Group Decision and Negotiation
https://doi.org/10.1007/s10726-024-09878-w

Coalition Stability in International Environmental Matching Agreements

Charlotte Süring[1] · **Hans-Peter Weikard**[2]

Accepted: 11 February 2024

Abstract

This study presents empirically calibrated simulations of three different variants of environmental matching agreements aimed at reducing global greenhouse gas emissions. We determine whether matching agreements can produce larger stable coalitions and increase abatement contributions and payoffs as compared to standard agreements. The matching agreements we analyze feature uniform matching rates by which coalition members match the unconditional contributions of (i) the other coalition members, (ii) all other players, or (iii) only non-members, while non-members do not commit to any matching and maximize their individual payoffs. The simulation considers twelve asymmetric world regions with linear abatement benefits and quadratic costs, calibrated based on the STACO 3 model, and uses emissions data from the shared socioeconomic pathways database. We find that the first variant of the matching game fails to produce any stable coalitions and thus performs worse than the standard agreement that produces a stable two-player coalition. The second variant produces a stable grand coalition and significantly increases the abatement and payoff levels beyond the non-cooperative Nash baseline. Partial coalitions are unstable in this game. The third variant produces a two-player coalition similar to the standard coalition formation game, but with different members and higher abatement and payoff levels due to the matching mechanism.

Keywords Matching games · International environmental agreements · Coalition formation and stability · Global public goods · STACO model

✉ Charlotte Süring
cs@ifro.ku.dk

[1] Department of Food and Resource Economics, University of Copenhagen, Frederiksberg, Denmark

[2] Department of Social Sciences, Wageningen University, Wageningen, The Netherlands

Published online: 23 March 2024

1 Introduction

As with many other transboundary environmental problems, greenhouse gas (GHG) abatement is a matter of public good provision, characterized by the lack of a central enforcement authority. In such cases, ill-defined property rights create an incentive for free-riding which is reinforced by the lack of credible punishment for non-cooperation. The continuing underprovision of national efforts to reduce global GHG emissions and the resulting failure of international environmental agreements (IEAs) is therefore all but surprising. In fact, Barrett (1994) demonstrated that cooperative behaviour in large coalitions of (symmetric) parties is inherently unlikely when the gains from cooperation are large.

One mechanism that has been put forward to address this issue of collective action without the introduction of a supranational authority is Guttman's (1978) matching game. In this game, countries commit to supplementing other countries' emission reductions by a proportional matching contribution, determined by a predefined matching rate, before choosing their own flat contribution which is matched by the other countries (Guttman 1978). The mechanism has been shown to produce agreements that are Pareto-efficient and, therefore, increase abatement and collective welfare beyond the Nash equilibrium outcome that is obtained in the absence of an agreement (e.g. Guttman 1978; Guttman and Schnytzer 1992; Rübbelke 2006; Boadway et al. 2011; Fujita 2013; Molina et al. 2020).

Under the Kyoto Protocol, such matching contributions were announced by the EU and Australia, who indicated their willingness to increase their GHG mitigation efforts conditional upon collective increases in international climate ambitions (UNFCCC 2012). These conditional contributions were additional to the two countries' unconditional mitigation pledges. Although the framework of the Paris agreement would lend itself to such matching commitments in the form of conditional Nationally Determined Contributions (NDCs), current conditional contributions announced by member states are conditioned on financial and technological requirements (UNFCCC 2023), not on other countries' contributions.[1] The implementation of a matching mechanism therefore remains a consideration for future climate negotiations.

Most studies on matching agreements so far are theoretical and many are based on the assumption of homogeneous players (e.g. Guttman 1978; Rübbelke 2006; Boadway et al. 2007; Fujita 2013). This assumption is problematic not only because it does not reflect reality, but also because player heterogeneity can drastically change game outcomes due to the possibility of exploiting inexpensive abatement options or payoff transfers; see e.g., Mäler (1989) or Weikard (2009). Moreover, prior studies commonly consider matching agreements in the form of two-stage games where all

[1] As part of the COP21 negotiations in 2015, some countries indicated that they would make the full implementation of their NDCs conditional on "the level of effort undertaken by other Parties" (UNFCCC 2015). However, the required level of ambition by other countries was not specified in any of these cases (Day et al. 2016). To our knowledge, no such conditions have been announced in more recent negotiations.

players are assumed to accede to the agreement ab initio. A more realistic scenario would allow players to make autonomous decisions on whether to join or abstain from an agreement.

To date, a single study has investigated a numerical application of a matching mechanism in the context of IEAs: Kawamata and Horita (2014) simulated a matching game with six regions based on the stability of coalitions (STACO) model's calibration of benefit and cost functions. They found that the matching mechanism produces significantly higher abatement and collective welfare than what is obtained in the absence of an agreement. Although this is an encouraging result, the study does not demonstrate whether the matching agreement is stable in the sense that no region has an incentive to opt out.

The present study thus sets out to analyze matching games between heterogeneous players with an initial coalition accedence stage. We assume that coalition members negotiate a uniform collective payoff-maximizing matching rate by which they match other players' abatements, while non-members follow Nash strategies. A comparison of the matching agreements' equilibrium abatement and payoff levels with those of the no-agreement Nash baseline and the socially optimal grand coalition is used to indicate the matching mechanisms' potential to overcome the free-rider problem in IEAs.

The remainder of this paper is organized as follows. Section 2 provides an overview of the most relevant game-theoretic literature on coalition stability and matching agreements. Section 3 outlines the mathematical models. Section 4 presents the calibrations used for our simulations and the underlying data sources. Section 5 presents the simulation results and Sect. 6 discusses their implications. Section 7 concludes.

2 Literature Review

Hoel (1992), Carraro and Siniscalco (1993) and Barrett (1994) were likely the first to model IEAs as cartel formation games in an effort to explain the widespread failure of such agreements. While Barrett (1994) analyzed a Stackelberg abatement game between symmetric players, Hoel (1992) and Carraro and Siniscalco (1993) investigated emissions games in which the symmetric players act simultaneously. In these games, environmental agreements are equivalent to coalitions in which members determine their contributions cooperatively while the decision to join the coalition is made in a non-cooperative manner by each player. An agreement is thus considered stable when coalition members have no incentive to leave (internal stability), and non-members have no incentive to join (external stability) the coalition. In general, the literature studying the stability of IEAs according to these conditions seems to agree that large and effective agreements are highly unlikely due of strong free-rider incentives.

