Efficient, Fair and Stable Agreements for Marine Plastic Pollution

[Latest Version]

Sreoshi Banerjee* Christopher Stapenhurst†

December 3, 2024

Abstract

In 2022, the UN environmental agency resolved to establish an international agreement for abating marine plastic pollution (MPP). We use income and MPP transition data from 120 countries across the globe to derive an efficient abatement policy that maximises collective welfare. This efficient policy is both unfair (poor countries bear the costs while rich countries reap the benefits) and unstable (costs outweigh benefits for some countries). We develop a novel game-theoretic approach to design a fair and stable compensation scheme that redistributes gains from cooperation to the countries who create the most value. We develop new methods to for quantifying the robustness of our results and offer suggestions for policy.

Keywords: Transboundary pollution, marine plastic pollution, nucleolus, m-core, environmental agreements.

JEL codes: C71, D62, H23, Q53, Q58.

---

*Email: banerjee.sreoshi@gtk.bme.hu

†Email: c.stapenhurst@edu.bme.hu

1 Introduction

Marine plastic pollution (MPP) is a global phenomenon with large impacts on marine ecosystems (Benson et al., 2022; Watts et al., 2015; Cole et al., 2011), coastal tourism (Hayati et al., 2020; Jang et al., 2014), the marine industry and fishing activities (McIlgorm et al., 2011); and with significant adverse consequences on human health (Efferth and Paul, 2017; Amato-Lourenço et al., 2021; Ragusa et al., 2022). Carney Almroth and Eggert (2019) and Li et al. (2016) discuss in detail the potential hazards of MPP and the need for government intervention to address this concern. Beaumont et al. (2019) estimate the total economic value of ecological damages caused by MPP to be \$0.5–2.5 trillion per year in 2007 dollars, amounting to approximately 0.6–3.2% of global income. Merkl and Charles (2022) estimate that the legal liabilities from health issues caused by MPP could amount to over \$100bn/year, with a ten fold increase after 2030 due to the development of better legal instruments.

The problems posed by MPP have recently attracted the attention of the intergovernmental organisations such as the UN, WHO and the OECD, as well as national governments, citizens and the scientific community (Zhao et al., 2023; Soós et al., 2022; Carroll et al., 2014; NOAA, 2021). The 14th Sustainable Development goal of UN is to “prevent and significantly reduce marine pollution of all kinds, in particular from land-based activities, including marine debris and nutrient pollution, by 2025”. The UN Environment Assembly has recognised marine litter as a top priority since its first session in 2014 (Chassignet et al., 2021). In 2022, it resolved to produce an international agreement to tackle plastic pollution (UNEP, 2022).

Such an agreement is necessitated by the fact that winds and ocean currents transport plastic pollution from one country’s Exclusionary Economic Zone (EEZ) to anothers’ (Borrelle et al., 2017; Vince and Hardesty, 2018). This creates negative pollution externalities, and a tragedy of the commons in the market for pollution abatement. Global welfare is maximised when each country abates at the level that sets marginal cost equals the sum of all countries’ marginal benefits. But upstream countries do not reap the full benefits of their own abatement efforts, so they under-abate relative to this global optimum. Downstream countries are then exposed to increased emissions from the upstream countries, so they respond by over-abating.

International agreements are crucial for internalizing externalities by prescribing an abatement policy that internalises these externalities, thereby maximizing global welfare. However, this can create imbalances: upstream countries may bear disproportionate abatement burdens, while downstream countries gain more benefits, leading to unfair and unstable agreements. Some countries may benefit by refusing the participate in the agreement and free-riding on the efforts of others to reduce pollution. This hampers incentives to cooperate at the global level. Monetary transfers can help to design fair and stable agreements that compensate countries for any additional abatement costs, and disincentive important players from leaving the agreement (Weikard et al., 2006; Carraro et al., 2006).

We make five contributions to the problem of designing an international agreement for MPP: First, we use global marine litter tracking data to construct a transition matrix describing the current movement of plastics between different 169 countries' EEZs. This matrix is available in the supplementary information.

Second, we use this data together with country level income data to estimate an efficient abatement policy that maximises the collective welfare of 120 countries. We find that it abates 32.2% of global MPP, and generates a surplus equal to 36.7% of the total willingness to pay (WTP) of all countries to abate all MPP. If we assume that the total WTP is equal to the economic damages estimated by Beaumont et al. (2019) then this amounts to $68 billion dollars per annum. The United States emerges as the biggest winner with a 40% share of the total surplus from this policy (roughly $27 billion). More generally, high-income and downstream continents tend to gain the most from the efficient policy both because they experience the largest decreases in MPP, and because their higher WTP means that they value these decreases the most. We refer to these as "demand" continents since they benefit the most from abatement and thus make higher levels of abatement more desirable. Low-income and upstream continents gain the least because they are required to carry out the highest levels of costly abatement, and because they have relatively low WTP for any decreases in MPP that they experience. We refer to these as the "supply" continents because they supply abatement.

Third, we develop new notions of the fairness and stability of an agreement based on both how much each country (or coalition) contributes to a global agreement, and how much each country (or coalition) can obtain by free-riding on the efforts of others. An *agreement* specifies both an abatement policy a system of monetary transfers between countries. Specifically, we develop a new partial equilibrium concept, in which the rest of the world best responds *collectively* to a deviation by a coalition of countries. This concept captures the idea of free riding because deviating coalitions benefit from the increased abatement efforts of non-deviating countries. The *excess* that a coalition gets from a given agreement is the total payoff its members from that agreement net of the surplus that it would get from its partial agreement equilibrium. We say that an agreement is *stable* if every coalition has a positive excess. This ensures that no coalition has an incentive to deviate from the agreement to its partial equilibrium in which it free-rides. We say that an agreement is *fair* if it lexicographically maximises the excess of all coalitions.[1] This definition captures the Rawlsian notion of maximising the welfare of the worst off in society, but here the "worst off" may be either an individual country or a coalition of countries. It also guarantees maximum stability in the sense that a fair agreement is guaranteed to be stable whenever a stable agreement exists.

These definitions coincide with the (pre-)nucleolus (Schmeidler, 1969) and the core (Gillies, 1959) of the characteristic form game in which the worth of a coalition is given by its partial equilibrium payoff. Below, we compare stability concept with γ core of Chander and Tulkens (1995), and the m- and s-cores of Hafalir (2007).

Fourth, we use the available data to calculate the efficient abatement policy for nine continental blocks, identify the monetary transfers required to implement the fair agreement between these continents, and verify that it is stable. We therefore conclude that this agreement is *efficient, fair and stable* (EFS). This agreement redistributes surplus away from demand continents (North America, Europe, and South

---

[1] An agreement lexicographically maximises the excesses if, for all coalitions C it gives a higher payoff to C than all any other agreement that gives the same of higher excess to all coalitions C' that have a lower excess than C. I.e., if another agreement gives a higher excess to C than it must give a lower excess to at least one coalition C' with an excess less than C.

and Southeast Asia) and towards supply continents (Middle America, North and West Asia, and Africa). North America and Europe remain the biggest beneficiaries of the agreement, but the fair transfers make Middle America, North and West Asia, and Africa the third, fourth and fifth greatest beneficiaries. Without transfers, these three supply continents would all be made worse off by the efficient policy. Despite having the highest WTP, South and Southeast Asia does not benefit much from the EFS agreement because it receives less than 10% of its MPP from other continents, and it emits little pollution to other high-income continents. The same is true for South America and Oceania.

Fifth, we develop two new types of sensitivity test to verify the robustness of the EFS agreement. The first of these tests measures the elasticity of each country's actual payoff with respect to each of the model parameters (plastic transport coefficients and WTP estimates). If these elasticities are small then each player's true payoff is likely to be somewhat close to the fair payoffs predicted by the model. The second test is a kind of stability stress test: we calculate the smallest error in our parameter estimates that would result in a coalition getting a negative excess from the EFS agreement. The larger the necessary error, the more confident we can be that the EFS agreement will be stable. The results of these tests show that the EFS agreement is robust. Moreover, we hope that these tests will be useful in future research.

The remainder of the paper is structured as follows. The remainder of the introduction reviews the literature; section 2 defines the model and proves our theoretical results; section 3 describes our data sources, and the construction of the transport matrix; section 4 discusses the main results; section 5 conducts the robustness checks; and section 6 concludes by outlining policy implications and directions for future research.

Related Literature To the best of our knowledge, Beaumont et al. (2023) is the only other article that estimates efficient abatement for MPP. They estimate MPP transition between 16 North Atlantic countries, and develop a model based on Mäler (1989)'s 'Acid Rain game', to calculate the efficient abatement policy for these countries. They find that the gains from international cooperation can be substantial. But their estimated efficient abatement policy leaves the three lowest-income countries worse off than they would be without an agreement. They solve this problem by proposing a constrained efficient policy that requires every country to be at least as well off as it is under the status quo. However, they find that this constraint leads to large reductions in the gains from cooperation.

Chander and Tulkens (1995) are the first to use partial agreement equilibria to define a characteristic form game. They refer to the core of this game as the γ-*core* and propose the use of monetary transfers to produce outcomes in the γ-core. But their notion of partial equilibrium assumes that the rest of the world choose abatement policies that maximise their *individual* payoffs given the policies of the deviating coalition. We see two drawbacks of their assumption that the rest of the world responds to a deviation by maximising their individual payoffs. First, we do not observe this in practice. When the US withdrew from the Paris Agreement in 2017, the rest of the world did not nullify the agreement and go their separate ways. Rather, they continued to act collectively under the Paris Agreement.

Second, their assumption does not adequately model free-riding since it does not allow deviating countries to benefit from the increased abatement efforts brought about by continued cooperation between countries remaining in the agreement. If the rest of the world continues to cooperate, then it is likely to carry out more abatement than if cooperation ceases and each country acts in its own interests. This higher level of abatement increases the payoff of the deviating coalition, and therefore makes deviation relatively more attractive than Chander and Tulkens (1995) implicitly assume. Our definition of partial-equilibria assumes that non-deviating countries do continue to cooperative and therefore tends to produce a more robust notion of stability. For uniformly mixing pollutants, we show that every agreement that is stable according to our definition is also stable according to theirs, but not every agreement that is stable according to their definition is stable according to ours. This is not necessarily true for non-uniformly mixing pollutants, such as MPP, but it hold continue to hold when mixing is "uniform enough".

Both Chander and Tulkens' game and ours are closely related to a pair of games defined by Hafalir (2007). Whilst our starting points are non-cooperative games between coalitions, Hafalir's is a partition function form game, specifying the payoffs of each coalition in a given partition of players. She defines the *m-core* (core with merging expectations) to be the set of payoff vectors which give every coalition a greater payoff than it gets in the partition where all other players form the complementary coalition. She defines the *s-core* (core with separating expectations) to be the set of payoff vectors which give every coalition a greater payoff than it gets in the partition where all other players form single coalitions. Thus Chander and Tulkens' γ-core stability captures the same intuition as her s-core, whereas our stability concept captures the same intuition as her m-core.

The (pre-)nucleolus has been used in other literature on environmental agreements, e.g. Van Steenberghe (2004) uses it to study Core-stable and equitable allocations of greenhouse gas emission permits, and Lejano and Li (2019) uses the proportional nucleolus to study international carbon reduction agreements. The literature has also proposed other methods for distributing the value of the grand coalition, including the Shapley value (Shapley, 1952; Winter, 2002; Fernandez, 2009), the equal division rule (Moulin, 1987), and McGinty (2011)'s division rule. But unlike the nucleolus, these allocations are not guaranteed to be stable.

Zhao et al. (2023) study fair and stable MPP abatement technology sharing agreements between China, Indonesia, and Malaysia. They estimate both the value of a grand coalition between all three countries and the values of three sub-coalitions containing only two of the three countries. They use these coalitional values to define a cooperative game and calculate monetary transfers that implement the Shapley Value (a well-known fairness concept) of the game. They then show that the Shapley Value is a core allocation (a well-known stability concept) of the game, and hence conclude that their agreement is both fair and stable. But their analysis assumes that the countries abate *all* of their pollution and ignores the transboundary nature of MPP. Not only is this likely to lead to an inefficient outcome, but it also precludes any strategic decisions that the countries face about how much MPP to abate (question 2).

2 Model and Theoretical Results

2.1 The Abatement Game

Consider a set of players $N = \{1, 2, \ldots, n\}$ where a generic player is denoted by $i \in N$ and a generic subset of players, or coalition, is denoted by $S \subseteq N$. A player can either be a country or a group of countries (e.g. European Union). Each player i chooses to abate a fraction, $a_i \in [\underline{a}, 1]$, of its status quo MPP emissions,[2] where $\underline{a} \leq 0$ is a lower bound on the faction of abatement that a player can perform. If $a_i = 0$, then player i continues its status quo emissions; if $a_i < 0$, then player i increases its emissions relative to the status quo. We can interpret $\underline{a}$ as an upper bound on the quantity of MPP that a player emits. A *global abatement policy* is a vector $a \coloneqq (a_i)_{i \in N}$ which describes the abatement choice of each player. A *coalitional abatement policy* is a vector $a_C \coloneqq (a_i)_{i \in C}$ stating the abatement choices made by the members of the coalition $C \subset N$.[3] We denote the set of complementary players of C by $N \setminus C$.

Player i's cost of abating a_i is given by a cost function $c_i(a_i) \coloneqq -\gamma_i \ln(1 - a_i)$, where $\gamma_i \geq 0$ is i's cost parameter. This class of functions satisfies the following desirable properties: (i) there is no cost of maintaining status quo emissions ($c_i(0) = 0$); (ii) the cost and the marginal cost are increasing in abatement ($c_i' > 0$ and $c_i'' > 0$); (iii) the cost and the marginal cost of full abatement are infinite ($\lim_{a_i \to 1} c_i(a_i) = \lim_{a_i \to 1} c_i'(a_i) = \infty$). Property (iii) guarantees interior solutions. This functional form also has the convenient feature that the marginal cost of abatement at the status quo level is equal to the cost parameter ($c_i'(0) = \gamma_i$).

