7):1518–1523.
- Nordhaus, W. D. and Boyer, J. (2000). *Warming the world: economic models of global warming*. MIT press.
- Nordhaus, W. D. and Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, 86(4):741–765.
- Ponti, G. and Rodriguez-Lara, I. (2015). Social preferences and cognitive reflection: evidence from a dictator game experiment. *Frontiers in Behavioral Neuroscience*, 9:146.
- Rogna, M. and Vogt, C. (2020). *Coalition formation with optimal transfers when players are heterogeneous and inequality averse*. Number 865. Ruhr Economic Papers.
- Rosen, R. A. (2015). IAMs and peer review. *Nature Climate Change*, 5(5):390–390.

- Roughgarden, T. and Schneider, S. H. (1999). Climate change policy: quantifying uncertainties for damages and optimal carbon taxes. *Energy Policy*, 27(7):415–429.
- Summers, N. (2016). Ethical consumerism in global perspective: A multilevel analysis of the interactions between individual-level predictors and country-level affluence. *Social Problems*, 63(3):303–328.
- Tol, R. S. (1994). The damage costs of climate change: a note on tangibles and intangibles, applied to DICE. *Energy Policy*, 22(5):436–438.
- Tol, R. S. (2019). A social cost of carbon for (almost) every country. *Energy Economics*, 83:555–566.
- van der Pol, T., Weikard, H.-P., and van Ierland, E. (2012). Can altruism stabilise international climate agreements? *Ecological Economics*, 81:112–120.
- Vogt, C. (2016). Climate coalition formation when players are heterogeneous and inequality averse. *Environmental and Resource Economics*, 65(1):33–59.
- Weitzman, M. L. (2012). GHG targets as insurance against catastrophic climate damages. *Journal of Public Economic Theory*, 14(2):221–244.
- Yang, Z. et al. (2008). *Strategic bargaining and cooperation in greenhouse gas mitigations: an integrated assessment modeling approach*. MIT Press.

Appendix

A1 RICE v2013: Model's sets, parameters and variables

Sets

Countries: $N = \{\text{US, EU, JPN, RUS, EUR, CHI, IND, MEST, AFR, LAM, OHI, OTH}\}$.

Time periods: $T = \{1, 2, \dots, 17\}$, equivalent to 2015, 2020, ..., 2100.

Endogenous Variables

W_i = Present value of the sum of discounted utility of country i ,

$U_{i,t}$ = Discounted utility of country i at time t ,

$C_{i,t}$ = Consumption of country i at time t ,

$K_{i,t}$ = Capital of country i at time t ,

$I_{i,t}$ = Investments of country i at time t ,

$Q_{i,t}$ = Gross output of country i at time t ,

$Y_{i,t}$ = Output, net of environmental damages and abatement expenditures, of country i at time t ,

$B_{i,t}$ = Abatement expenditures of country i at time t ,

$D_{i,t}$ = Environmental damages (proportion of output) for country i at time t ,

$E_{i,t}^{ind}$ = Industrial (economic driven) emissions of country i at time t ,

E_t^{tot} = Total emissions (sum over all countries) at time t ,

M_t^{at} = Atmospheric CO₂ concentration at time t ,

M_t^{up} = CO₂ concentration in biosphere/shallow oceans at time t ,

M_t^{lo} = CO₂ concentration in deep oceans at time t ,

T_t = Atmospheric temperature rise over 1900 level at time t ,

T_t^{lo} = Temperature of deep oceans at time t ,

F_t = Radiative force at time t ,

$\mu_{i,t}$ = Proportion of abatement of country i at time t (control variable).

Parameters and Exogenous Variables

Symbol	Description	Dependency
ts	time step (5 years)	Invariant
$L_{i,t}$	Population and labour force (exogenous)	Country—time dependent
η_i	Marginal Utility of Consumption	Country dependent
ρ_i	Coefficient of inter-temporal discounting	Country dependent
δ_i^k	Capital depreciation rate	Country dependent
$\sigma_{i,t}^I$	Rate of saving—investment	Country—time dependent
$a_{i,t}$	Cobb—Douglas efficiency parameter	Country—time dependent
γ_i	Cobb-Douglas share parameter	Country dependent
$\theta_{i,t}^1, \theta_{i,t}^2$	Coefficients determining the abatement costs	Country—time dependent
$d_i^1, d_i^2, d_i^3, d_i^4, d_i^5$	Environmental damages parameters	Country dependent
sl_t	Sea level rise coefficient	Time dependent
$\sigma_{i,t}^e$	Emissions intensity coefficient	Country—time dependent
$E_{i,t}^{land}$	Emissions from land use change (exogenous)	Country—time dependent
$b_{11}, b_{2,1}, b_{1,2}, b_{2,2},$ $b_{3,2}, b_{2,3}, b_{3,3}$	CO ₂ concentration determinants	Invariant
c_1, c_2, c_3, c_4	Temperature determinants	Invariant
ξ	Radiative force determinant	Invariant
M_{1900}^{at}	Temperature in 1900	Invariant
F_t^{ex}	Exogenous radiative force	Invariant