One mechanism that promises to be effective at increasing contributions by players behaving non-cooperatively was proposed by Guttman (1978). In his matching game, players commit to reciprocating other players' contributions according to a self-determined matching rate before choosing their own unconditional

contributions. Barrett (1990) was possibly the first to suggest applying the matching mechanism in the context of GHG abatement. To our knowledge, Rübbelke (2006) was the first to study matching contributions in a coalitional GHG abatement game.

Much of the literature on matching schemes adopts standard public goods or externality formulations, including Guttman and Schnytzer (1992), Rübbelke (2006), Buchholz et al. (2011; 2012; 2014), Liu (2018; 2019), and Buchholz and Liu (2020). Our study is closest to Liu (2018) who analyzes the profitability of matching coalitions using an aggregative game approach and examines their stability according to the internal and external stability conditions. The game he analyzes allows for multiple heterogeneous players and features an exogenously determined, uniform matching rate. Coalition members only match other coalition members' flat contributions, while non-members have no matching commitment and only abate to maximize their individual payoffs. Liu (2018) finds that the matching mechanism can produce Pareto-improving outcomes, but that stability conditions are hard to satisfy. In particular, larger matching rates produce more profitable but less stable coalitions and vice versa. He also notes that stability issues may be overcome by introducing a reputation mechanism.

Other literature features models that are similar to the emissions game delineated by Finus (2001) or its abatement game analogue. In two such studies, Fujita (2013) and Wood and Jotzo (2015) examine the stability of matching coalitions with symmetric players, where coalition members not only match each other's contributions, but also non-members' abatements. In Fujita's (2013) study, coalition members negotiate a common matching rate by which all players' abatements are matched. Here, the grand coalition is found to be stable, and the agreement is thus 'self-enforcing' and efficient. In Wood and Jotzo (2015), matching rates are determined non-cooperatively by each coalition member. Since all players are assumed to have identical abatement benefit and cost functions (quadratic ones, in this case), all matching rates towards coalition members are the same, as are the matching rates towards non-members. Wood and Jotzo (2015) observe that the matching rate towards non-members increases as more players join the coalition, while the matching rates towards members approach zero. The authors conclude that the matching game produces multiple stable coalitions, including the grand coalition.

As the first theoretical analysis of an emissions game with a matching mechanism between multiple heterogeneous players, the study by Boadway et al. (2011) serves an important point of reference for our study. Their game considers players with generalized benefit and cost functions that match each other's emissions reductions at different, non-cooperatively determined rates. Boadway et al. (2011) show that the matching mechanism leads to emissions reductions relative to the baseline and that the subgame perfect equilibrium is Pareto-efficient and unique. The game is also extended to include an emissions trading stage leading to an optimal allocation of contributions among players.

Practical applications of the matching mechanism are still scarce in the literature. In the IEA literature, Kawamata and Horita's (2014) simulation study is so far the only numerical application of a matching game that we are aware of. Their study demonstrates that the matching mechanism as introduced by Boadway et al. (2011) induces higher abatement levels and collective welfare than the Nash baseline case

without an agreement. Our paper builds on this literature but also examines whether matching coalitions are stable, which ultimately determines the effectiveness of matching agreements in practice.

3 Matching Models

The main aim of this paper is to assess the effectiveness of the matching mechanism in the context of IEAs. To do so, we conduct simulations with different variants of environmental matching agreements between multiple asymmetric players and investigate whether the matching mechanism can increase players' abatements and payoffs as compared to a game without matching. Our analysis grants players the possibility to opt out of the respective agreements, allowing us to draw conclusions about the agreements' stability. In particular, we compare the outcome of the standard game of coalition formation (Sect. 3.1) to those of matching games in which coalition members commit to (i) matching other coalition members' flat abatements (Sect. 3.2.1), (ii) matching all other players' flat abatements (Sect. 3.2.2) and (iii) only matching non-members' abatements (Sect. 3.2.3). A comparison of each agreement's abatement level and payoff with those of the no-agreement Nash baseline and the fully cooperative ('grand coalition') outcome will give an indication of the respective agreement variants' effectiveness in overcoming potential free-rider incentives.

3.1 Stable Coalitions in the Standard IEA Game

The baseline model of coalition formation without matching considers a two-stage, non-cooperative game between n heterogeneous countries or regions that generate damaging GHG emissions; see Hagen et al. (2020) for a recent survey. Emissions are assumed to be perfect substitutes in each region's damage function. Let $N = \{1, \ldots, n\}$ denote the set of regions. We write our model as an abatement game where the abatement of emissions is a global public good. Each region incurs region-specific costs from their own abatement efforts q_i and region-specific benefits from collective abatement $Q = \sum_{j \in N} q_j$. In the present study, we consider linear benefit and quadratic cost functions that monotonously increase in abatement. Region i's payoff π_i is given by

$$\pi_i = b_i \sum_{j \in N} q_j - \frac{1}{2} c_i q_i^2, \tag{1}$$

where b_i and c_i are region-specific abatement benefit and cost parameters.

In the first stage of the game, regions decide whether or not to accede to an agreement and join a unique coalition $S \subseteq N$, the set of signatories. We assume that the agreement is an 'open-membership agreement', meaning that any region can freely join without permission of other coalition members (Finus et al. 2005). Regions that join, the members, commit to abatement strategies which maximize the coalition's total payoff, while non-members remain singleton players who set their

contributions non-cooperatively. In the second stage, the abatement game, both the coalition, acting jointly, and the singletons are assumed to play a Nash game and have dominant strategies due to the linear benefits of abatement, meaning that their actions are independent of the actions by other players. Any singleton player i thus selects its abatement level q_i such that $\frac{\partial \pi_i}{\partial q_i} = 0$. Given the linear benefits and quadratic costs, we have

$$q_i = \frac{b_i}{c_i}, i \notin S. \tag{2}$$

Coalition members cooperate to maximize the sum of all coalition members' payoffs. Each signatory thus sets its abatement level q_i such that its marginal cost of abatement is equal to the sum of all members' marginal benefits, i.e.,

$$q_i = \frac{\sum_{j \in S} b_j}{c_i}, i \in S. \tag{3}$$

In stage 1, when regions decide to join the agreement or not, members must reap higher payoffs than what they would incur as singletons. Similarly, non-members must reap higher payoffs than if they joined the agreement. Put differently, a Nash equilibrium in stage 1 is characterized by the following conditions:

$$\pi_i(S) \geq \pi_i(S \setminus \{i\}), i \in S, \tag{4}$$

$$\pi_i(S) \geq \pi_i(S \cup \{i\}), i \notin S. \tag{5}$$

These conditions are commonly referred to as internal and external stability conditions, respectively (Carraro and Siniscalco 1993).

With n players, there are $2^n - n$ possible coalitions. The coalition that is formed when all players join, $S = N$, is called the 'grand coalition' and results in the globally optimal level of abatement and an equalization of all players' marginal abatement costs.