Transboundary externalities are described by an $N \times N$ matrix T whose i, jth entry denotes the quantity of plastic transitioning from player i' EEZ to player j's EEZ in a year. The total MPP emissions of country i are given by $e_i \coloneqq \sum_{k \in N} T_{ik}$. In our model, what matters for each player i is not the quantity of plastic that it receives from each country, rather the fraction of MPP in its EEZ that originated from each

---

[2] "Status quo" meaning a country's current emission level.

[3] We use the notation ":=" to mean that the symbol on the left is defined by the expression on the right.

other player j. We therefore define the *backward transition matrix*, $\overleftarrow{T}$ (Brouard, 2019) to have entries $\overleftarrow{T}_{ij} \coloneqq T_{ij}/\sum_{k\in N} T_{kj}$. The jth column of the matrix is a probability distribution of MPP origins, whose elements sum to one. The use of a backward transition matrix both simplifies notation and eliminates the need for estimating emission quantities to achieve our main results.

Similarly, the *forward transition matrix* $\overrightarrow{T}$, defined by $\overrightarrow{T}_{ij} \coloneqq\coloneqq T_{ij}/e_i$, describes the fraction of player i's MPP that transitions to country j. The i'th row of the forward transition matrix describes the distribution of destinations of i's MPP emissions. If emissions from each country are equally distributed across all countries, then $\overrightarrow{T}_{ij} = 1/n$ and we say that the pollutant is *uniformly mixing.* In this special case, the columns of $\overleftarrow{T}$ are all identical and equal to the vector of emissions quantities $e \coloneqq (e_i)_{i\in N}$.

If players commit to a global policy a, then the amount of MPP transitioning to player j falls by a fraction $R_j(a) \coloneqq \sum_{i\in N} a_i \overleftarrow{T}_{ij}$.[4] The fraction of MPP abated from all players' EEZs is $\sum_{j\in N}\sum_{i\in N} a_i T_{ij}/\sum_{j\in N}\sum_{i\in N} T_{ij}$.

Player i's WTP to have a fraction R_i of MPP abated from its EEZ is described by a benefit function $b_i(R_i) \coloneqq \beta_i R_i\,(2 - R_i)\,$, where $\beta_i \geq 0$ is player i's WTP (or "benefit") parameter. Quadratic benefits are popular in previous literature (Alvarado-Quesada and Weikard, 2017). This particular quadratic functional form is constructed to have the following desirable properties: (i) players have no WTP for no abatement ($b_i(0) = 0$); (ii) WTP is increasing in the fraction of abatement received ($b_i'(R_i) > 0$); (iii) there are diminishing marginal returns to abatement ($b_i''(R_i) < 0$); (iv) the marginal WTP for full abatement is zero ($b_i'(1) = 0$). It also has the convenient property that the benefit parameter is equal to both half of the marginal WTP for no abatement ($\beta_i = b_i'(0)/2$), and the WTP for full abatement ($\beta_i = b_i(1)$).

The *total economic damages* caused by MPP is equal to the total WTP of all countries to abate all MPP. We denote it by $B \coloneqq \sum_{i\in N}\beta_i$.

Player i's *monetary value* for abatement policy a for player $i \in N$ is given by $b_i(R_i(a)) - c_i(a_i)$. Our main unit of analysis is player i's *normalised value* of policy a,

---

[4] Under policy a, $(1 - a_i)T_{ij}$ units of MPP transition from i to j, so the quantity of MPP transition to j falls by $\sum_{i\in N}(T_{ij} - (1 - a_i)T_{ij})/\sum_{i\in N} T_{ij} = \sum_{i\in N} a_i T_{ij}/\sum_{i\in N} T_{ij}$.

which expresses values as a fraction of the total economic damages, B. It is defined by

$$v_i(a) = (b_i(R_i(a)) - c_i(a_i))/B. \tag{1}$$

The normalised value yields the same optimal policies as the monetary value, but with the advantage that it makes our main results less sensitive to potential measurement error of the level of the β_i estimates. we denote the vector of values of players in a coalition C by $v_C(a) \coloneqq (v_i)_{i \in C}$.

The value of a policy a for a coalition $C \subseteq N$ is given by the sum of its members' values, $V_C(a) \coloneqq \sum_{i \in C} v_i(a)$. The *global value* of a is $V_N(a)$.

2.2 Equilibrium Abatement policies

Our notions of fairness and stability are based on the value that a given coalition of countries can receive if they collectively deviate from a global agreement. It seems natural to assume that the deviating coalition will choose to maximise the collective payoffs of its members, but what about non-deviating players? The Chander and Tulkens (1995) assume that non-deviating players give up on the agreement altogether and best respond unilaterally, rather than as a coalition. This seems natural if negotiations start from a disagreement equilibrium, or if there is no coordinated attempt to produce a global agreement. But if there is an existing agreement, or some expectation that one will be formed, then it seems more likely that non-deviating countries would choose to continue negotiating an agreement that both internalised externalities between members of the coalition, whilst best responding to the strategy of the deviating players. Moreover, non-deviating countries will tend to do more abatement when they act collectively than individually, thereby allowing the deviating coalition to free-ride on their increased abatement efforts. This kind of free riding cannot be captured if non-deviating do not internalise their mutual externalities. We therefore define coalition C's *merge* partial-equilibrium (or m-equilibrium) to be the Nash equilibium of the two-player game played between the players in a deviating coalition C and the non-deviating players outside of C. In these equilibria, members of C collectively choose a policy a_C that maximises their combined value given the policy of the players in $N \setminus C$.

Together, each $\langle \overleftarrow{T}, \beta, \gamma \rangle$ defines a family of *pollution games*. The tuple $\langle N, a, v; \overleftarrow{T}, \beta, \gamma \rangle$ defines a non-cooperative game between all n players. For a each coalition $C \subseteq N$, the tuple $\langle \{C, N \setminus C\}, (a_C, a_{N\setminus C}), (v_C, v_{N\setminus C}); \overleftarrow{T}, \beta, \gamma \rangle$ defines a two player game between C and $N \setminus C$.

Definition 1 (Equilibrium policies)**.**

1. A *disagreement equilibrium* policy, a^D, is a Nash equilibrium of the game played between all n-players.

2. The *globally efficient policy*, a^* is the policy that maximises the global value.

3. A *merge-partial equilibrium policy* for coalition C (or *m-equilibrium for* C), a^C, is defined by

$$a^C_C \in \arg\max_{a_C} V_C(a_C; a^C_{N\setminus C}) \tag{2}$$

$$a^C_{N\setminus C} \in \arg\max_{a_{N\setminus C}} V_C(a_{N\setminus C}; a^C_C). \tag{3}$$

The Chander and Tulkens (1995) partial equilibrium concept is analogous, but instead of assuming that complementary players in $N \setminus C$ best respond collectively by maximising their joint value, they assume that complementary players best respond individually by choosing $a_i \in \arg\max_{a_i} V_C(a_i; a^C_{-i})$. We refer to it as a separating-partial equilibrium with respect to C, or simply an s-equilibrium.

The remainder of this subsection presents results that characterise and compare the behaviour of coalitions in different kinds of equilibria.

Lemma 1 (Equilibrium Existence)**.**

1. *A disagreement policy exists and satisfies the following first order conditions (FOCs):*

$$\gamma_i/(1 - a^D_i) = 2\beta_i \overleftarrow{T}_{ii}(1 - R_i(a^D)) \qquad \forall i \in N. \tag{4}$$

2. *The globally efficient policy exists, is unique, and satisfies*

$$\gamma_i/(1-a_i^*) = 2\sum_{j\in N}\beta_j \overleftarrow{T}_{ij}(1-R_j(a^*)) \qquad \forall i \in N.$$

3. *For every coalition $C \subset N$, an m-equilibrium policy for C exists and satisfies*

$$\gamma_i/(1-a_i^C) = 2\sum_{j\in C}\beta_j \overleftarrow{T}_{ij}(1-R_j(a^C)) \qquad \forall i \in C \tag{5}$$

$$\gamma_i/(1-a_i^C) = 2\sum_{j\in N\setminus C}\beta_j \overleftarrow{T}_{ij}(1-R_j(a^C)) \qquad \forall i \in N\setminus C. \tag{6}$$

Proof.

1. Strict concavity of the benefit functions and strict convexity of the cost functions implies that $v_i(a)$ is strictly concave in a. The choice set $[-\underline{a}, 1]$ is a convex and compact subset of the Euclidean space $\mathbb{R}^n$. Theorem 1 of Rosen (1965) then implies that the game between the players in N has an equilibrium characterised by FOCs.

2. Theorem 2 of Luenberger et al. (1984) implies that the function $v_N(\cdot)$ has a unique global maximum whenever the FOCs are satisfied.

3. For all $C \subseteq N$ the function $v_C(a)$ is strictly concave in a. The conditions of the Rosen's Existence theorem are satisfied for

$$\langle\{C, N\setminus C\}, (a_C, a_{N\setminus C}), (V_C, V_{N\setminus C}); \overleftarrow{T}, \beta, \gamma\rangle.$$

$\square$

The FOC for players in the disagreement equilibrium equates that player's marginal cost of abatement (the left side of Equation 4) with its marginal benefit of abatement *carried out by itself* (the right side). The qualifier "carried out by itself" is important because each player can only impact its received MPP in proportion to its fraction of *self-pollution*, $\overleftarrow{T}_{ii}$.

By contrast, the FOC for a player in C equates its marginal cost with the sum of the marginal benefits of countries in C (the right side of Equation 5). This is exactly what it means to "internalise player i's externality on player j". The benefit of a fraction of abatement by player i on player j in proportion to $\overleftarrow{T}_{ij}$, which quantifies the size of the externality that i has on j. Similarly, in m-equilibria, the FOC for a player in $N \setminus C$ equates its marginal cost with the sum of the marginal benefits of countries in $N \setminus C$ (the right side of Equation 6).

Proposition 1. *Let $C \subseteq N$ and a^C be an m-equilibrium policy with respect to C. In a pollution game with concave benefits and convex costs, at least one of the coalitions, C or $N \setminus C$, attains a higher welfare in the m-equilibrium a^C than in the disagreement equilibrium a^D.*

Proof. A coalition must benefit from changes in the abatement policy of its own members in a^C relative to a^D, but it may suffer from changes in the abatement policies of countries in the complementary coalition. There are two cases to consider. If coalition C abates more in a^C than in a^D, then the $N \setminus C$ must get higher welfare from a^C than from a^D. If coalition C abates less in a^C than in a^D, then it must be best responding to an increase in the abatement of $N \setminus C$. Hence C must get higher welfare from a^C than from a^D. □

Uniform Mixing

In the uniform mixing case, Chander and Tulkens (1995) prove that coalition formation has positive spillovers on complementary countries: they abate less and have higher welfare in the γ-partial equilibrium, than in the disagreement equilibrium. However, this turns out not to be true for m-partial equilibria. Example 1 shows that in the m-partial equilibrium of a game with uniform mixing,

1. complementary countries may abate more and have lower welfare,
2. coalition countries may abate less,
3. coalition countries may have lower welfare,

than in the disagreement equilibrium.

Example 1. Negative spillovers and Free-riding (uniform mixing). Consider three countries, A, B, and C who benefit equally from full abatement, so that $\beta_A = \beta_B = \beta_C = 1$. In the disagreement equilibrium, countries A and B contribute 10% of total emissions, country C contributes the other 80%. Mixing is uniform, so all the columns of T equal (0.10.10.8). Countries A and B have much lower abatement costs than C. To be precise, $\gamma_A = \gamma_B = 0.2$, and $\gamma_C = 1.6$. The numbers are chosen so that $a^D = 0$ (no country does any abatement in the disagreement equilibrium).

The m-partial equilibrium for coalitions $\{A, B\}$ and $\{C\}$ has abatement profile

$$a^{\{A,B\}} = a^{\{C\}} = (0.47, 0.47, -0.05)$$

(the two coalitions are complementary so they induce the same m-partial equilibrium). Members of $\{A, B\}$ internalise their mutual externalities by abating more than in the disagreement equilibrium. Country C *free rides* on A and B's increased abatement efforts. C receives less pollution from A and B, so it marginal benefit of abatement falls below its marginal cost in the disagreement equilibrium. Country C brings its marginal cost in line with the coalition's marginal benefit by *decreasing* its abatement effort.

Whilst $\{A, B\}$ benefit from internalising their mutual externalities, they suffer from C's increased abatement. Since C is such a big polluter, the latter effect dominates and $\{A, B\}$ is worse off overall. Indeed, $v_{\{A,B\}}(a^D) = 0$ in the disagreement equilibrium, whilst $v_{\{A,B\}}(a^{\{C\}} = -0.02$ in the m-partial equilibrium. Thus, with the m-partial equilibrium concept, the formation of the $\{C\}$ coalition has *negative spillovers* on the A and B because it pushes them into a coalition, which prompts $\{C\}$ to decrease abatement.[5] The γ-partial equilibrium concept cannot capture this phenomenon because it precludes the formation of the $\{A, B\}$ coalition.

Policy implication 1. No agreement may be better than an agreement that ex-

---

[5] An example with a singleton coalition formation may seem contrived, but it is easy to produce examples with non-trivial coalitions by adding dummy countries with low WTP and low status quo emissions.

cludes big emitters.

Chander and Tulkens (1995) prove that, under uniform mixing, total abatement is higher in the γ-partial equilibrium relative to the disagreement equilibrium.