A2 RICE v2013: Model's equations (with exogenous investment\ saving rate)

$$W_i = \sum_{t \in T} U_{i,t}, \quad (A1)$$

$$U_{i,t} = \frac{1}{1-\eta_i} \left(\frac{C_{i,t}}{L_{i,t}} \right)^{1-\eta_i} + 1, \quad (A2)$$

$$K_{i,t} = tsI_{i,t-1} + (1 - \delta_i^k)^{ts} K_{i,t-1}, \quad (A3)$$

$$I_{i,t} = \sigma_{i,t}^I Y_{i,t}, \quad (\text{A4})$$

$$Q_{i,t} = a_{i,t} K_{i,t}^{\gamma_i} L_{i,t}^{1-\gamma_i}, \quad (\text{A5})$$

$$Y_{i,t} = Q_{i,t} - D_{i,t} \frac{Q_{i,t}}{1 + D_{i,t}^{ts}} - B_{i,t}, \quad (\text{A6})$$

$$B_{i,t} = \theta_{i,t}^1 \mu_{i,t}^{\theta_{i,t}^2} Q_{i,t}, \quad (\text{A7})$$

$$D_{i,t} = 0.01 \left(d_i^1 * T_t + d_i^2 T_{i,t}^{d_i^3} \right) + 2 \left(d_i^4 s l_{t-1} + d_i^5 s l_{t-1}^2 \right) \left(\frac{Q_{i,t}}{Y_{i,0}} \right)^{0.25}, \quad (\text{A8})$$

$$C_{i,t} = Y_{i,t} - I_{i,t}, \quad (\text{A9})$$

$$E_{i,t}^{ind} = \sigma_{i,t}^e (1 - \mu_{i,t}) Q_{i,t}, \quad (\text{A10})$$

$$E_t^{tot} = \sum_{i \in N} \left(E_{i,t}^{ind} + E_{i,t}^{land} \right), \quad (\text{A11})$$

$$M_t^{at} = ts E_{t-1}^{tot} + b_{11} M_{t-1}^{at} + b_{21} M_{t-1}^{up}, \quad (\text{A12})$$

$$M_t^{up} = b_{12} M_{t-1}^{at} + b_{22} M_{t-1}^{up} + b_{32} M_{t-1}^{lo}, \quad (\text{A13})$$

$$M_t^{lo} = b_{23} M_{t-1}^{up} + b_{33} M_{t-1}^{lo}, \quad (\text{A14})$$

$$T_t = T_{t-1} + c_1 \left[F_t - c_2 T_{t-1} - c_3 \left(T_{t-1} - T_{t-1}^{lo} \right) \right], \quad (\text{A15})$$

$$T_t^{lo} = T_{t-1}^{lo} + c_4 \left(T_{t-1} - T_{t-1}^{lo} \right), \quad (\text{A16})$$

$$F_t = \xi \frac{\log\left(\frac{M_t^{at}}{M_{1900}^{at}}\right)}{\log(2)} + F_t^{ex}. \quad (\text{A17})$$

A3 Other sensitivity analysis results

Table A31: Calibrated α values for different β s and US α s

β	α	US	US	EU	JAP	RUS	EUR	CHI
0.1	0.11	0.11	0.11	0.149	0.1183	0.4592	1.2507	1.686
0.1	0.21	0.21	0.21	0.149	0.1203	0.4592	1.2507	1.686
0.1	0.5	0.5	0.5	0.149	0.1261	0.4592	1.2507	1.686
0.1	1	1	1	0.149	0.136	0.4592	1.2507	1.686
0.1	2	2	2	0.149	0.1558	0.4592	1.2507	1.686
0.1	4	4	4	0.149	0.1955	0.4592	1.2507	1.686
0.2	0.01	0.01	0.01	0.0444	0.0144	0.3489	1.1257	1.5529
0.2	0.21	0.21	0.21	0.0444	0.0183	0.3489	1.1257	1.5529
0.2	0.5	0.5	0.5	0.0444	0.0241	0.3489	1.1257	1.5529
0.2	1	1	1	0.0444	0.034	0.3489	1.1257	1.5529
0.2	2	2	2	0.0444	0.0538	0.3489	1.1257	1.5529
0.2	4	4	4	0.0444	0.0935	0.3489	1.1257	1.5529
		IND	MEST	AFR	LAM	OHI	OTH	
0.1	0.11	3.5533	0.7112	5.0128	0.75	0.1225	2.6116	
0.1	0.21	3.5533	0.7112	5.0128	0.75	0.1225	2.6116	
0.1	0.5	3.5533	0.7112	5.0128	0.75	0.1227	2.6116	
0.1	1	3.5533	0.7112	5.0128	0.75	0.1229	2.6116	
0.1	2	3.5533	0.7112	5.0128	0.75	0.1233	2.6116	
0.1	4	3.5533	0.7112	5.0128	0.75	0.1241	2.6116	
0.2	0.01	3.3856	0.5962	4.8181	0.6343	0.0184	2.4614	
0.2	0.21	3.3856	0.5962	4.8181	0.6343	0.0185	2.4614	
0.2	0.5	3.3856	0.5962	4.8181	0.6343	0.0186	2.4614	
0.2	1	3.3856	0.5962	4.8181	0.6343	0.0188	2.4614	
0.2	2	3.3856	0.5962	4.8181	0.6343	0.0192	2.4614	
0.2	4	3.3856	0.5962	4.8181	0.6343	0.02	2.4614	

Table A32: Temperature increase in year 2100 for different calibrations of α and β considering extreme cases (total and not-at-all cooperation)

β	α US	Cooperative case (Celsius degrees)	Non-cooperative Case (Celsius degrees)
0.1	0.11	2.210883	3.29798
0.1	0.21	2.210885	3.29798
0.1	0.5	2.210888	3.29798
0.1	1	2.210893	3.297981
0.1	2	2.210903	3.297983
0.1	4	2.210924	3.297987
0.2	0.01	2.217292	3.295185
0.2	0.21	2.217282	3.295185
0.2	0.5	2.217285	3.295186
0.2	1	2.21729	3.295187
0.2	2	2.2173	3.295189
0.2	4	2.21732	3.295193