3.2 International Environmental Matching Agreements

This section considers IEAs that adopt a matching mechanism. The basic assumptions about players' abatement benefits and costs are retained, but the second stage of the standard game is replaced by a matching game.

The matching game itself consists of two stages that will be preceded by a coalition formation stage in our analysis. In the first stage of the matching game, the coalition members negotiate a common matching rate m by which every member $i \in S$ must match the unconditional 'flat' abatements determined in the following stage. Since, ideally, collaborative agreements aspire to achieve a collective optimum, we assume joint payoff maximization by the cooperating players, i.e., the matching rate

is selected such that it maximizes the collective payoff of the coalition.[2] After the negotiation of the matching rate, players set their unconditional 'flat' contributions a_i non-cooperatively, considering the matching rate announced in the previous stage. In addition to these unconditional flat contributions, members abate according to their conditional matching commitments. Player i's total abatement contribution q_i is thus

$$q_i = a_i + m \sum_{j \in M} a_j, i \in S, \tag{6}$$

where M is the set of regions whose flat contributions a_j are matched. The matching mechanism signals to matched players that every unit of flat abatement they provide will trigger additional abatements in the form of matching contributions by the coalition members. This incentivizes larger contributions beyond the Nash equilibrium abatement levels by increasing the marginal benefits associated with every flat contribution. We consider three matching mechanism variants. The coalition matches the flat contributions of (i) only fellow members, (ii) all other regions, and (iii) only singleton regions. In summary, we consider the following stages.

1. *Coalition formation* Players non-cooperatively decide to join the matching agreement or to remain singleton players. As in the standard game, we consider an open-membership agreement that can freely be joined by any player.
2. *Negotiation of the matching rate* Coalition members cooperatively set a common matching rate by which every coalition member must match the flat abatements determined in stage 3.
3. *Abatement* All players choose their flat contributions non-cooperatively. In addition, members abate according to their matching commitment while non-members do not match other players' contributions.

3.2.1 Variant 1: Members Match Members

To determine whether matching agreements can effectively curb the general freeriding problem of IEAs, they must be examined in a context of voluntary participation. For the first matching variant, we adopt Fujita's (2013) matching rule in which coalition members only match other members' flat contributions. We generalize Fujita's analysis by allowing for heterogeneous players.

The remainder of this section reports the analytical solution of the game. Payoffs are defined by Eq. 1 as before. Members' total abatements are given by Eq. 6, with the specification $M = S \setminus \{i\}$. Let s denote the number of signatories of coalition $S \subseteq N$. Note then that every members' flat contribution is matched by $s - 1$ other members. We can rewrite the payoffs as

[2] This assumption is customary in the IEA literature, see also the baseline model in Sect. 3.1. However, it should be noted that if players can default on their conditional matching commitments, i.e., match at a smaller rate than the negotiated matching rate, the free-riding problem reemerges. Additional sanctioning mechanisms may be needed to curb this risk in practice.

$$\pi_i = \begin{cases} b_i(1+m(s-1))\sum\limits_{j\in S} a_j + b_i \sum\limits_{j\notin S} a_j - \frac{1}{2}c_i\left(a_i + m\sum\limits_{j\in S\setminus\{i\}} a_j\right)^2 & (i \in S) \\ b_i(1+m(s-1))\sum\limits_{j\in S} a_j + b_i \sum\limits_{j\notin S} a_j - \frac{1}{2}c_i a_i^2 & (i \notin S). \end{cases} \tag{7}$$

We solve the game by backward induction.

Stage 3: Choice of Flat Abatements At this stage, all players have made their choices on whether to join the agreement or not, and the coalition members have negotiated a matching rate. All players now choose their flat contributions non-cooperatively. As the singleton players have dominant strategies, their flat abatements are equivalent to the Nash singleton contributions in the standard game of coalition formation (Eq. 2):

$$a_i = q_i = \frac{b_i}{c_i}, i \notin S. \tag{8}$$

Assuming an interior solution coalition, members $i \in S$ maximize their payoffs defined by Eq. 7. Since non-signatories have dominant strategies, their abatements do not show up in the first order condition for coalition members:

$$\frac{\partial \pi_i}{\partial a_i} = 0 = b_i(1+m(s-1)) - c_i\left(a_i + m\sum_{j\in S\setminus\{i\}} a_j\right). \tag{9}$$

Solving for a_i yields

$$a_i = \frac{b_i}{c_i}(1+m(s-1)) - m\sum_{j\in S\setminus\{i\}} a_j \tag{10}$$

which means that

$$q_i = \frac{b_i}{c_i}(1+m(s-1)), i \in S. \tag{11}$$

In equilibrium, the coalition members will choose their flat contributions such that each members' payoff function is maximized. We can also show that the sum of all coalition members' equilibrium flat contributions is equivalent to the sum of their would-be singleton abatements. Let $\mu = 1 + m(s-1)$. Then, from Eq. 11, summing both sides of the equation over the set of members, we have

$$\sum_{i\in S} q_i = \mu \sum_{i\in S} \frac{b_i}{c_i} \tag{12}$$

From Eq. 10, summing over all members, we obtain

$$\sum_{i\in S} a_i = \mu \sum_{i\in S} \frac{b_i}{c_i} - m(s-1)\sum_{i\in S} a_i \tag{13}$$

and hence,

$$\mu \sum_{i\in S} a_i = \mu \sum_{i\in S} \frac{b_i}{c_i}. \tag{14}$$

This result implies that the existence of a stable matching coalition inevitably increases the global abatement level beyond that of the no-agreement baseline, provided that the matching rate is positive. Since the sum of flat abatements is equal to the sum of the non-cooperative Nash abatement levels, the sum of all coalition members' matching contributions is equivalent to the additional global abatement above the baseline that is induced by the matching agreement. Each member i's individual flat contribution can be obtained from 10 and 14 as

$$a_i = \frac{1}{1-m}\left(\mu\frac{b_i}{c_i} - m\sum_{j\in S}\frac{b_j}{c_j}\right). \tag{15}$$

Stage 2: Negotiation of the Matching Rate In this stage, coalition members negotiate a common matching rate that maximizes the coalition's total payoff, taking the stage 3 equilibrium abatements as given. Disregarding non-members' abatements as they have dominant strategies, the coalition's collective payoff function is:

$$\pi_S = \sum_{j\in S} b_j \sum_{k\in S} q_k - \sum_{j\in S}\left(\frac{1}{2}c_j q_j^2\right). \tag{16}$$