Proposition 2. *The deviating coalition is better off in the m-partial equilibrium of a uniform-mixing game than in the γ-partial equilibrium.*

- total abatement is higher in the m-equilibrium
- the deviating coalition does less abatement in the m-equilibrium

Non-uniform Mixing

We get even more striking results when we depart from uniform mixing: Folmer and von Mouche (2015) prove 1 by way of example.[6]

Example 2. Total abatement decreases (non-uniform mixing). Three countries, A, B, and C, all benefit equally from full abatement and have identical cost parameters. Specifically, $\beta_A = \beta_B = \beta_C = 1$ and $\gamma_A = \gamma_B = \gamma_C = 1/6$. In the status quo, A and B each emit 25% of total MPP, whilst C emits 50%. Backward transition probabilities are described by the following matrix:

$$\overleftarrow{T} = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{7}{16} \\ \frac{1}{8} & \frac{1}{8} & \frac{7}{16} \\ \frac{3}{4} & \frac{3}{4} & \frac{1}{8} \end{bmatrix}$$

Thus, A and B pollute C, but exchange little pollution with each other; C sends a large proportion of pollution to A and B, but does not pollute itself very much. In the disagreement equilibrium, all countries continue to emit at the status quo level, i.e. $a_A^D = a_B^D = a_C^D = 0$.

If A and B form a coalition, then they internalise their small mutual externalities by increasing their abatement. This has a large positive spillover on country C,

---

[6] In practice, we find no evidence of multiple equilibria in any of our numerical calculations.

which best responds by increasing its emissions, producing a large negative spillover on A and B. The $\{A, B\}$ coalition responds by increasing their abatement efforts further. In both the m-partial equilibria with respect to $\{A, B\}$, we have $a_A^{\{A,B\}} = a_B^{\{A,B\}} \approx 67\%$ and $a_C^{\{A,B\}} \approx -90\%$. The net result is that emissions increase from $2 \times 8 + 16 = 24$ tons to $2 \times (1 - 0.67) \times 8 + (1 + 0.9) \times 16 \approx 35$ ton. Equivalently, total abatement decreases by roughly $35/24 - 1 \approx 46\%$.

Policy implication 2. No agreement may be better than an agreement that excludes countries that exchange large amounts of MPP with member countries.

Example 3. Negative spillovers (non-uniform mixing)[label=ex:river] Three countries, A, B, and C, all value abatement equally. Their costs are given by $\gamma_A = 2$, $\gamma_B = \gamma_C = 0.2$. They are situated along a river with backward transition matrix:

$$\overleftarrow{T} = \begin{bmatrix} 1 & 0.9 & 0 \\ 0 & 0.1 & 0.9 \\ 0 & 0 & 0.1 \end{bmatrix}.$$

In the disagreement equilibrium, all countries continue to emit at the status quo level, and get a value of 0 i.e. $a_i^D = v_i(a^D) = 0$ for $i = 1, 2, 3$. If upstream countries A and B form a coalition then A increases its abatement effort to internalise the externality it exerts on B. Since it receives less MPP from A relative to the disagreement equilibrium policy, country B reduces its abatement effort to save on costs. Thus, in the m-partial equilibrium with respect to $\{A, B\}$, we have $a_A^{\{A,B\}} \approx 29\%$ and $a_B^{\{A,B\}} \approx -30\%$. Country C is unaffected by A's increased abatement, but suffers an increase in emissions from B. Its value in the partial equilibrium is therefore lower than in the disagreement equilibrium. To be precise, $v_C(a^{\{A,B\}}) = -0.6$.

Policy implication 3. Downstream countries cannot always free-ride on agreements between upstream countries.

Together, these examples prove the following proposition:

Proposition 3. *There are pollution games with concave benefits and convex costs in which:*

1. *m- and disagreement equilibria are not unique;*

2. *total abatement in the m-equilibrium is lower than in the disagreement equilibrium;*

3. *(negative spillovers) complementary countries may abate more and have lower welfare in the m-equilibrium than in the disagreement equilibrium;*

4. *a coalition gets a higher payoff in its m-equilibrium than in its γ-equilibrium.*

Whilst examples 1–**??** are contrived to illustrate the mechanics of m-equilibria, Section 4.3 gives empirical examples which demonstrate that free riding and negative abatement are also practical concerns.

2.3 Efficient, Fair and Stable Agreements

To produce fair and stable agreements, players must compensate each other using monetary transfers. We, therefore, define an *agreement* to be a pair (a, t) that specifies a global abatement policy a and a vector of monetary transfers $t := (t_i)_{i \in N}$ for each player, satisfying budget balance $\sum_{i \in N} t_i = 0$. Player i's payoff from an agreement (a, t) is given by $u_i(a, t) = t_i + v_i(a)$. We denote the vector of payoffs by $u(a, t) := (u_i(a, t))_{i \in N}$. Let $t_C = \sum_{i \in C} t_i$ and $u_C(\cdot) = \sum_{i \in C} u_i(\cdot)$. The *excess* (or *satisfaction*) that coalition C gets from agreement (a, t) is its payoff from (a, t) net of its m-equilibrium payoff:

$$\begin{aligned} \mathrm{ex}_C(a, t) &= u_C(a, t) - u_C(a^C, 0) && (7) \\ &= t_C + V_C(a) - V_C(a^C). && (8) \end{aligned}$$

If the excess is large and positive, then C gets a much better payoff from the agreement then it could get from deviating to its m-equilbrium. If it is negative, then C can potentially increase its payoff by committing to play its m-equilibrium policy a^C_C. If the remaining players $N \setminus C$ remain in a coalition, then their best response, $a^C_{N \setminus C}$. This ensures that the players in C collectively get $V_C(a^C)$. The excess therefore describes how satisfied each coalition is with their agreed allocation of the surplus.

Definition 2 (Efficient, fair, and stable agreements)**.** An agreement (a, t) is

- *efficient* if a is a globally efficient policy;
- *fair* if t lexicographically maximises the satisfaction of the least satisfied coalitions, i.e. for all other budget-balanced $t' \in \mathbb{R}^n$ with $\mathrm{ex}_C(a, t') > \mathrm{ex}_C(a, t)$ for some coalition C, there exists another coalition C' with

$$\mathrm{ex}_{C'}(a, t') < \mathrm{ex}_{C'}(a, t) \leq \mathrm{ex}_C(a, t).$$

- *stable* if every coalition is satisfied (has a positive satisfaction), i.e. $\mathrm{ex}_C(a, t) \geq 0$ for all $C \subseteq N$.

Our definition of stability is completely analogous to the γ-core stability of Chander and Tulkens (1995). They use the payoffs that each coalition C gets from their γ-partial equilibrium, $V_C(\tilde{a}^C)$, to define a characteristic form game $\langle N, (V_C(\tilde{a}^C))_{C \subseteq N} \rangle$. They say that an agreement (a, t) is stable if $u(a, t)$ is in the core of this game. They refer to the set of core stable agreements as the γ-core. We simply replace γ-partial equilibrium with m-equilibrium, and define fair agreements to those in the pre-nucleolus of the game.

The only difference between their definition and ours is that they their *partial equilibrium with respect to* C, $\tilde{a}^C$, differs from our m-equilibrium for C, denoted a_C. $\tilde{a}^C$ is defined to be a Nash equilibrium of the $|N \setminus C| + 1$ player pollution game between C and the remaining $|N \setminus C|$ individual players. By contrast, we define our m-equilibrium for C to be the Nash equilibrium of the two player game between Not only do we find this assumption more realistic, but we also find in Section 5.3 that it leads to more stable agreements. agreements which are obtained using the m-equilibrium notion are stable under both the m- and Chander and Tulkens (1995, 1997) equilibrium notions, but agreements which are obtained using the γ-equilibrium notion are not. Say that an agreement is m-core stable if it is stable according to 2.

Corollary 1.

1. *An efficient and fair agreement exists and is generically unique.*
2. *Efficient and stable agreements may or may not exist.*
3. *The efficient and fair agreement is stable if and only if an efficient and stable agreement exists.*
4. *If the transition matrix* $\overleftarrow{T}$ *is uniformly mixing, then every agreement that is m-core stable is also* γ*-core stable, but not every* γ*-core stable agreement is m-core stable.*

Proof.

The first statement follows immediately from standard results on the pre-nucleolus (Maschler, 1992).

The second statement is proved by example. Our main result in section 4 gives an example of an efficient and stable agreement. Example 5 gives an example of a pollution game with no stable agreements.

Example 4.

The third statement is a standard result. If the pre-nucleolus is in the core then it is an example of a payoff vector is core stable. Otherwise, if the pre-nucleolus is not core stable, then it must give the most dissatisfied coalition a lower payoff than their coalitional value. Every other payoff vector must produce at least one coalition with a greater level of dissatisfaction than the most dissatisfied coalition under the nucleolus. Therefore every other payoff vector must produce at least one coalition with a lower payoff than its coalition value. Hence no other payoff vector lies in the core.

The fourth statement is a corollary of 2. Since coalitions get a higher payoff in their m-equilibrium Smaller coalitions internalise fewer externalities and therefore abate less pollution than larger coalitions. Every coalition's value is increasing in the amount of abatement carried out by complementary players, so the coalitional values are lower when complementary players respond unilaterally.

□

Example 5. Negative spillovers (non-uniform mixing)[continues=ex:river] The first part of the example calculated that $a^D = (0,0,0)$ for all $i \in N$, and that $a^{\{A,B\}} = (0.3, -0.3, 0.2)$ is the m-equlbrium for $\{A, B\}$. The coalitional values are given by $V_{\{A,B\}}(a^{\{A,B\}}) = 0.3$ and $V_{\{C\}}(a^{\{A,B\}} = -0.6$. The unique m-equilibrium for downstream players $\{B, C\}$ is $a^{\{B,C\}} = (0, 0.8, -1.4)$, which yields values $V_{\{A\}}(a^{\{B,C\}}) = 0$ and $V_{\{B,C\}}(a^{\{B,C\}}) = 0.8$. Players A and C exchange no externalities, so their m-equilibrium is the same as the disagreement equilibrium $a^{\{A,C\}} = (0,0,0)$, with values $V_{\{A,C\}}(a^{\{A,C\}}) = V_{\{B\}}(a^{\{A,C\}}) = 0$.

Finally, if all three players form a coalition then the unique equilibrium has $a^* = (-1.3, 0.8, 0.3)$. Player 3 internalised its externality on 2; 2 internalises its externality on 1; and 1 reduces its costs by increasing its emissions. The payoffs are given by $v(a^*) = (1, 0.2, -0.2)$, so $V_N(a^\gamma(N)) = V_N(a^m(N)) = 1$.

Now consider the allocation $x = (-0.1, 1, 0.1)$. The payoffs sum to 1, so it is efficient. However $-0.1 > V_{\{1\}}(a^m(\{2,3\})) = -0.6$, so it is individually rational for player 1 in the m-CFG. Similarly, $1 > V_{\{2\}}(a^m(\{1,3\})) = 0$ and $0.1 > V_{\{3\}}(a^m(\{1,2\})) = 0$, so it is individually rational for all players. Finally, $x_{\{1,2\}} = 0.9 > V_{\{1,2\}}(a^m(\{1,2\})) = 0.8$, $x_{\{1,3\}} = 0 = V_{\{1,3\}}(a^m(\{1,3\}))$ and $x_{\{2,3\}} = 1.1 > V_{\{2,3\}}(a^m(\{2,3\})) = 0.3$, so it is in the m-core. It gives player 1 a negative payoff, which is lower than $v_{\{1\}}(a^\gamma(\{1\}) = 0$, so it cannot be in the γ-core,

[7] The key point is that players 2 and 3 have a negative externality on player 1 when they form a coalition, because player 3's increase in abatement causes player 2 to decrease abatement. player 1 is therefore worse off in the m-CFG, so it is easier to keep player 1 in the grand coalition.

In our empirical application, we find that the m-nucleolus *does* lies in the γ-core, but that the γ-nucleolus *does not* lie in the m-core.

---

[7] All inequalities hold strictly with the unrounded figures.

3 Data

3.1 Transition probabilities

We obtain the backward probability matrix from the dataset of Chassignet et al. (2021).[8] They use particle tracking simulation to model the movements of mismanaged plastic waste (MPW; defined as "plastic material littered, illdisposed, or from uncontrolled landfills" p.2; Soós et al. (2022) estimate that this accounts for 83% of MPP (p.31)) through the world's oceans over the ten year period from 2010 to 2019. They seed their model with estimates of the quantities of buoyant land-based macro-plastics emissions for 145 countries and territories taken from previous literature — Lebreton and Andrady (2019) for coastal regions (3t globally per year) and Lebreton et al. (2017) for rivers (0.9t globally per year). Each particle emitted by a country i is then tracked until it either decays (at an annual rate of 18%) or beaches on the shores of some country j. The authors then count the total quantity of emissions from each country i that beach on the shores of each country j over the ten year period. They find that around 75% of non-decayed MPW beaches in one of 165 countries or territories during the 10-year period, whilst the other 25% remains at sea, for instance in subtropical gyres.

We first aggregate the emissions sent and received by non-sovereign territories with their sovereign states.[9] Doing so leaves us with 169 states that either emit or receive MPW. For a given set of players N we obtain each $\overleftarrow{T}_{ij}$ by dividing the quantity of emissions that player i receives from player j by the total quantity of emissions player i receives from all 169 countries. The full matrix T for 169 countries is available in the supplementary material. Table 1 shows the backward transition probabilities when the player set N is given by nine regional blocks: Africa (Af), East & Southeast Asia (ESAs), Europe (Eu), Middle America (MAm), North America

---

[8] The raw data is available in .pdf or .csv formats from the following webpage: `https://www.frontiersin.org/articles/10.3389/fmars.2021.667591/full#supplementary-material` (accessed 31/05/2023).

[9] French southern and antarctic lands, French Polynesia, New Caledonia, St. Martin (French part), Faroe Islands, Greenland, Falkland islands, Guernsey, Jersey, British Virgin Islands, Isle of Man.

(NAm), North & West Asia (NWAs), Oceania (Oc), South America (SAm), South Asia (SAs). The membership of these blocks is given in B. We present the EFS agreement for these nine regional blocks in section 4.2.

Table 1: Regional blocks backward transition matrix. All figures are given in percentages rounded to one decimal place.

	NAm	MAm	SAm	Eu	Af	SAs	ESAs	NWAs	Oc
NAm	12.4	0.5	0.0	1.8	0.0	0.0	0.0	0.0	0.0
MAm	58.6	64.2	3.4	4.7	0.0	0.0	0.1	0.0	1.5
SAm	3.4	13.6	94.0	0.2	0.2	0.0	0.0	0.0	5.1
Eu	0.1	0.0	0.0	27.8	1.9	0.0	0.0	6.7	0.0
Af	4.2	2.3	0.4	27.6	80.1	2.4	0.4	16.7	5.5
SAs	0.0	0.0	0.0	0.0	1.7	84.2	6.8	1.1	0.2
ESAs	6.1	0.0	0.0	0.0	7.5	13.0	91.0	1.7	14.5
NWAs	0.8	0.0	0.0	37.0	7.1	0.0	0.0	62.8	0.0
Oc	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	72.7

3.2 Costs

We utilise the "status quo equilibrium assumption" from Mäler (1989) and Beaumont et al. (2023) that the status quo abatement choices constitute a disagreement equilibrium. This implies that $a^D = 0$, so the FOC (Equation 4) characterising the disagreement equilibrium reduces to

$$\gamma_i = 2\beta_i \overleftarrow{T}_{ii} \tag{9}$$

for all $i \in N$, whence cost estimates can easily be inferred from benefit and backward transition probability estimates. Note, equation (9) implies that i's cost parameter is directly proportional to its benefit parameter and its level of self pollution.

3.3 Benefits

Another advantage of the status quo equilibrium assumption is that equation (9) implies that each benefit parameter β_i appears on both sides of the FOCs characterising the efficient and equilibrium global policies. For instance, a^* is characterised by the FOCs $\beta_i \overleftarrow{T}_{ii}/(1-a_i^*) = \sum_{j\in N} \beta_j (1 - a^* \cdot \overleftarrow{T}_{\cdot j}) \overleftarrow{T}_{ij}$ for all $i \in N$. It follows that these global policies do not depend on the levels the benefit parameters β_i, but only on their relative values, β_i/B. Furthermore, the normalised values, v_i, are also independent of the levels of the benefit parameters. Thus, we only require estimates of countries' "relative" benefits from abatement to obtain and state our main results.

We therefore adopt the same method as Beaumont et al. (2023) for estimating countries relative benefits from MPP abatement. They take a benefit transfer approach which effectively assumes that each country's WTP is proportional to their national income, up to an elasticity parameter ϵ. Specifically, $\beta_i/y_i^\epsilon = \beta_j/y_j^\epsilon$ for any pair of countries i and j with national incomes y_i and y_j. This implies that

$$\beta_i/B = y_i^\epsilon / \sum_{j\in N} y_j^\epsilon. \tag{10}$$

Our main results use the assumption that $\epsilon = 1$. This is motivated by Czajkowski et al. (2017)'s finding that $\epsilon = 1$ gives a reasonable estimate of benefit transfers between households in different countries.[10]

We explore the consequences of using different values of ϵ in section 5.1.2. We use The World Bank (2023) measures of Purchasing Power Parity (PPP) adjusted Gross National Income (GNI) as estimates of national income. The World Bank provides GNI data for 120 of the 169 countries included in the Chassignet et al. (2021) dataset. Of the excluded countries, the biggest emitters are Venezuela (261k tons over ten years), Taiwan (135k tons) and Comoros (41k tons). Together they emit 528k tons over ten years, which is around 3.2% of global emissions.

3.4 Software

We solve FOCs numerically with *Mathematica*'s *FindRoot* command, which uses the Newton–Raphson method (Wolfram Research, Inc., 2020). We use the *TUGames*

---

[10] Applying the benefit transfer method at the household level yields the estimate $\beta_i = h_i^{1-\epsilon} y_i^\epsilon B$, where h_i is the number of households in country i. The national level method yields $\beta_i = y_i^\epsilon B$. These are the same (similar) when WTP is (close to being) unit elastic with respect to income.

package to calculate the pre-nucleolus and check the core membership of the fair agreement allocation (Meinhardt, 2022).

4 Results

4.1 The Globally Efficient Policy

We find that the globally efficient policy for the 120 countries included in the sample abates 32.2% of all MPP. The breakdown of how much abatement each individual country should carry out, and how much abatement each country receives is given in table A1 in appendix A. We observe the following patterns: (1) Low WTP countries do more abatement. This is because countries with lower marginal WTP have lower marginal costs in the status quo through equation (9). For these countries, abatement is cheaper than for those with higher marginal WTP (and hence, higher costs). Further, these countries tend to emit more in the status quo, so each additional fraction abated corresponds to a larger amount of abatement in absolute terms. (2) Countries who transport a larger fraction of their MPP to other countries are prescribed to do more abatement. This is because they have less incentive to abate in the absence of an agreement, and hence are required to do relatively more under the agreement. (3) Countries who transport MPP to higher WTP countries do more abatement than those who transport to lower WTP ones. This is because abatement by these countries is more valuable. (4) Some countries are required to do negative abatement (i.e. to increase their emissions). This is because the efficient agreement allows them to outsource abatement to countries who can do it more cheaply and/or have a larger impact on the amount of MPP transported to those countries. Therefore they can save on their costs by reducing their abatement efforts.

The globally efficient policy resolves 52.7% of the economic damages caused by MPP. The total cost of carrying out the abatement required by this policy is equal to approximately 16% of the total damages, so the net value of the agreement is equal to 36.7% of the total damages. The biggest winner from the globally efficient policy is USA, which obtains 39.5% percent of the net value, followed by Germany, which obtains 7.2% (table A1 column $v(a^*)/v_N(a^*)$). One the other hand, the biggest losers are the Philippines (-4.4% of the net value), and the Netherlands (-1.7%) These are indicative of three general patterns. First, high WTP countries gain the most because they receive the highest levels of abatement. Second, downstream countries that receive a high fraction of their MPP from others gain the most because the agreement internalises the damages imposed. This is best illustrated by comparing USA and China — China has a higher WTP than USA (19.3% of the total vs 17.0%), but obtains only 6.1% of the net value of the policy because it receives only 11% of its MPP from other countries, whereas USA receives 89.5%. Third, the countries who lose the most are not always lowest income countries (because they also have low abatement costs), but those who both receive a high fraction of self-pollution (so that they have high costs relative to WTP), and who are responsible for a fraction of MPP of high WTP countries (who therefore demand a lot of abatement).

4.2 Fair and Stable Transfers

Computational constraints would mean that it is not feasible for us to perfectly calculate the characteristic function for all 120 countries.[11] Instead, we follow the common approach (Sziklai et al., 2020; Eyckmans and Finus, 2006) of partitioning the players into the 'continental blocks' (table 2 column *Blocks*), and applying our model with N equal to this set of blocks. Columns a^* and R^* of table 2 shows the abatement performed and received by blocks under the globally efficient policy.[12]

We first observe that the agreement under globally efficient policy without transfers (column $v(a^*)$) is not stable because Middle America gets a negative payoff of -3.1%, less than both the 0% it gets from the disagreement equilibrium and the 6.1% it gets from its m-equilibrium policy. The total value of the globally efficient

---

[11] Though see section 6 for a discussion of approximation methods.

[12] We note that the globally efficient policy and values for the regional blocks are not the same as we would obtain from aggregating the globally efficient policy and values across the 120 individual countries because the former only internalises externalities between blocks, but not within them, whereas the latter internalises externalities both within and between blocks. We can circumvent this issue by allowing blocks to choose different policies for all of their members, but doing so obscures the connection between each block's characteristics and their fair allocation.

Table 2: The ESF agreement for Regional Blocks.

Blocks	β	γ	a^*	R^*	$v(a^*)$	$u(a^*, t^*)$	t^*	η
N America	18.5	4.6	-47.3	43.4	14.3	8.6	-5.7	1.7
M America	2.5	3.2	79.2	52.5	-3.1	2.6	5.7	-1.2
S America	4.7	8.9	5.4	8.1	0.2	0.3	0.1	0.8
Europe	18.3	10.2	-20.1	25.9	10.1	4.4	-5.8	2.3
Africa	4.6	7.3	52.1	44.6	-2.2	0.9	3.1	-2.6
S Asia	9.1	15.4	13.5	13.1	0.	0.4	0.4	0.
S&E Asia	32.1	58.4	3.1	4.1	0.7	0.3	-0.4	2.2
N&W Asia	8.9	11.2	38.5	31.7	-0.7	1.9	2.6	-0.4
Oceania	1.3	1.9	2.8	6.9	0.1	0.1	0.	1.

policy is equal to 19% of the total economic damages, so the equal sharing rule gives each player a payoff of 2.2%. This is not stable either: Middle America would still get -0.9%.

On the other hand, we find that the fair agreement is stable. Column $u(a^*, t^*)$ of table 2 shows that it allocates the most value to North America, Europe Middle America, and North and West Asia. It is clearly important to reward Middle America and North & West Asia because they are supplying the most abatement. North America and Europe are rewarded because they generate the most demand for abatement due to their relatively high WTP (18.5% and 18.3% of the total), and the fact that they receive a high proportion of their MPP from other continents (87.6 and 72.2% resp.). East & South East Asia has the highest WTP (32.1%), but it only receives 9% of its MPP from other regions, and it does not export much MPP to other high WTP regions, so it does not generate much value in the agreement.

The fair allocation can be implemented with the monetary transfers in column t^*. Thus the agreement (a^*, t^*) is efficient, fair and stable. These transfers reallocate surplus away from high WTP, downstream regions, who receive more abatement than they perform (North America, Europe, South and East Asia, and South America), towards low WTP, upstream regions, who perform more abatement than they receive (Middle America, Africa, North & West Asia, South Asia). The biggest loser is North America, which pays a transfer of 5.7% of the total economic damages in the

EFS agreement. This is unsurprising since the efficient policy both reduces North America's abatement costs (it performs -47.3% abatement), whilst increasing the abatement that it receives (43.4%). The biggest winner is Middle America, which receives a transfer of 5.7%. It performs the most abatement (79.2%) and receives less abatement than it performs (52.5%). Continents that give and receive little MPP relative to income (e.g. S&E Asia) generate least surplus and therefore get the smallest allocation from the fair agreement.

4.3 Partial Equilibria

In Section 2.2 we used theoretical examples to demonstrate how non-uniform mixing can lead to negative spillovers and even losses from cooperation. Here we show that these phenomena are not purely theoretical.

Example 1 demonstrates that countries can actually be made *worse off* by cooperating if complementary countries decrease their abatement in response to the coalition's increased abatement. This turns out to be the case if China and Malaysia form a coalition and the remaining countries best respond to them unilaterally. Less that 0.5% of China's MPP originates in Malaysia, and only 6% of Malaysia's MPP comes from China. Yet China has very high WTP, so in their γ-partial equilibrium, Malaysia abates 10% of its MPP. This has a small impact on China, but a large positive spillover on the rest of the world. The rest of the world, particularly the Philippines, best responds by increasing their collective emissions, resulting in an increase in the MPP received by China. Since China is such a high GNI country, it suffers very high losses from this, which wipes out any gains from cooperating with Malaysia and makes the value of the coalition negative.

Negative spillovers, such as those illustrated in Example 1 and **??** are common. For example, if Middle America and Europe form a coalition, then Middle America increases abatement by 26.7%, and Europe decreases abatement by 0.7%. The total amount of MPP transitioning to Africa and N&W Asia increases by 157 and 124 tons respectively. There are 37 other examples in the regional blocks dataset: 32 contain Europe, 26 contain Middle America, none contain Africa.

We also find an example in the data that produces second order negative spillovers. If North America, South America, South Asia, South & South East Asia and Oceania form a coalition, then they respectively increase abatement by 0.1, 7.8, 13.4, 3.7 and 4.6%. If the complementary continents (Middle America, Europe, Africa and North & West Asia) best respond unilaterally then they respectively *decrease* abatement by 0.6, -0.1, 0.3, and 0.1%. As a result, Europe gets a payoff of -0.05%.

We failed to find any empirical instances where coalition formation may cause total abatement to decrease (Example 2).

5 Sensitivity Analysis

We develop two new types of sensitivity test for three different parameters of our model: each country's marginal willingness to pay for abatement, β_i; the degree of heterogeneity of countries' willingness to pay, ϵ; and the quantity of plastic transitioning between each pair of countries, $\overleftarrow{T}_{ij}$. The first of these tests measures the elasticity of each country's actual payoff with respect to each of these parameters. If these elasticities are small then each player's true payoff is likely to be somewhat close to the fair payoffs predicted in Table 2. Our second test is a kind of stability stress test: we calculate the smallest error in our parameter estimates that would result in a coalition being dissatisfied with the agreement proposed in Table 2 (in the sense of receiving a negative excess). The details are specific to each type of parameter.

5.1 Payoff Sensitivity

5.1.1 Marginal WTP

Our main results assume that country i's marginal WTP for abatement, β_i, is directly proportional to its GNI. Not only is GNI subject to measurement error, but it seems likely that other factors such as the size of the marine economy and social attitudes towards the marine environment are likely to play an important role in determining a country's WTP for abatement.