With $q_i = \mu\frac{b_i}{c_i}$ (Eq. 11) from stage 3, we have

$$\pi_S = \sum_{j\in S} b_j \sum_{k\in S}\frac{b_k}{c_k}\mu - \sum_{j\in S}\left(\frac{1}{2}c_j\left(\frac{b_j}{c_j}\mu\right)^2\right). \tag{17}$$

With the partial derivative of μ with respect to m being

$$\frac{\partial\mu}{\partial m} = s-1 \tag{18}$$

maximizing π_S yields

$$\frac{\partial\pi_S}{\partial m} = 0 = \sum_{j\in S} b_j \sum_{k\in S}\frac{b_k}{c_k}(s-1) - \sum_{j\in S}\left(\frac{b_j^2}{c_j}(s-1)\mu\right) \tag{19}$$

$$\implies 0 = \sum_{j\in S} b_j \sum_{k\in S}\frac{b_k}{c_k}(s-1) - \sum_{j\in S}\left(\frac{b_j^2}{c_j}(s-1)\right)(1+m(s-1)). \tag{20}$$

Solving for m gives the optimal matching rate

$$m = \frac{\sum_{j\in S} b_j \sum_{k\in S} \frac{b_k}{c_k}}{\sum_{j\in S} \frac{b_j^2}{c_j}(s-1)} - \frac{1}{(s-1)} \tag{21}$$

for every given coalition S.

Stage 1: Coalition Formation In this stage, all players non-cooperatively decide whether to join the agreement and commit to matching or to remain singletons. Anticipating the equilibrium outcomes of the subsequent stages, each player compares the payoff it would reap as a coalition member to the payoff it would get as a singleton.

Formally, the equilibrium conditions are the internal stability condition (Eq. 4: $\pi_i(S) \geq \pi_i(S\backslash\{i\}), i \in S$) and the external stability condition (Eq. 5: $\pi_i(S) \geq \pi_i(S \cup \{i\}), i \notin S$). As a signatory, player i anticipates the subgame perfect Nash equilibrium of the matching game with $q_i = \mu\frac{b_i}{c_i}$ (Eq. 11). As a non-signatory, i plays a Nash strategy with $q_i = a_i = \frac{b_i}{c_i}$ (Eq. 8). The solution of this stage identifies stable coalitions S which satisfy both stability conditions. While Fujita (2013) solves this stage analytically for a matching game with symmetric players, analytical solutions are difficult to obtain for games with heterogeneous players. We therefore resort to examining coalition stability of this agreement numerically by means of a simulation (Sect. 5.2). If less than two players join the agreement, matching in the second stage becomes irrelevant and the standard abatement game without matching is played.

3.2.2 Variant 2: Members Match All Other Players

In this variant of the matching game, we assume that coalition members not only match other members' flat abatements, but those of all other players. That is, we specify $M = N \setminus \{i\}$. To analyse the game, we follow the same steps as in the previous subsection. The payoff to player $i \in S$ or $i \notin S$ is given by

$$\pi_i = \begin{cases} b_i(1+m(s-1))\sum_{j\in S} a_j + b_i(1+ms)\sum_{j\notin S} a_j - \frac{1}{2}c_i\left(a_i + m\sum_{j\in N\setminus\{i\}} a_j\right)^2 & (i \in S) \\ b_i(1+m(s-1))\sum_{j\in S} a_j + b_i(1+ms)\sum_{j\notin S} a_j - \frac{1}{2}c_i a_i^2 & (i \notin S). \end{cases} \tag{22}$$

Stage 3: Choice of Flat Abatements At this stage, all players have made their choices on whether to join the agreement or opt out, and the coalition members have

negotiated the matching rate. All players now choose their flat contributions non-cooperatively, so as to maximize their respective payoffs.

$$\frac{\partial \pi_i}{\partial a_i} = 0 = \begin{cases} b_i(1 + m(s-1)) - c_i\left(a_i + m \sum\limits_{j \in N \setminus \{i\}} a_j\right) & (i \in S) \\ b_i(1 + ms) - c_i a_i & (i \notin S). \end{cases} \tag{23}$$

Solving for a_i gives

$$a_i = \begin{cases} \frac{b_i}{c_i}(1 + m(s-1) - m \sum\limits_{j \in N \setminus \{i\}} a_j & (i \in S) \\ \frac{b_i}{c_i}(1 + ms) & (i \notin S). \end{cases} \tag{24}$$

Setting $\mu = 1 + m(s-1)$ and $\eta = 1 + ms$, we have

$$a_i = \begin{cases} \frac{b_i}{c_i}\mu - m \sum\limits_{j \in N \setminus \{i\}} a_j & (i \in S) \\ \frac{b_i}{c_i}\eta & (i \notin S) \end{cases} \tag{25}$$

and

$$q_i = \begin{cases} \frac{b_i}{c_i}\mu & (i \in S) \\ \frac{b_i}{c_i}\eta & (i \notin S). \end{cases} \tag{26}$$

Since we can assume $s \geq 1$ and we have $\eta > \mu$, matching ($m > 0$) incentivizes not just additional abatement but singletons abate more than members with similar benefit and cost functions.

Stage 2: Negotiation of the Matching Rate As in the previous variant, all members to the matching agreement negotiate a common matching rate that maximizes the coalition's total payoff, taking the stage 3 equilibrium abatements as given. We can write the coalition payoff as

$$\pi_S = \sum_{j \in S} b_j \left(\sum_{k \in S} q_k + \sum_{k \notin S} q_k \right) - \sum_{j \in S} \left(\frac{1}{2} c_j q_j^2 \right) \tag{27}$$

Using results from stage 3 (Eq. 26) we have

$$\pi_S = \sum_{j \in S} b_j \left(\sum_{k \in S} \frac{b_k}{c_k}\mu + \sum_{k \notin S} \frac{b_k}{c_k}\eta \right) - \sum_{j \in S} \left(\frac{1}{2} c_j \left(\frac{b_j}{c_j}\mu \right)^2 \right) \tag{28}$$

With the partial derivatives

$$\frac{\partial \mu}{\partial m} = s - 1 \tag{29}$$

and

$$\frac{\partial \eta}{\partial m} = s \tag{30}$$

we maximize π_S. This gives first order conditions as follows

$$\frac{\partial \pi_S}{\partial m} = 0 = \sum_{j \in S} b_j \left(\sum_{k \in S} \frac{b_k}{c_k}(s-1) + \sum_{k \notin S} \frac{b_k}{c_k} s \right) - \sum_{j \in S} \left(\frac{b_j^2}{c_j}(s-1)\mu \right) \tag{31}$$

$$= \sum_{j \in S} b_j \left(\sum_{k \in S} \frac{b_k}{c_k}(s-1) + \sum_{k \notin S} \frac{b_k}{c_k} s \right) - \sum_{j \in S} \left(\frac{b_j^2}{c_j}(s-1) \right)(1 + m(s-1)) \tag{32}$$

$$\implies m = \frac{\sum_{j \in S} b_j \sum_{k \in S} \frac{b_k}{c_k}}{\sum_{j \in S} \frac{b_j^2}{c_j}(s-1)} + \frac{\sum_{j \in S} b_j \sum_{k \notin S} \frac{b_k}{c_k} s}{\sum_{j \in S} \frac{b_j^2}{c_j}(s-1)^2} - \frac{1}{s-1} \tag{33}$$

for every given coalition S.