Player i's utility depends only on its own WTP parameter β_i, so it is unaffected by measurement error in any other country j's WTP parameter. The elasticity of player i's utility with respect to β_i is

$$\begin{aligned}
\eta_i &:= \frac{\partial u_i(a^*, t^*; \beta_i)}{\partial \beta_i} \frac{\beta_i}{u_i(a^*, t^*; \beta_i)} && (11) \\
&= \frac{v_i(a^*; \beta_i)}{\beta_i} \frac{\beta_i}{u_i(a^*, t^*; \beta_i)} && (12) \\
&= \frac{v_i(a^*; \beta_i)}{u_i(a^*, t^*; \beta_i)} && (13) \\
&= 1 - \frac{t_i^*}{u_i(a^*, t^*; \beta_i)}. && (14)
\end{aligned}$$

This elasticity therefore provides a good measure of the sensitivity of country i's payoff to poor estimation of β_i. Equation (1) shows that player i's final payoff is linear in β_i. So if player i's true WTP is $\alpha = (\beta_i - \beta_i')/\beta_i$ percent of our estimate β_i, then, ceteris paribus, its true utility is $(u_i(a^*, t^*; \beta_i') - u_i(a^*, t^*; \beta_i))/u_i(a^*, t^*; \beta_i) = \alpha\eta_i(\beta_i)$ percent higher than its predicted utility $u(a^*, t^*; \beta_i)$,

The vector of elasticities $(\eta_i)_{i\in\mathbb{N}}$ is reported in column η of Table 2. The continents with the most sensitive payoffs are Africa (-2.6), Europe (2.3), and Southeast & East Asia (2.2). Europe receives a lot of abatement from the agreement, and it has high WTP for abatement, so its value, $v_{\text{Europe}}(a^*)$ is high. But much of this value is positive spillover from the reduction that Middle America carries out for the benefit of North America. Hence Europe's relative importance to the agreement, as reflected by the fair allocation, is not in proportion to its value. A similar explanation applies to Southeast & East Asia. Although it does not receive as much abatement as Europe, it has a high WTP, so it gets a relatively large value from positive spillovers.

Africa's story is slightly different. It is required to carry out a large amount of abatement to internalise its externality on Europe, which carries high costs. Since our estimated cost are directly proportional to WTP through Equation (9), an underestimate of WTP would imply that actual costs are higher than estimated costs, thereby substantially decreasing Africa's payoff (yielding a negative elasticity). Yet it doesn't receive as big a transfer as, say Middle America, to mitigate this sensitivity.

The least sensitive continents are South Asia (0), North and West Asia (-0.4), and South America (0.8). They are upstream of high WTP continents and downstream of low cost continents, so they benefit both from compensating transfers and from a greater level of overall abatement.

5.1.2 WTP Heterogeneity

Our main results assume that WTP is given by Equation 10 with income elasticity parameter $\epsilon = 1$. Higher ϵ implies that WTP varies more with national income, i.e. players are more heterogeneous in their WTP. In the limit, only the richest player has any WTP. At the other extreme, if $\epsilon = 0$, then all players have equal WTP, irrespective of their income. Czajkowski et al. (2017) find that the transfer error for household level transfers (the difference between the true WTP and that estimated by the benefit transfer method) is minimised by the value 1.32 (p.108). But our benefit transfer is carried out using the national elasticity, which may be higher or lower than that of the representative household. On the other hand, if countries benefit from abatement of plastics outside of their own waters (Börger et al., 2023), then this will generally act to equalise the benefits of abatement in different waters, which is captured by lower values of ϵ. Thus, there are good reasons to think that countries' true benefits might be better estimated with either a higher or lower value.

Figure 1 shows how each player's utility from (a^*, t^*) depends on ϵ by plotting the percentage difference between i's utility predicted by the model and i's actual utility,

$$\frac{u_i(a^*, t^*; \epsilon) - u_i(a^*, t^*; 1)}{u_i(a^*, t^*; 1)} = \eta_i \frac{y_i^\epsilon / (\sum_{j \in N} y_j^\epsilon) - \beta_i}{\beta_i}. \tag{15}$$

The biggest overestimates of utilities would occur if ϵ turns out to be much lower than 1. This would mean that low income countries have higher relative WTP, and high income countries have relatively lower WTP. For example, if ϵ is as low as 0.5 then M America and Africa respectively get 159% and 182% less than their intended utility. However, falls in utility are less than 100% for all values of ϵ between 0.75 and 1.4. The direction is reversed for values of $\epsilon > 1$, with the utility of the highest income continents being overestimated, but the magnitude is lower.

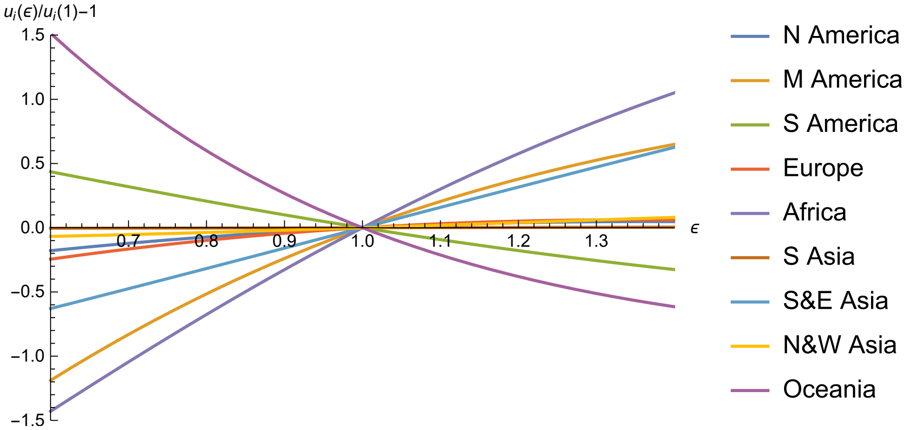

Figure 1: Agreements for the regional blocks under different elasticity parameters.

5.1.3 Plastic transport estimates

Plastic transport estimates are subject to errors. For example, Chassignet et al. (2021) omit MPP emissions from land locked countries and sea-based sources (mostly discarded fishing gear), and emissions that either decay or sink during the ten year period. These omissions only affect our results to the extent that they have different backward transition probabilities from the modelled MPW emissions. A change in The i, jth entry of Table 3 quantifies the change in country j's payoff as a percentage of their predicted payoff if T_{ij} is 1% higher than estimated. For instance, if North America's self-pollution is under-estimated by 1% then its true value is 0.1% lower than we estimated in Table 2. The elasticity of country j's payoff with respect to the quantity of MPP transitioning from i to j is $\frac{\partial u_j(a^*,t^*;T)}{\partial T_{i,j}} \frac{T_{i,j}}{u_j(a^*,t^*;T)}$.

Table 3: Payoff Sensitivities with respect to plastic transition estimates. All figures are given in percentages rounded to one decimal place.

	NAm	MAm	SAm	Eu	Af	SAs	ESAs	NWAs	Oc
NAm	-0.1	0.	0.	-0.1	0.	0.	0.	0.	0.
MAm	0.4	-0.5	0.8	0.1	0.	0.	0.2	0.	0.2
SAm	0.	0.2	-0.8	0.	0.	0.	0.	0.	0.
Eu	0.	0.	0.	-0.5	0.	0.	0.	0.	0.
Af	0.	0.	0.1	0.3	-0.9	0.5	0.4	0.7	0.5
SAs	0.	0.	0.	0.	0.1	-0.7	1.6	0.	0.
ESAs	-0.1	0.	0.	0.	0.3	0.2	-2.2	0.	0.
NWAs	0.	0.	0.	0.1	0.4	0.	0.	-0.8	0.
Oc	0.	0.	0.	0.	0.	0.	0.	0.	-0.7

5.2 Stability Stress Testing

5.2.1 Marginal WTP

What is the smallest percentage by which the relative WTP of a coalition would have to change in order to produce an unstable coalition? If the true WTP is a fraction α of our estimated β, then the excess is

$$\begin{aligned} \mathrm{ex}_C(\alpha\beta_C) &= t^*_C + v_C(a^*;\alpha\beta_C) - v_C(a^C;\alpha\beta_C) \qquad (16) \\ &= t^*_C + \alpha(v_C(a^*;\beta_C) - v_C(a^C;\beta_C)) \qquad (17) \end{aligned}$$

which is negative if and only if

$$\bar{\alpha}_C := -t_C/(v_C(a^*;\beta_C) - v_C(a^C;\beta_C)) > \alpha. \qquad (18)$$

Thus, coalition C is better off playing their m-equilibrium than joining the fair agreement (i.e. has a negative excess) if and only if the true WTP of a coalition is less than $\bar{\alpha}_C$ times our estimate β_C. The coalitions closest to a negative excess are: $\alpha_{\{\text{Europe}\}} = 0.91$, $\alpha_{\{\text{S. America,Europe}\}} = 0.90\%$, $\alpha_{\{\text{Europe,Oceania}\}} = 0.90$, and $\alpha_{\{\text{S America,Europe,Oceania}\}} = 0.89$. This means that Europe's WTP would be need to be 9% lower than estimated in order for the fair agreement to be unstable. But this seems unlikely to be an issue in practice: Europe, South America and Oceania are all keen on conservation at an international level. Overall, 99% of coalitions would need their value be underestimated by more than 11.7%, and 95% would need their value to be underestimated by more than 22.0%. About 10% of coalitions cannot be destabilised because they have negative thresholds. This means that they receive positive transfers even though they are better off under the efficient abatement policy than their partial equilibrium.

5.2.2 WTP Heterogeneity

If the true benefit parameters are described by Equation 10 with $\epsilon \neq 1$, then the true excess of coalition C from the proposed agreement is given by

$$\begin{aligned} \mathrm{ex}_C(\epsilon) &= t_C + v_C(a^*;\epsilon) - v_C(a^C;\epsilon) \qquad (19) \\ &= t_C + \frac{\beta_C(\epsilon)}{\beta_C(1)}(v_C(a^*;1) - v_C(a^C;1)) \qquad (20) \end{aligned}$$

where $\beta_i(\epsilon) \coloneqq y_i^\epsilon / \sum_{j\in N} y_j^\epsilon$.

Figure 2 plots the excesses of all the singleton coalitions for values of ϵ between 0.6 and 1.4. In general, the excess is increasing in ϵ. The least satsified coalition is {Africa, Oceania}, which is dissatisfied for $\epsilon \leq 0.83$, followed by three coalitions containing Africa, Oceania and South America at $\epsilon = 0.81$ For low income continents, small changes in ϵ will increase their WTP dramatically. Africa has a high estimated excess, but it is quite sensitive to ϵ. The most sensitive country is Middle America, even though has an even higher estimated excess than Africa. By contrast, South America and Ocean's excess is not very sensitive but it is very small (it is barely distinguishable from the horizontal axis). The next least stable coalition is {Europe, S&E Asia} at 0.78, all other coalitions have a threshold at or below 0.70 Europe is slightly above average WTP so falls comparatively slowly when ϵ falls.

Overall, it overestimating ϵ seems to lead to less stable agreements, so propose that a cautious approach should assumed lower income elasticity.

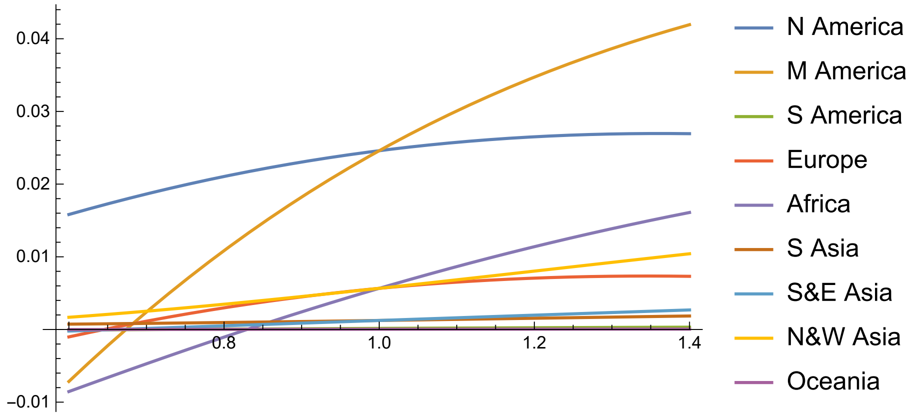

Figure 2: Excess as a function of the elasticity parameters.

5.2.3 Plastic transition quantities

The first order Taylor approximation of the excess arising from a transition quantity T'_{ij} is

$$\mathrm{ex}_C(T'_{ij}) = \mathrm{ex}_C(T_{ij}) + (T'_{ij} - T_{ij})\frac{\partial \mathrm{ex}_C}{\partial T_{ij}}$$

where $\frac{\partial \mathrm{ex}_C}{\partial T_{ij}} = \frac{\partial v_C(a^*;T_{ij})}{\partial T_{ij}} - \frac{\partial v_C(a^C;T_{ij})}{\partial T}$. Coalition C is dissatisfied if and only if

$$\mathrm{ex}_C(T_{ij}) < (T_{ij} - T'_{ij})\frac{\partial \mathrm{ex}_C}{\partial T_{ij}}.$$

If C's satisfaction is increasing in T_{ij}, then $\frac{\partial \mathrm{ex}_C}{\partial T_{ij}} > 0$ so the condition reduces to

$$\zeta_{ij}(C) := \mathrm{ex}_C(T)/(T_{ij}\frac{\partial \mathrm{ex}_C}{\partial T_{ij}}) < 1 - T'_{ij}/T_{ij},$$

which says that our estimate of T_{ij} would have to be at overestimated by a fraction of at least $\zeta_{ij}(C)$ in order for coalition C to be unstable. This is the minimum error (in percentage terms) required to produce a dissatisfied coalition. Similarly, if $\frac{\partial \mathrm{ex}_C}{\partial T_{ij}} < 0$ then our estimate of T_{ij} would have to be at underestimated by a fraction of at least $-\zeta_{ij}(C)$ in order for coalition C to be unstable.

Table 4 therefore identifies $\zeta_{ij}(\arg\max_{C\subseteq N} |\zeta_{ij}(C)|)$ for each i, j. It shows that the fair agreement is highly stable with respect to most of the transition quantities. The least stable coalitions are:

- {North America, Middle America, South America, Europe, Africa, North & West Asia}: responsible for all values below 100 in the North America, Africa, and North & West Asia columns, and for the lowest values in the Europe column.
- {North America, Middle America}: responsible for all of the Middle America column, and the lowest value in the North America column.
- {South Asia, East & Southeast Asia}: responsible for most values in the South Asia column, including all of the values below 100.
- {East & Southeast Asia}: lowest values in the East & Southeast Asia column.

Table 4: Stability Sensitivities with respect to plastic transition estimates. All figures are given in percentages rounded to one decimal place.

	NAm	MAm	SAm	Eu	Af	SAs	ESAs	NWAs	Oc
NAm	25	593	4337	21	11261	∞	3700	4553	51345
MAm	-10	-23	-110	-88	17602	-429146	-684	-28079	-268
SAm	35	47	115	570	1140	54212	2832	49917	327
Eu	6844	51141	379340	-35	134	∞	∞	-100	418527
Af	49	459	1149	7	-13	171	414	25	-145
SAs	∞	∞	4636642	∞	160	-17	23	1192	-20566
ESAs	36	3038644	37541	85141	27	19	-24	283	185
NWAs	-1046	∞	∞	-10	49	-19398	-8925	-39	∞
Oc	3090	∞	34923	702	3072	3332	34	4689	-258

5.3 γ-Core Stability

Allocations in the m-core deter coalitional deviations based on the belief that complementary countries best respond as a coalition. By contrast, previous literature has used the γ-core, which rules out coalitional deviations based on the belief that complementary countries best respond unilaterally. Corollary 1 implies that, under uniform mixing, any agreement that is stable according to our definition must also be stable according to the **?** definition, even though the reverse is not true. However, **??** shows that γ-core stability is not always weaker than m-core stability in the case of non-uniformly mixing pollutants, such as MPP. Despite the theoretical possibility that efficient and fair agreements might not be stable under the γ-core notion of stability, we find that them to be so in all of the empirical cases we studied.

6 Conclusion

To conclude, we estimate the efficient marine plastic abatement policy for 120 countries, and find that it abates 32.2% of MPP and generates a surplus equal to 36.7% of the total damages caused by MPP. We calculate transfers to implement a fair allocation and verify its core stability for a set of nine continental blocks. We find that producing an efficient, fair and stable agreement requires redistribution of value away from countries who benefit the most from the efficient policy, towards countries who bear the highest costs of abatement. Finally, we incorporate exogenous abatement cost estimates and find that even in the absence of an agreement, individual countries would be better off carrying more abatement than existing levels.

Our results have important implications for policy. Specifically, agreements should include (i) the biggest polluters, (ii) upsteam countries who exert large externalities on members, (iii) downstream countries who receive externalities from agreement members, and therefore benefit from their increased abatement.

Ours is one of the first studies to address the problem of MPP in the economic literature on transboundary pollution, and as such there are many potential avenues for future research. We briefly mention three. First, limitations in computing power prevented us from estimating fair and stable transfers for all 120 countries in our sample. Doing so would require calculation of 2^{120} equilibrium abatement policies (almost twice the number of stars in the universe), so computing power is unlikely to advance enough to address this in the foreseeable future. Instead, future research can experiment with using approximation methods, such as those provided by Aas et al. (2021) and Castro et al. (2009), to calculate the Shapley value and check core stability.

Second, our model assumes that players' costs, benefits and abatement choices are common knowledge. In actual fact, there is a high degree of both incomplete and private information about them. Private information is likely to be particularly problematic since it can lead to countries adversely selecting which agreement terms to accept. Future research can address this by taking a mechanism design approach to incentivise truthful reporting (e.g. Candel-Sánchez, 2006; Martimort and Sand-Zantman, 2016; Kaitala et al., 2006).

Finally, our static model assumes that costs, benefits and transition probabilities are fixed. In practice, they are constantly evolving over time as technologies and the needs of different economies change. countries make one-off and instant abatement decisions that lead to instance reductions of MPP stocks. In practice, abatement choices take time to implement, and the consequences can take years to realise. we have modelled our model can be improved by incorporating dynamic decisions along the lines of De Frutos and Martín-Herrán (2019); de Frutos and Martín-Herrán (2019); Mäler and De Zeeuw (1998)

Acknowledgements

The research was funded by the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA).

We would like to thank Ramses Abul Naga, Tobias Börger, Jim Clark, Andrew Clausen, Nick Hanley, Robert Johnston, Lászlo Kóczy, András Pora, Luca Sandrini, Róbert Somogyi, Csete Mária Szalmáné, Frans de Vries for their helpful suggestions and comments.

References

Aas, K., Jullum, M., Løland, A., 2021. Explaining individual predictions when features are dependent: More accurate approximations to shapley values. Artificial

Intelligence 298, 103502.

Alvarado-Quesada, I., Weikard, H.P., 2017. International environmental agreements for biodiversity conservation: a game-theoretic analysis. International Environmental Agreements: Politics, Law and Economics 17, 731–754.

Amato-Lourenço, L.F., Carvalho-Oliveira, R., Júnior, G.R., dos Santos Galvão, L., Ando, R.A., Mauad, T., 2021. Presence of airborne microplastics in human lung tissue. Journal of Hazardous Materials 416, 126124. URL: https://www.sciencedirect.com/science/article/pii/S0304389421010888, doi:https://doi.org/10.1016/j.jhazmat.2021.126124.

Beaumont, N., Börger, T., Clark, J., Hanley, N., Johnston, R.J., Meginnis, K., Stapenhurst, C., de Vries, F., 2023. Significant benefits from international cooperation over marine plastic pollution. preprint (version 1) available at research square. doi:https://doi.org/10.21203/rs.3.rs-3328986/v1.

Beaumont, N.J., Aanesen, M., Austen, M.C., Börger, T., Clark, J.R., Cole, M., Hooper, T., Lindeque, P.K., Pascoe, C., Wyles, K.J., 2019. Global ecological, social and economic impacts of marine plastic. Marine Pollution Bulletin 142, 189–195. URL: https://www.sciencedirect.com/science/article/pii/S0025326X19302061, doi:https://doi.org/10.1016/j.marpolbul.2019.03.022.

Benson, N.U., Agboola, O.D., Fred-Ahmadu, O.H., De-la Torre, G.E., Oluwalana, A., Williams, A.B., 2022. Micro (nano) plastics prevalence, food web interactions, and toxicity assessment in aquatic organisms: a review. Frontiers in Marine Science 9, 291.

Borrelle, S.B., Rochman, C.M., Liboiron, M., Bond, A.L., Lusher, A., Bradshaw, H., Provencher, J.F., 2017. Why we need an international agreement on marine plastic pollution. Proceedings of the National Academy of Sciences 114, 9994–9997.

Brouard, N., 2019. Chapter 10 - theory and applications of backward probabilities and prevalences in cross-longitudinal surveys, in: Srinivasa Rao, A.S., Rao, C. (Eds.), Integrated Population Biology and Modeling, Part B. Elsevier. volume 40 of *Handbook of Statistics*, pp. 435–486. URL: https://www.sciencedirect.com/science/article/pii/S0169716118301020, doi:https://doi.org/10.1016/bs.host.2018.11.009.

Börger, T., Hanley, N., Johnston, R.J., Meginnis, K., Ndebele, T., Ali Siyal, G.E., de Vries, F., 2023. Equity preferences and abatement cost sharing in international environmental agreements. American Journal of Agricultural Economics n/a. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ajae.12392, doi:https://doi.org/10.1111/ajae.12392, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajae.12392.

Candel-Sánchez, F., 2006. The externalities problem of transboundary and persistent pollution. Journal of Environmental Economics and Management 52, 517–526.

Carney Almroth, B., Eggert, H., 2019. Marine plastic pollution: Sources, impacts, and policy issues. Review of Environmental Economics and Policy 13, 317–326. URL: https://doi.org/10.1093/reep/rez012, doi:10.1093/reep/rez012, arXiv:https://doi.org/10.1093/reep/rez012.

Carraro, C., Eyckmans, J., Finus, M., 2006. Optimal transfers and participation decisions in international environmental agreements. The Review of International Organizations 1, 379–396.

Carroll, C., Sousa, J., Thevenon, F., 2014. Plastic debris in the ocean : the characterization of marine plastics and their environmental impacts, situation analysis report. Technical Report. URL: https://portals.iucn.org/library/node/44966, doi:https://doi.org/10.2305/IUCN.CH.2014.03.en.

Castro, J., Gómez, D., Tejada, J., 2009. Polynomial calculation of the shapley value based on sampling. Computers & Operations Research 36, 1726–1730.

Chander, P., Tulkens, H., 1995. A core-theoretic solution for the design of cooperative agreements on transfrontier pollution. International tax and public finance 2, 279–293.

Chander, P., Tulkens, H., 1997. The core of an economy with multilateral environmental externalities. International journal of game theory 26, 379–401.

Chassignet, E.P., Xu, X., Zavala-Romero, O., 2021. Tracking marine litter with a global ocean model: where does it go? where does it come from? Frontiers in Marine Science 8, 667591. URL: https://www.frontiersin.org/articles/10.3389/fmars.2021.667591/full#supplementary-material. accessed 31/05/2023.

Cole, M., Lindeque, P., Halsband, C., Galloway, T.S., 2011. Microplastics as contaminants in the marine environment: A review. Marine Pollution Bulletin 62, 2588–2597. URL: https://www.sciencedirect.com/science/article/pii/S0025326X11005133, doi:https://doi.org/10.1016/j.marpolbul.2011.09.025.

Czajkowski, M., Ahtiainen, H., Artell, J., Meyerhoff, J., 2017. Choosing a Functional Form for an International Benefit Transfer: Evidence from a Nine-country Valuation Experiment. Ecological Economics 134, 104–113. URL: https://ideas.repec.org/a/eee/ecolec/v134y2017icp104-113.html, doi:10.1016/j.ecolecon.2017.0.

De Frutos, J., Martín-Herrán, G., 2019. Spatial effects and strategic behavior in a multiregional transboundary pollution dynamic game. Journal of Environmental Economics and Management 97, 182–207.

Efferth, T., Paul, N.W., 2017. Threats to human health by great ocean garbage patches. The Lancet. Planetary health 1 8, e301–e303.

Eyckmans, J., Finus, M., 2006. Coalition formation in a global warming game: How the design of protocols affects the success of environmental treaty-making. Natural Resource Modeling 19, 323–358.

Fernandez, L., 2009. Wastewater pollution abatement across an international border. Environment and Development Economics 14, 67–88.

Folmer, H., von Mouche, P., 2015. Nash equilibria of transboundary pollution games, in: Handbook of Research methods and Applications in Environmental Studies. Edward Elgar Publishing, pp. 504–524.

de Frutos, J., Martín-Herrán, G., 2019. Spatial vs. non-spatial transboundary pollution control in a class of cooperative and non-cooperative dynamic games. European Journal of Operational Research 276, 379–394.

Gillies, D.B., 1959. Solutions to general non-zero-sum games. Contributions to the Theory of Games 4, 47–85.

Hafalir, I.E., 2007. Efficiency in coalition games with externalities. Games and Economic Behavior 61, 242–258.

Hayati, Y., Adrianto, L., Krisanti, M., Pranowo, W.S., Kurniawan, F., 2020. Magnitudes and tourist perception of marine debris on small tourism island: Assessment of tidung island, jakarta, indonesia. Marine Pollution Bulletin 158, 111393. URL: `https://www.sciencedirect.com/science/article/pii/S0025326X20305117`, doi:`https://doi.org/10.1016/j.marpolbul.2020.111393`.

Jang, Y.C., Hong, S., Lee, J., Lee, M.J., Shim, W.J., 2014. Estimation of lost tourism revenue in geoje island from the 2011 marine debris pollution event in south korea. Marine pollution bulletin 81, 49–54.

Kaitala, V., Mäler, K.G., Tulkens, H., 2006. The acid rain game as a resource allocation process, with application to negotiations between Finland, Russia and Estonia. Springer.

Lebreton, L., Andrady, A., 2019. Future scenarios of global plastic waste generation and disposal. Palgrave Communications 5, 1–11.

Lebreton, L.C., Van Der Zwet, J., Damsteeg, J.W., Slat, B., Andrady, A., Reisser, J., 2017. River plastic emissions to the world's oceans. Nature communications 8, 15611.

Lejano, R.P., Li, L., 2019. Cooperative game-theoretic perspectives on global climate action: Evaluating international carbon reduction agreements. Journal of Environmental Economics and Policy 8, 79–89.

Li, W.C., Tse, H., Fok, L., 2016. Plastic waste in the marine environment: A review of sources, occurrence and effects. Science of the total environment 566, 333–349.

Luenberger, D.G., Ye, Y., et al., 1984. Linear and nonlinear programming. volume 2. Springer.

Mäler, K.G., 1989. The acid rain game, in: Folmer, H., van lerland, E. (Eds.), Valuation Methods and Policy Making in Environmental Economics. Elsevier. volume 36 of *Studies in Environmental Science*, pp. 231–252. URL: https://www.sciencedirect.com/science/article/pii/S0166111608700359, doi:https://doi.org/10.1016/S0166-1116(08)70035-9.

Mäler, K.G., De Zeeuw, A., 1998. The acid rain differential game. Environmental and Resource Economics 12, 167–184.

Martimort, D., Sand-Zantman, W., 2016. A mechanism design approach to climate-change agreements. Journal of the European Economic Association 14, 669–718.

Maschler, M., 1992. Chapter 18 the bargaining set, kernel, and nucleolus, Elsevier. volume 1 of *Handbook of Game Theory with Economic Applications*, pp. 591–667. URL: https://www.sciencedirect.com/science/article/pii/S1574000505800219, doi:https://doi.org/10.1016/S1574-0005(05)80021-9.

McGinty, M., 2011. A risk-dominant allocation: Maximizing coalition stability. Journal of Public Economic Theory 13, 311–325.