Stage 1: Coalition Formation As in the previous variant, in this stage, all players non-cooperatively decide whether to join the agreement and commit to the matching or to opt out and play their Nash strategy. Anticipating the equilibrium outcomes of the subsequent stages, each player evaluates whether to join or stay out of the agreement. An agreement is considered stable if it fulfills the internal and external stability conditions (Eqs. 4 and 5). We solve this stage numerically by means of a simulation in Sect. 5.3.

3.2.3 Variant 3: Members Match Non-members

In the last matching game variant, coalition members only match non-members' flat abatements. That is, we specify $M = N \backslash \{S\}$. In the grand coalition, no outsiders are left whose flat contributions can be matched, $M = 0$. Thus, when all players join the coalition, all abatements are the non-cooperative Nash abatements. This variant is similar to the 'unilateral' matching game that Buchholz and Liu (2020) analyze. We assume that members' flat contributions are non-negative $a_i \geq 0$, such that that their total abatements cannot be smaller than their matching contributions.[3] The payoff to player $i \in S$ or $i \notin S$ is given by

[3] If coalition members were to set their flat contributions freely, they would set the matching rate m as large as possible and resort to setting flat contributions such that their total contributions equal Nash baseline abatements $q_i = \frac{b_i}{c_i}$. We therefore constrain $a_i \geq 0$ for $i \in S$. Buchholz and Liu (2020) call this a "commitment device" and assume that $a_i = \frac{b_i}{c_i}$.

$$\pi_i = \begin{cases} b_i \sum_{j\in S} a_j + b_i(1+ms) \sum_{j\notin S} a_j - \frac{1}{2} c_i \left(a_i + m \sum_{j\notin S} a_j \right)^2 & (i \in S) \\ b_i \sum_{j\in S} a_j + b_i(1+ms) \sum_{j\notin S} a_j - \frac{1}{2} c_i a_i^2 & (i \notin S). \end{cases} \tag{34}$$

Stage 3: Choice of Flat Abatements At this stage, all players have made their choices on whether to join the agreement or opt out, and the coalition members have negotiated the matching rate. Non-members now choose their flat contributions non-cooperatively. The first order condition gives

$$\frac{\partial \pi_i}{\partial a_i} = 0 = b_i(1+ms) - c_i a_i \quad (i \notin S) \tag{35}$$

$$\Longrightarrow \; a_i = q_i = \frac{b_i}{c_i}(1+ms) \quad (i \notin S). \tag{36}$$

Coalition members have to fulfill their commitment but choose a zero flat contribution as long as their marginal abatement costs are larger than their marginal benefits, i.e. whenever their commitment $m \sum_{j\notin S} \frac{b_j}{c_j}(1+ms)$ exceeds what they would abate as a singleton $\frac{b_i}{c_i}$. We then have

$$q_i = max\left(\frac{b_i}{c_i}, m \sum_{j\notin S} \frac{b_j}{c_j}(1+ms) \right) (i \in S). \tag{37}$$

The analysis of stage 2 below shows that m is always chosen sufficiently high such that the commitment exceeds the singleton abatement.

Stage 2: Negotiation of the Matching Rate As in the previous variants, all members of the matching agreement negotiate a common matching rate that maximizes the coalition's total payoff, anticipating the stage 3 equilibrium abatements.

$$\pi_S = \sum_{j\in S} b_j \left(\sum_{k\in S} q_k + \sum_{k\notin S} q_k \right) - \sum_{j\in S} \frac{1}{2} c_j q_j^2 \tag{38}$$

Using Eqs. 36 and 37, this is equivalent to

$$\pi_S = \sum_{j\in S} b_j \left(sm \sum_{k\notin S} \frac{b_k}{c_k}(1+ms) + \sum_{k\notin S} \frac{b_k}{c_k}(1+ms) \right) - \sum_{j\in S} \frac{1}{2} c_j \left(m \sum_{k\notin S} \frac{b_k}{c_k}(1+ms) \right)^2 \tag{39}$$

or

$$\pi_S = \sum_{j\in S} b_j \sum_{k\notin S} \frac{b_k}{c_k}(sm+1)^2 - \sum_{j\in S} \frac{1}{2} c_j \left(\sum_{k\notin S} \frac{b_k}{c_k} \right)^2 (sm^2+m)^2. \tag{40}$$

Maximizing π_S with respect to m gives

$$\frac{\partial \pi_S}{\partial m} = 0 = \sum_{j\in S} b_j \sum_{k\notin S} \frac{b_k}{c_k} 2s(sm+1) - \sum_{j\in S} c_j \left(\sum_{k\notin S} \frac{b_k}{c_k} \right)^2 m(sm+1)(2sm+1) \tag{41}$$

$$\implies \sum_{j\in S} b_j 2s = \sum_{j\in S} c_j \sum_{k\notin S} \frac{b_k}{c_k} m(2sm+1) \tag{42}$$

$$\implies \frac{\sum_{j\in S} b_j 2s}{\sum_{j\in S} c_j \sum_{k\notin S} \frac{b_k}{c_k}} = 2sm^2 + m \tag{43}$$

$$\implies 0 = m^2 + \frac{1}{2s} m - \frac{\sum_{j\in S} b_j 2s}{\sum_{j\in S} c_j \sum_{k\notin S} \frac{b_k}{c_k}} \tag{44}$$

$$\implies m = -\frac{1}{4s} + \sqrt{\frac{1}{16s^2} + \frac{\sum_{j\in S} b_j}{\sum_{j\in S} c_j \sum_{k\notin S} \frac{b_k}{c_k}}}. \tag{45}$$

Substituting Eq. 45 into Eq. 37, we can show that the matching commitment is always larger than the singleton abatement level.

Stage 1: Coalition Accedence As in the previous variants, we solve this stage numerically by means of a simulation (Sect. 5.4).

4 Data and Simulation Model Calibration

To generate numerical results for the different matching variants introduced in Sects. 3.2.1 to 3.2.3, as well as the non-cooperative Nash baseline and the grand coalition outcomes, we calibrate a general simulation model consisting of estimates of regional abatement benefit and cost functions. As the data underlying the different simulations are the same, their results are directly comparable and can be used to draw conclusions about the agreements' potential to overcome the general freeriding problem of IEAs.

Table 1 Regions, model notation, and benefit and cost parameters

Region	Model notation	b_i	c_i
USA	USA	48.886	0.0239
Japan	JPN	6.415	0.1315
EU27 & EFTA	EUR	62.022	0.0387
Other High Income	OHI	1.824	0.0956
Rest of Europe	ROE	1.890	0.0985
Russia	RUS	1.923	0.0448
High Income Asia	HIA	5.649	0.0777
China	CHN	21.176	0.0080
India	IND	18.633	0.0386
Middle East	MES	2.877	0.0532
Brazil	BRA	0.784	0.1343
Rest of the World	ROW	53.398	0.0498

Source: Own calibration based on Dellink et al. (2015)

4.1 The Stability of Coalitions (STACO) Model

The simulation data are calibrated based on the third version of the stability of coalitions (STACO) model (Dellink et al. 2015). STACO 3 is a dynamic, multi-region model which serves to investigate the formation and stability of international climate agreements in the time frame from 2010 to 2110. It considers a two-stage, non-cooperative game of coalition formation and GHG abatement (as described in Sect. 3.1) between twelve world regions with linear abatement benefits and cubic abatement costs. Detailed specifications of STACO 3 can be found in the technical manual by Dellink et al. (2015). Lessmann et al. (2015) have shown that STACO 3 results are comparable to other integrated assessment models that explore the stability and performance of international climate coalitions.

4.2 Simulation Model Set-up

The current study's general simulation model is a static representation of the dynamics that are inherent in the STACO 3 model. This simplification was made to increase the model's conceptual fit with the one-shot matching game as presented in the literature and to reduce its emphasis on the complex growth-climate interactions. The procedure we followed to calibrate the static model is described in Appendices C and D in Supplementary material. We include the different players' cost and benefit parameters in Table 1. In the simulations, players make a decision on their accedence to the analyzed agreement before 2020, and then set their abatement strategies for the time period from 2020 up to and including 2100. While the duration of the agreement and players' abatement strategies are limited to 80 years, the time horizon for the benefits of abatement extends beyond the agreement term to reflect the long-term impacts of climate change. The calibration of our static game presents annual costs and benefits as the payoff space. STACO 3 comprises the twelve regions presented in Table 1 and Appendix A in Supplementary material.

4.3 GDP, Population and Emissions Data

The baseline data underlying the STACO 3 model (Dellink et al. 2015) are updated for the simulations of the current study. We use more recent GDP, population and business as usual (BAU) GHG emissions projection estimates from the Shared Socioeconomic Pathways (SSP) database (IIASA 2018). The data were retrieved from the second version of the SSP database, which contains data on five possible trajectories of global socioeconomic and climatic development in the 21st century. The different scenarios each consist of a narrative of future development which is substantiated with quantitative data from economic, demographic and integrated assessment models (IAMs) (Riahi et al. 2017). Specifically, we use baseline emissions data from the SSP2 'marker' model MESSAGE-GLOBIOM, economic development data from Dellink et al.'s (2017) SSP2 marker model and population projections based on Samir and Lutz (2017). SSP2 describes a "middle-of-the-road" scenario in which "social, economic, and technological trends do not shift markedly from historical patterns" (Riahi et al. 2017). We use data from this scenario to avoid incorporating assumptions about drastic future socioeconomic change in our model. Graphical representations of the STACO model's regional emissions, GDP and population projections over the 2020 to 2100 time horizon can be found in Appendix B in Supplementary material.

5 Results

This section reports the simulation results. To compare the different matching agreement variants in terms of abatement and payoff levels, we calculate so-called 'closing-the-gap indices' (CGIs) for each variant. Here, the gap is the difference between the non-cooperative Nash baseline ('all-singletons') and the globally optimal ('grand coalition') outcome, and the index expresses to what extent each agreement can bridge the divide between the respective global abatement and payoff levels. In Table 7, we report CGIs of the different agreements, as well as the agreements' implied temperature increases above 1990 levels at the end of the 21st century.

Broadly speaking, the globally optimal abatement level requires a 93.0% reduction of the annual mean global BAU emissions over the time horizon 2020–2100, while the all-singletons abatement level only achieves 10.2% (Table 5). In the non-cooperative outcome, abatement contributions strongly depend on players' marginal abatement benefits, while in the fully cooperative outcome the players' marginal abatement costs are equalized, thus inducing the players with the smallest cost parameters to make the largest contributions (Table 5). Although the globally optimal outcome leads to the highest possible collective payoff, six out of the twelve players experience negative payoffs in the grand coalition (Table 6). This makes such an agreement highly improbable in practice unless effective redistribution mechanisms are put in place, as it means that countries with low abatement costs would incur large net losses from their climate mitigation efforts for the benefit of some other countries. Additionally, the grand coalition requires some players

to make abatement contributions that exceed their average BAU emissions making negative emission technologies necessary to achieve the abatement targets.

We now turn to examine the numerical results of the stable standard agreement, as well as the three different matching agreements in more detail. Table 4 presents the players' marginal abatement costs under the different agreements. Table 5 contains the annual abatement values for the different agreement variants and players. Table 6 reports the annual payoffs for the different agreement variants and players.

5.1 The Standard Agreement

With twelve players, there are $2^{12} - 12 = 4,084$ possible unique coalitions, including the all-singletons outcome. In our simulated standard game of coalition formation (Sect. 3.1), the only coalition which satisfies both stability conditions (Eqs. 4 and 5) is the one between EUR and ROW. In our model, these regions have the highest marginal benefits of abatement and both have moderate marginal abatement costs (see Table 1 and Appendices C and D in Supplementary material), making it lucrative for each of them to maximize their joint payoffs. With relatively lower marginal benefits of abatement, all other regions have incentives to freeride when the {EUR, ROW} coalition is formed.

In terms of global abatement, the standard agreement leads to total annual abatement levels which are equivalent to 15.1% of the world's mean annual BAU emissions (as compared to 10.2% in the all-singletons case and 93.0% under the grand coalition) (Table 5). The coalition {EUR, ROW} thus manages to narrow the abatement gap between Nash baseline levels and globally optimal levels by 5.9% and the payoff gap by 7.3% (Table 7).