McIlgorm, A., Campbell, H.F., Rule, M.J., 2011. The economic cost and control of marine debris damage in the asia-pacific region. Ocean & Coastal Management 54, 643–651. URL: https://www.sciencedirect.com/science/article/pii/S0964569111000688, doi:https://doi.org/10.1016/j.ocecoaman.2011.05.007.

Meinhardt, H.I., 2022. Tugames. URL: URLofthepackage'sdocumentationorsourcecode. version 3.1.0.

Merkl, A., Charles, D., 2022. The Price of Plastic Pollution: Social Costs and Corporate Liabilities. Technical Report. Minderoo Foundation. URL: https://cdn.minderoo.org/content/uploads/2022/10/14130457/The-Price-of-Plastic-Pollution.pdf.

Moulin, H., 1987. Equal or proportional division of a surplus, and other methods. International Journal of Game Theory 16, 161–186.

NOAA, 2021. Marine Debris Program Accomplishments Report. Technical Report. URL: https://marinedebris.noaa.gov/report/accomplishments-report.

Ragusa, A., Notarstefano, V., Svelato, A., Belloni, A., Gioacchini, G., Blondeel, C., Zucchelli, E., De Luca, C., D'Avino, S., Gulotta, A., et al., 2022. Raman microspectroscopy detection and characterisation of microplastics in human breastmilk. Polymers 14, 2700.

Rosen, J.B., 1965. Existence and uniqueness of equilibrium points for concave n-person games. Econometrica 33, 520–534. URL: http://www.jstor.org/stable/1911749.

Schmeidler, D., 1969. The nucleolus of a characteristic function game. SIAM Journal on applied mathematics 17, 1163–1170.

Shapley, L.S., 1952. A Value for N-Person Games. RAND Corporation, Santa Monica, CA. doi:10.7249/P0295.

Soós, R., Whiteman, A., Gavgas, G., 2022. The cost of preventing ocean plastic pollution. Technical Report 190. URL: https://www.oecd-ilibrary.org/content/paper/5c41963b-en, doi:https://doi.org/https://doi.org/10.1787/5c41963b-en.

Sziklai, B.R., Kóczy, L.A., Csercsik, D., 2020. The impact of nord stream 2 on the european gas market bargaining positions. Energy Policy 144, 111692. doi:https://doi.org/10.1016/j.enpol.2020.111692.

The World Bank, 2023. GNI, PPP (current international $). URL: https://data.worldbank.org/indicator/NY.GNP.MKTP.PP.CD.. (accessed 31/05/2023).

UNEP, 2022. End plastic pollution: Towards an international legally binding instrument URL: https://wedocs.unep.org/handle/20.500.11822/38913.

Van Steenberghe, V., 2004. Core-stable and equitable allocations of greenhouse gas emission permits .

Vince, J., Hardesty, B.D., 2018. Governance solutions to the tragedy of the commons that marine plastics have become. Frontiers in Marine Science 5. URL: https://www.frontiersin.org/articles/10.3389/fmars.2018.00214, doi:10.3389/fmars.2018.00214.

Watts, A.J., Urbina, M.A., Corr, S., Lewis, C., Galloway, T.S., 2015. Ingestion of plastic microfibers by the crab carcinus maenas and its effect on food consumption and energy balance. Environmental science & technology 49, 14597–14604.

Weikard, H.P., Finus, M., Altamirano-Cabrera, J.C., 2006. The impact of surplus sharing on the stability of international climate agreements. Oxford Economic Papers 58, 209–232.

Winter, E., 2002. The shapley value. Handbook of game theory with economic applications 3, 2025–2054.

Wolfram Research, Inc., 2020. Mathematica. URL: https://www.wolfram.com/mathematica/. version 12.2.

Zhao, C., Hou, Y., Liu, M., Gong, Y., Wang, J., 2023. Research on the cooperative network game model of marine plastic waste management. Marine Policy 149, 105504.

A Tables

Tables A2 and A3 in appendix A show sub-matrices of the global transition matrix for the North Atlantic countries, and for China, Indonesia and Malaysia.

Table A1: Main results. All figures are rounded to two decimal places.

Country	β_i	T_{ii}	γ_i	a^*	A^*	$v(a^*)$	$v(a^*)/v_N(a^*)$
ALB	0.03	68.57	0.04	81.84	63.11	-0.05	-0.13
DZA	0.39	81.12	0.63	69.9	68.1	-0.41	-1.11
AGO	0.16	99.23	0.31	6.55	6.74	0.0	0.0
ARG	0.73	4.47	0.07	16.63	73.79	0.67	1.83
AUS	1.07	3.51	0.07	11.92	62.11	0.9	2.46
BHS	0.01	1.06	0.0	99.43	85.54	0.01	0.02
BGD	0.82	55.99	0.92	34.95	29.66	0.02	0.05
BEL	0.51	1.84	0.02	-13.55	61.57	0.43	1.18
BLZ	0.0	12.31	0.0	97.2	56.09	0.0	0.0
BEN	0.03	6.47	0.0	96.39	57.21	0.01	0.04
BRA	2.47	99.12	4.9	2.63	2.95	0.01	0.04
BRN	0.02	1.69	0.0	68.34	59.54	0.02	0.05
BGR	0.14	0.37	0.0	88.16	35.93	0.08	0.21
CPV	0.0	16.13	0.0	98.31	69.38	0.0	0.0
KHM	0.06	1.1	0.0	75.65	29.82	0.03	0.07
CMR	0.08	19.03	0.03	5.21	-3.39	-0.01	-0.02
CAN	1.42	8.99	0.26	-4.95	14.17	0.39	1.05
CHL	0.36	84.3	0.61	0.02	3.16	0.02	0.06
CHN	19.26	89.02	34.3	7.35	13.52	2.24	6.09
COL	0.61	16.61	0.2	58.49	43.05	0.23	0.64
COG	0.01	61.95	0.02	42.72	28.94	0.0	-0.01
CRI	0.08	6.52	0.01	60.27	46.61	0.05	0.14
CIV	0.11	90.68	0.2	6.65	10.76	0.01	0.02

Continued on next page

Table A1 – continued from previous page

Country	β_i	T_{ii}	γ_i	a^*	A^*	$v(a^*)$	$v(a^*)/v_N(a^*)$
HRV	0.1	32.01	0.06	32.25	43.02	0.04	0.11
CYP	0.03	0.05	0.0	94.84	32.46	0.01	0.04
DNK	0.29	0.4	0.0	95.46	44.84	0.2	0.54
DJI	0.0	45.4	0.0	83.25	73.11	0.0	-0.01
DMA	0.0	6.9	0.0	99.66	75.88	0.0	0.0
DOM	0.15	91.5	0.27	88.08	86.68	-0.43	-1.17
ECU	0.15	80.03	0.24	1.89	14.08	0.03	0.09
EGY	1.0	41.16	0.82	4.16	30.7	0.49	1.32
SLV	0.04	50.62	0.04	43.42	59.18	0.01	0.03
GNQ	0.02	15.63	0.0	22.07	23.88	0.01	0.01
EST	0.04	1.93	0.0	91.09	22.5	0.01	0.03
FJI	0.01	35.2	0.01	89.37	68.55	-0.01	-0.01
FIN	0.24	6.29	0.03	64.55	18.45	0.05	0.13
FRA	2.64	3.99	0.21	0.32	56.52	2.14	5.81
GAB	0.03	7.23	0.0	61.92	29.43	0.01	0.03
GMB	0.0	49.1	0.0	97.19	87.17	-0.01	-0.03
GEO	0.04	1.68	0.0	93.97	33.92	0.02	0.05
DEU	3.82	21.95	1.68	-13.45	39.81	2.65	7.2
GHA	0.14	77.47	0.22	47.75	39.8	-0.05	-0.14
GRC	0.24	7.29	0.04	70.39	35.16	0.1	0.26
GTM	0.12	18.1	0.04	63.85	65.51	0.06	0.16
GIN	0.03	52.09	0.03	-14.95	20.73	0.01	0.04
GNB	0.0	55.53	0.0	87.74	85.64	0.0	-0.01
GUY	0.01	57.97	0.01	81.57	84.55	-0.01	-0.03
HTI	0.03	46.84	0.03	99.41	92.74	-0.11	-0.29
HND	0.04	28.99	0.02	58.28	56.9	0.01	0.03
HKG	0.37	4.67	0.03	81.48	17.94	0.06	0.17
ISL	0.02	4.04	0.0	61.5	42.82	0.01	0.03

Continued on next page

Table A1 – continued from previous page

Country	β_i	T_{ii}	γ_i	a^*	A^*	$v(a^*)$	$v(a^*)/v_N(a^*)$
IND	7.11	75.74	10.77	-0.83	11.98	1.69	4.6
IDN	2.57	36.07	1.86	8.96	42.31	1.54	4.19
IRN	1.07	62.78	1.34	25.12	22.22	0.03	0.09
IRL	0.29	5.32	0.03	84.82	64.61	0.19	0.52
ISR	0.29	1.44	0.01	30.71	31.75	0.15	0.41
ITA	2.07	5.81	0.24	-79.07	60.01	1.88	5.12
JAM	0.02	28.39	0.01	99.19	89.99	-0.04	-0.1
JPN	4.36	4.01	0.35	-11.03	31.77	2.37	6.43
KEN	0.19	4.46	0.02	70.18	48.07	0.12	0.33
KOR	1.87	16.57	0.62	18.75	16.91	0.45	1.22
LVA	0.05	23.68	0.02	19.8	23.05	0.01	0.04
LBN	0.06	12.77	0.02	79.74	36.21	0.01	0.03
LBR	0.01	58.56	0.01	46.97	50.14	0.0	0.0
LBY	0.1	26.38	0.05	62.36	63.37	0.03	0.09
LTU	0.09	3.46	0.01	89.66	15.85	0.01	0.03
MDG	0.03	1.81	0.0	84.28	30.95	0.02	0.04
MYS	0.71	27.27	0.39	66.97	63.84	0.19	0.51
MRT	0.02	89.45	0.04	44.8	45.3	-0.01	-0.02
MUS	0.02	0.11	0.0	84.8	33.91	0.01	0.03
MEX	1.81	28.85	1.04	39.1	51.54	0.87	2.35
MAR	0.24	93.49	0.44	66.75	66.49	-0.28	-0.75
MOZ	0.03	44.42	0.03	89.73	53.66	-0.04	-0.11
MMR	0.22	25.86	0.11	85.0	38.65	-0.08	-0.21
NAM	0.02	81.34	0.03	2.95	4.18	0.0	0.0
NLD	0.82	66.04	1.09	69.94	58.97	-0.62	-1.7
NZL	0.18	7.74	0.03	-17.16	27.03	0.09	0.24
NIC	0.03	5.33	0.0	97.77	41.55	0.01	0.02
NGA	0.82	63.55	1.05	-6.78	20.11	0.37	1.0

Continued on next page

Table A1 – continued from previous page

Country	β_i	T_{ii}	γ_i	a^*	A^*	$v(a^*)$	$v(a^*)/v_N(a^*)$
NOR	0.29	2.76	0.02	5.99	46.36	0.2	0.55
OMN	0.12	1.3	0.0	89.08	32.99	0.06	0.16
PAK	0.94	98.39	1.86	18.33	18.27	-0.06	-0.17
PAN	0.09	9.97	0.02	88.55	51.14	0.03	0.08
PNG	0.03	54.43	0.03	67.39	73.9	-0.01	-0.02
PER	0.31	98.88	0.61	8.89	8.92	0.0	-0.01
PHL	0.79	91.1	1.44	80.75	77.65	-1.63	-4.42
POL	1.03	42.15	0.87	-0.31	8.51	0.17	0.46
PRT	0.28	13.15	0.07	37.78	61.5	0.21	0.56
PRI	0.06	0.71	0.0	92.98	63.14	0.05	0.14
QAT	0.2	0.49	0.0	97.24	8.99	0.03	0.08
ROU	0.51	19.34	0.2	-12.89	32.18	0.3	0.81
RUS	3.43	55.71	3.82	12.7	24.32	0.95	2.57
WSM	0.0	66.67	0.0	91.25	73.07	0.0	-0.01
STP	0.0	66.67	0.0	74.41	51.79	0.0	0.0
SAU	1.33	4.22	0.11	24.98	9.24	0.2	0.55
SEN	0.05	73.26	0.07	77.88	79.37	-0.06	-0.16
SLE	0.01	64.74	0.01	25.26	38.76	0.0	0.01
SLB	0.0	75.0	0.0	99.21	90.89	-0.01	-0.03
SOM	0.02	3.26	0.0	92.86	40.34	0.01	0.02
ZAF	0.63	64.56	0.81	4.16	20.98	0.2	0.55
ESP	1.45	8.31	0.24	37.36	64.27	1.16	3.14
LKA	0.24	11.12	0.05	79.26	25.32	0.02	0.06
LCA	0.0	100.0	0.0	96.77	96.77	-0.01	-0.03
SDN	0.14	43.84	0.12	28.23	17.45	0.0	0.01
SUR	0.01	93.77	0.01	89.99	86.57	-0.02	-0.06
SWE	0.48	8.68	0.08	54.76	41.15	0.25	0.67
TZA	0.13	44.48	0.11	81.37	51.1	-0.09	-0.26

Continued on next page

Table A1 – continued from previous page

Country	β_i	T_{ii}	γ_i	a^*	A^*	$v(a^*)$	$v(a^*)/v_N(a^*)$
THA	0.99	48.53	0.96	-1.02	28.32	0.49	1.34
TLS	0.01	9.0	0.0	95.76	41.86	0.0	0.0
TGO	0.01	39.47	0.01	88.75	64.05	-0.01	-0.03
TTO	0.03	38.28	0.02	98.54	84.37	-0.06	-0.18
TUN	0.1	37.05	0.07	79.0	70.78	-0.02	-0.07
TUR	1.84	84.79	3.13	31.32	33.02	-0.16	-0.43
UKR	0.45	83.37	0.75	54.26	48.3	-0.26	-0.69
ARE	0.53	26.0	0.28	-6.81	21.6	0.22	0.6
GBR	2.48	46.51	2.31	14.53	40.79	1.25	3.39
USA	17.04	10.52	3.59	-1.62	61.05	14.52	39.46
URY	0.06	86.36	0.1	76.84	67.27	-0.1	-0.27
VUT	0.0	16.08	0.0	99.0	71.46	0.0	0.0
VNM	0.82	23.33	0.38	71.15	39.08	0.04	0.11

B Block membership

North America: Canada, United States.
Middle America: Bahamas, The, Mexico, Nicaragua, Panama, Puerto Rico, El Salvador, Guatemala, Honduras, Belize, Costa Rica, Dominica, Dominican Republic, Haiti, Jamaica, St. Lucia, Trinidad and Tobago.
South America: Argentina, Brazil, Chile, Colombia, Ecuador, Guyana, Peru, Suriname, Uruguay.
Europe: Albania, Belgium, Bulgaria, Croatia, Denmark, Estonia, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Spain, Sweden, Ukraine, United Kingdom.
Africa: Algeria, Angola, Benin, Cabo Verde, Cameroon, Congo, Rep., Cote d'Ivoire, Djibouti, Egypt, Arab Rep., Equatorial Guinea, Gabon, Gambia, The, Ghana, Guinea, Guinea-Bissau, Kenya, Liberia, Libya, Madagascar, Mauritania, Mauritius,

Table A2: North Atlantic transition matrix. Rounded to closest percent. Empty cells correspond to 0s.

	BE	CA	DK	DO	FR	DE	HT	IR	MX	MA	NL	PT	ES	SE	GB	US
BE	2															
CA		9														
DK														5		
DO			1	91	13	1	51	12	5	1	2	15	6	1	7	11
FR	4		1		4	1					2			1	1	
DE			12			22					1			13		
HT	21		3	2	16	1	47	23	13	1	3	2	6	2	1	24
IR								5							6	
MX									29							9
MA			1	2	1	1	1	11	1	93	2	18	22	1	6	1
NL	55	1	52			51					66			35	3	
PT					3							13	2			
ES					9					1		2	8		1	
SE														9		
GB	13		24		1	18		6			14			23	47	
US	4	48	2		15	2		17	2	1	4	14	5	2	9	11

Table A3: China-Indonesia-Malaysia transition matrix.

	China	Indonesia	Malaysia	Other
China	89.0	3.1	6.3	7.6
Indonesia	0.1	36.1	5.0	2.1
Malaysia	0.1	10.5	27.3	1.6
Other	8.9	49.9	60.7	88.6

Table A4: Black Sea countries RM-transition matrix.

	BGR	GEO	ROU	RUS	TUR	UKR
BGR	0.4	0.0	0.9	0.0	0.0	0.0
GEO	0.4	1.7	0.6	0.3	0.2	0.1
ROU	4.5	0.5	19.3	0.2	0.6	1.3
RUS	1.0	2.7	1.5	60.8	0.5	8.6
TUR	66.2	85.1	39.3	23.2	90.9	6.6
UKR	27.6	9.9	38.4	15.5	7.8	83.4

Morocco, Mozambique, Namibia, Nigeria, Sao Tome and Principe, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Tanzania, Togo, Tunisia.

South Asia: Bangladesh, India, Pakistan, Sri Lanka.

East and Southeast Asia: Cambodia, Myanmar, Indonesia, Malaysia, Brunei Darussalam, Timor-Leste, Vietnam, Philippines, Thailand, China, Hong Kong SAR, China, Japan, Korea, Rep.

North and West Asia: Cyprus, Turkiye, Georgia, Iran, Islamic Rep., Israel, Lebanon, Oman, Qatar, Saudi Arabia, United Arab Emirates, Russian Federation.

Oceana: Australia, Fiji, New Zealand, Papua New Guinea, Samoa, Solomon Islands, Vanuatu.

B.1 North Atlantic Policies

GE- and NA-policies Table A5 compares the GE-policy of the North Atlantic (NA) countries (column a^*) with the γ-equilibrium policy of the NA (the *NA-policy*; column a^{NA}). The NA components of the transition matrix are given in table A2 in appendix A. We see two main differences between the GE-policy and the NA-policy. First, countries who export a large fraction of their MPP to complementary countries (e.g. Germany and Sweden) carry out much less abatement in the NA- than the GE-policy, because they are no longer internalising the externalities that they are exerting on the complementary countries (from -13.4 to -35.5% and from 54.8 to -33.2% respectively). Second, the NA countries tend to do more abatement overall in the NA-policy (79.0%) than in the GE-policy (73.0%). This is natural since the complementary countries do not internalise any externalities when they act unilaterally, and therefore do less abatement (-0.3% rather than 32.4%) — the NA countries, especially USA, compensate by doing more abatement themselves.

Table A5: North Atlantic Policies.

Country	β	T_{ii}	$\tilde{T}_{ii}$	a^*	a^{NA}	$\tilde{a}$	$v(a^*)$	$v(a^{\mathrm{NA}})$	$v(\tilde{a})$	$\tilde{v}(\tilde{a})$
BEL	0.5	1.8	1.8	-13.5	-14.5	-8.7	0.4	0.4	0.4	0.4
CAN	1.4	9.0	15.3	-4.9	1.5	10.8	0.4	0.3	-0.1	-0.2
DNK	0.3	0.4	0.4	95.5	93.2	93.4	0.2	0.2	0.2	0.2
DOM	0.1	91.5	95.6	88.1	92.0	91.9	-0.4	-0.5	-0.5	-0.6
FRA	2.6	4.0	5.6	0.3	14.3	-10.2	2.1	1.6	1.4	2.0
DEU	3.8	22.0	22.5	-13.4	-35.5	-34.3	2.6	2.7	2.6	2.7
HTI	0.0	46.8	47.5	99.4	99.6	99.6	-0.1	-0.1	-0.1	-0.1
IRL	0.3	5.3	7.0	84.8	86.1	83.0	0.2	0.1	0.1	0.2
MEX	1.8	28.8	57.2	39.1	59.3	27.0	0.9	0.1	0.4	0.7
MAR	0.2	93.5	96.9	66.8	72.7	73.6	-0.3	-0.4	-0.4	-0.4
NLD	0.8	66.0	69.5	69.9	71.0	70.6	-0.6	-0.7	-0.7	-0.7
PRT	0.3	13.1	15.8	37.8	56.1	51.5	0.2	0.2	0.2	0.2
ESP	1.5	8.3	16.7	37.4	53.0	15.9	1.2	0.5	0.6	1.1
SWE	0.5	8.7	9.4	54.8	-33.2	-34.7	0.2	0.3	0.3	0.3
GBR	2.5	46.5	52.3	14.5	18.4	14.2	1.2	1.0	0.9	1.1
USA	17.0	10.5	18.9	-1.6	22.7	-8.2	14.5	9.4	9.3	15.5
NA	33.7	70.0	100.0	73.0	79.0	72.8	22.8	15.1	14.5	22.5
RoW	66.3	97.6	—	32.4	-0.3	100.0	14.0	0.3	0.3	$-\infty$

Looking at the values, we see that the GE-policy provides more value to the NA than the NA-policy. Columns $v(a^*)$ and $v(a^{\mathrm{NA}})$ show that this is true both collectively, and for each individual country with the exception of Germany and

Sweden. This is not surprising since the GE-policy internalises the externalities that complementary countries have on NA. This result is mostly driven by USA because it receives a large fraction of MPP from Middle and South America. The reversed trend for Germany is due to the fact that it receives half of its MPP from the Netherlands which increases its abatement in the NA-policy relative to the GE-policy. The same reasoning applies to Sweden.

The m-equilibrium policy has similar levels of abatement for NA countries as the γ-equilibrium policy. Rather, the main difference between the two policies is that the rest of the world abates 31.1% of its emissions in the m-equilibrium, but increases its emissions by 0.3% under the γ-equilibrium. This is as we would expect, because non-NA countries form a coalition and internalise externalities between one another in the m-equilibrium, but not in the γ-equilibrium. As a result, NA-countries obtain higher values under the m-equilibrium than under the γ-equilibrium, but not as high as under the GE-poliy.

Global and Regional Models Beaumont et al. (2023) also estimate an efficient abatement policy for the NA region.[13] However, unlike our *global model*, their model excludes the complementary set of countries. We therefore term theirs as the *regional model*. The key difference between the regional model and the global model is that the regional model only includes NA countries in its transition matrix, which we denote by $\tilde{T}$. Since the columns of the matrix are constrained to sum to 1, $\tilde{T}$ overestimates the fraction of MPP that NA countries receive from each other. We illustrate this in columns T_{ii} and $\tilde{T}_{ii}$ of table A5. Column T_{ii} shows the estimated fraction of MPP that NA countries receive from themselves in the global model. We see that 70% of MPP in the EEZ's of NA countries comes from other NA countries (the remaining 30% comes mostly from Venezuela (10.1%), Trinidad and Tobago (5.2%) and Jamaica (4.2%)). By contrast, column $\tilde{T}_{ii}$ shows the estimates that we obtain in the regional model. By construction, this matrix assumes that 100% of MPP in NA ocean comes from NA countries. Similarly, it inflates the estimated

---

[13] Efficiency in their model corresponds to global efficiency in our model with N equal to the set of NA countries.

fractions of MPP that individual countries receive from themselves. For example, it estimates that Mexico receives 57.2% of its MPP from itself, when in fact it only receives 28.8% from itself.

This has two effects on the policy estimated by regional model (column $\tilde{a}$).[14] Firstly, it overestimates country's abatement costs though equation (9), causing countries to abate less MPP than they should — 72.8% overall, rather than 79% in the NA-policy. This is particularly extreme for USA and France, since the policy under the regional model prescribes them *negative* abatement. Second, the regional model overestimates the capacity of one NA country to abate MPP from another, and therefore overestimates the benefits of abatement. For instance, the regional model estimates that USA receives 0.12% of its MPP from Canada, whereas the global model estimates it to be only 0.06%. Thus the NA-policy prescribes Canada to abate 1.5% of its emissions, whereas the regional model policy prescribes 10.8%, despite overestimating its abatement costs by a factor of 1.7.

The policy under the regional model would be a best response for NA countries if the complementary countries abated all of their MPP. But, full abatement is not a best response to the policy under the regional model for the complementary countries (they would get a value of negative infinity). If they best respond unilaterally then they abate -0.2% of their collective emissions. We measure the inefficiency of the policy under the regional model by comparing the values that NA countries get from this policy with their values from the NA-policy (columns $v(\tilde{a})$ and $v(a^{\mathrm{NA}})$ of table A5). We see that the values are actually very similar: 15.1% vs 14.5%. We conclude that the regional model does a good job of estimating an efficient policy for the NA.

Rather, the main weakness of the regional model is that it overestimates the value of the policy under the regional model (22.5%; column $\tilde{v}(\tilde{a})$), relative to the scenario where non-NA countries best respond unilaterally (14.5%; column $v(\tilde{a})$). This is because it overestimates the impact that the policy has on the fraction of abatement

---

[14] We cannot replicate their results exactly because the data they use to estimate their backward probability matrix is not in the public domain at the time of writing. We therefore only produce the results that we obtain from applying their model to our own data.

received, especially for USA. This bias is likely to have further knock on effects for deriving fair and stable transfers.

B.2 Fair and Stable Agreements for China, Indonesia and Malaysia

We compare the m-nucleolus of the characteristic function restricted to the coalition formed by China (C), Indonesia (I) and Malaysia (M) with the Myerson value of the same coalition obtained by Zhao et al. (2023). They find that the Myerson value distributes 49.2% of the gains from cooperation to China, 29.8% to Malaysia and 21% to Indonesia. By contrast, the m-nucleolus of our restricted game distributes 31.3% to Malaysia, 61.6% to Indonesia, and only 7.1% to China. These stark differences can be traced back to the underlying characteristic functions. In Zhao et al. (2023), the $\{C, I\}$ and $\{C, M\}$ coalitions respectively generate a 39% and 56% of the value of the $\{C, I, M\}$ coalition, whereas the $\{I, M\}$ coalition generates no value at all. In our characteristic function, the $\{I, M\}$ coalition generates 94% of the value of the $\{C, I, M\}$ coalition, whereas the $\{C, I\}$ and $\{C, M\}$ coalitions both generate *negative* values equal to -1% of the $\{C, I, M\}$ coalition value.

These differences are easily explained by the different modelling approaches. There are no negative externalities in Zhao et al. (2023), rather coalitions create value by sharing their abatement technologies. China has much better technology than Malaysia and Indonesia, so it generates a lot of extra value by forming coalitions with them. Malaysia and Indonesia have similar technologies, so they cannot create any extra value without including China in the coalition. It is therefore not surprising that China receives a large allocation under the Myerson value. We do not model the sharing of technology — the only cost saving comes from the fact that MPP emitters may be able to reduce emissions for a receiving country more cheaply than the receiving country itself. For instance, Indonesia receives 10.5% of its MPP from Malaysia and its marginal abatement cost is more than four times higher than Malaysia's (see table A3 in appendix A). Thus the benefit that Indonesia obtains from a marginal increase in Malaysia's abatement choice exceeds Malaysia's marginal costs, thereby generating value for the coalition. China does not benefit from forming a two-player coalition with either Indonesia or Malaysia because it receives only 0.1% of its total MPP from each of them. Indonesia and Malaysia both receive a more significant fraction of their MPP from China (3.1% and 6.3% respectively), but China's marginal abatement cost is around 18 times higher than Indonesia's, and 88 times higher than Malaysia's, so there is little scope for efficient abatement by China. Thus the m-nucleolus assigns a negligible value to China. These contrasting results demonstrate the importance of taking a holistic approach that internalises both negative externalities from the transboundry nature of MPP, and positive externalities of technology sharing.