5.2 Variant 1: Members Match Other Members

Our simulation of the first matching game variant (Sect. 3.2.1) does not produce any coalitions that satisfy both stability conditions (eqs. 4 and 5). When regions have the option to freely join and opt out of the matching agreement, all of them choose to remain singleton players and the global abatement and payoff levels are the Nash baseline levels. We therefore note CGIs of zero for this first matching agreement variant (Table 7).

5.3 Variant 2: Members Match All Other Players

We obtain more promising results for the second matching agreement variant, i.e., when coalition members commit to matching all other players' flat abatements (Sect. 3.2.2). Here, the grand coalition satisfies the internal stability condition (Eq. 4) and is stable, since for a grand coalition, external stability always holds. Partial coalitions are unstable for this matching game variant. In the stable grand coalition, 57.8% of the global mean annual BAU emissions are abated. The abatement gap between the Nash baseline and globally optimal abatement levels is narrowed by 57.5% and the payoff gap by 52.7% (Table 7).

Table 2 Stable variant 2 matching agreement: grand coalition ($m \approx 0.424$) (annual contributions in Mt CO2-e, annual payoffs in billion 2005 USD)

Region	Flat contribution a_i	Matching contribution $m \sum_{j \neq i} a_j$	Total contribution q_i	Payoff π_i
USA	14,157	-2,563	11,594	642,659
JPN	-5,502	5,778	277	290,046
EUR	9,815	-721	9,094	1,254,680
OHI	-5,794	5,903	108	83,359
ROE	-5,793	5,902	109	86,362
RUS	-5,559	5,803	243	87,145
HIA	-5,266	5,679	412	253,279
CHN	20,092	−5.081	15,011	73,346
IND	-1,234	3,968	2,733	712,789
MES	-5,450	5,756	306	129,830
BRA	-5,925	5,958	33	36,012
ROW	4,577	1,502	6,079	1,536,476
World	8,117	37,883	46,000	5,185,983

The matching rate that maximizes the grand coalition's collective payoff is $m \approx 0.424$. Individual members have no control over the size of their matching contributions since these are determined by the matching rate and the other regions' flat contributions. To adjust their total abatement commitments so as to maximize their individual payoffs, players with relatively low marginal benefits and high marginal costs of abatement are therefore forced to set negative flat abatements. However, as the USA, EUR, CHN and ROW have incentives to set large positive flat contributions, all but USA, EUR and CHN's matching contributions are positive, resulting in positive total contributions by all regions. The different regions' flat, matching and total contributions in the variant 2 matching agreement are reported in Table 2. Since for a grand coalition, matching all other players is the same as matching all members, Eq. 14 applies and we have that the sum of all players' flat abatements is equal to the non-cooperative Nash baseline abatement level, meaning that the global abatement level under the stable variant 2 matching agreement inevitably surpasses the Nash baseline abatement level.

Each region's abatement commitment increases more than five-fold compared to its singleton abatement level, leading to substantial payoff increases of 85 to 466% for all regions but CHN. The latter experiences a payoff decrease of 49%, giving it a clear preference for the all-singletons outcome or the standard agreement. In fact, the matching agreement demands that CHN, similar to the USA and EUR, make annual total abatement contributions in excess of its annual BAU emissions, implying that it must employ negative emission technologies to fulfill its commitments. The matching mechanism forces CHN to accept this payoff loss, since its payoff outside the agreement would be even lower. No region has incentives to opt out of the grand coalition as the remaining coalition would respond

Table 3 Stable variant 3 matching agreement $\{USA, CHN\}$ $(m \approx 0.686)$ (annual contributions in Mt CO2-e, annual payoffs in billion 2005 USD)

Region	Flat contribution a_i	Matching contribution $m \sum_{j \neq i} a_j$	Total contribution q_i	Payoff π_i
USA	0	5567	5567	570,795
JPN	116	0	116	122,613
EUR	3806	0	3806	914,092
OHI	45	0	45	35,023
ROE	46	0	46	36,286
RUS	102	0	102	36,794
HIA	172	0	172	107,607
CHN	0	5567	5567	283,763
IND	1144	0	1144	333,436
MES	128	0	128	54,944
BRA	14	0	14	15,090
ROW	2544	0	2544	866,903
World	8117	11,135	19,252	3,377,346

Table 4 Marginal abatement costs per region (2005 USD per ton CO2-e)

Region	All singletons	Grand coalition	Standard agreement	Variant 2 matching agreement	Variant 3 matching agreement
USA	48.9	225.5*	48.9	277.1*	133.0*
JPN	6.4	225.5*	6.4	36.4*	15.2
EUR	62.0	225.5*	142.9*	351.5*	147.1
OHI	1.8	225.5*	1.8	10.3*	4.3
ROE	1.9	225.5*	1.9	10.7*	4.5
RUS	1.9	225.5*	1.9	10.9*	4.6
HIA	5.6	225.5*	5.6	32.0*	13.4
CHN	21.2	225.5*	21.2	120.0*	44.5*
IND	18.6	225.5*	18.6	105.6*	44.2
MES	2.9	225.5*	2.9	16.3*	6.8
BRA	0.8	225.5*	0.8	4.4*	1.9
ROW	53.4	225.5*	142.9*	302.6*	126.7

*Signatory

by setting high matching rates and low or even negative flat abatements, to which the deviating region must respond with high (flat) contributions.

5.4 Variant 3: Members Match Non-members

Our simulation of the third matching game variant (Sect. 3.2.3), produces a single stable two-player coalition: {USA, CHN}. With this coalition, 24.2% of the global mean annual BAU emissions are abated, narrowing the gap between the Nash

Table 5 Total annual abatements per region (Mt CO2-e)

Region	All singletons		Grand coalition		Standard agreement		Variant 2		Variant 3	
	Mt CO2-e	% of mean annual BAU emissions	Mt CO2-e	% of mean annual BAU emissions	Mt CO2-e	% of mean annual BAU emissions	Mt CO2-e	% of mean annual BAU emissions	Mt CO2-e	% of mean annual BAU emissions
USA	2,046	28	9436*	131*	2,046	28	11,594*	160*	5,567*	77*
JPN	49	4	1715*	123*	49	4	277	20*	116	8
EUR	1,605	21	5834*	76*	3697*	48*	9094*	118*	3,806	49
OHI	19	1	2358*	154*	19	1	108*	7*	45	3
ROE	19	1	2290*	85*	19	1	109*	4*	46	2
RUS	43	2	5037*	189*	43	2	243*	9*	102	4
HIA	73	1	2902*	50*	73	1	412*	7*	172	3
CHN	2,649	19	28,203*	207*	2,649	19	15,011*	110*	5,567*	41*
IND	482	5	5836*	59*	482	5	2733*	28*	1,144	12
MES	54	1	4238*	99*	54	1	306*	7*	128	3
BRA	6	0.3	1678*	79*	6	0.3	33*	2*	14	1
ROW	1,073	5	4529*	22*	2870*	14*	6079*	29*	2544	12
World	8,117	10	74,055	93	12,006	15	5185,983	58	19,252	24

*Signatory

Table 6 Annual payoffs per region (billion 2005 USD)

Region	All singletons		Grand coalition		Standard agreement		Variant 2		Variant 3	
	Billion 2005 USD	% of global payoff	Billion 2005 USD	% of global payoff	Billion 2005 USD	% of global payoff	Billion 2005 USD	% of global payoff	Billion 2005 USD	% of global payoff
USA	346,787	21	2,556,541*	31*	536,923	25	642,659*	12*	570,795*	17*
JPN	51,909	3	281,657*	3*	76,857	4	290,046*	6*	122,613	4
EUR	453,645	27	3,935,359*	47*	480,550*	22*	1,254,680*	24*	914,092	27
OHI	14,790	1	-130,700*	-2*	21,885	1	83,359*	2*	35,023	1
ROE	15,323	1	-118,213*	-1*	22,674	1	86,362*	2*	36,286	1
RUS	15,569	1	-425,430*	-5*	23,050	1	87,145*	2*	36,794	1
HIA	45,649	3	91,213*	1*	67,622	3	253,279*	5*	107,607	3
CHN	143,831	9	-1,611,358*	-19*	226,191	10	73,346*	1*	283,763*	8*
IND	146,741	9	721,891*	9*	219,210	10	712,789*	14*	333,436	10
MES	23,271	1	-264,711*	-3*	34,460	2	129,830*	3*	54,944	2
BRA	6,365	0.4	-131,132*	-2*	9,416	0.4	36,012*	1*	15,090	0.4
ROW	404,773	24	3,443,787*	41*	436,064*	20*	1,536,476*	30*	866,903	26
World	1,668,654		8,348,905		2,154,903		5,185,983		3,377,346	

*signatory

Table 7 Closing-the-gap indices (CGI) and warming in 2100

		All singletons	Grand coalition	Standard agreement	Variant 1	Variant 2	Variant 3
Global annual abatement	Mt CO2-e	8117	74,055	12,006	8117	46,000	19,252
	% of mean BAU emissions	10.2	93.0	15.1	10.2	57.8	24.2
	CGI abatement	**0%**	**100%**	**5.9%**	**0%**	**57.5%**	**16.9%**
Global annual payoff	Billion 2005 USD	1,668,654	8,348,905	2,154,903	1,668,654	5,185,983	3,586,117
	CGI payoff	**0%**	**100%**	**7.3%**	**0%**	**52.7%**	**28.7%**
Warming in 2100	°C above 1990 level	3.71	1.89	3.60	3.71	2.66	3.40

baseline and globally optimal abatement levels by 16.9%. Similarly, the global payoff gap is narrowed by 28.7% (Table 7). This constitutes a significant improvement over the stable standard agreement.

The equilibrium matching rate in this agreement is $m \approx 0.686$. The different players' flat, matching and total contributions are reported in Table 3. Players' total abatement commitments more than double compared to the singleton abatement levels, leading to substantial payoff increases of 65 to 137% for all players.

6 Discussion

The aim of this study is to provide numerical evidence of the potential of three matching mechanisms to increase players' contributions and payoffs in IEAs. While prior research has substantiated that different versions of the matching game can produce Pareto-efficient outcomes given full participation by all players (e.g. Guttman 1978; Guttman and Schnytzer 1992; Rübbelke 2006; Boadway et al. 2011), only three theoretical studies have investigated how matching agreements fare with regard to coalition stability thus far (Fujita 2013; Wood and Jotzo 2015; Liu 2018). Given that enforcing participation is generally difficult in IEAs, coalition stability is decisive for the effectiveness of such agreements in practice.

The merit of this study as compared to the studies by Fujita (2013), Wood and Jotzo (2015) and Liu (2018) is three-fold. First, while these studies' approaches are purely theoretical, we offer numerical results to provide further insight into the practical implications of the matching games analyzed. In particular, the simulations of the different matching agreements we conducted convey a sense of magnitude of the matching mechanism's effect on abatements and payoffs relative to the non-cooperative Nash baseline and the desired socially optimal outcome. Second, we allow for player heterogeneity (as does Liu (2018)), while simultaneously considering cooperatively determined matching rates (as in Fujita (2013)), rather than exogenously set ones (as in Liu (2018)). These game characteristics, which had not been analyzed in combination before, are arguably more representative of a real-world setting in which players are heterogeneous and agreements are designed to reach a collective goal. Third, by analyzing three related variations of the matching game, we allow for direct comparison between the effects of the different agreement designs and manage to demonstrate that a matching mechanism where coalition members match all players' contributions (variant 2) may be more effective at counteracting free-rider incentives than ones where members only match other members or non-members' contributions.

Generally, our simulation results indicate that the adoption of a matching mechanism by countries aiming to mitigate climate change at a global level may be a promising approach to address the current underprovision of global mitigation efforts. In particular, matching mechanisms in which coalition members commit to supporting mitigation efforts of outsiders (as in variant 2 and 3) seem to outperform ones in which coalition members only match each others' contributions (as in variant 1). The promise by coalition members to effectively subsidize other countries' emission reductions by reciprocating their efforts with a matching contribution incentivizes

these countries to increase their total abatement levels, regardless of whether they join the coalition or not. Our results are generally consistent with the findings of other studies. As observed by Liu (2018), coalition stability seems to be difficult to achieve when coalition members only match other members' contributions (variant 1). In fact, the variant 1 matching agreement that we simulated performs worse in terms of stability than the standard agreement. In contrast, Fujita (2013) finds that the variant 1 matching mechanism produces stable and efficient outcomes when symmetric players are considered, suggesting that free-riding incentives emerge as a result of player heterogeneity. As predicted by Molina et al. (2020), players' equilibrium abatement contributions are higher than at the Nash baseline in in the variant 2 outcome. It remains to be shown whether more elaborate types of matching games such as the one analyzed by Boadway et al. (2011) and Molina et al. (2020), where players set their own matching rates non-cooperatively, can further improve the performance of matching games with asymmetric players, to the point that they may be able to bring about stable agreements with globally optimal abatement levels. The third matching game variant is most similar to Bucholz et al.'s (2015) and Buchholz and Liu's (2020) unilateral matching games, in which matching players match non-matching players' contributions. The authors show that the unilateral games lead to Pareto-improving outcomes given that the matching players abstain from reducing their flat contributions to below their initial (Nash) levels. We obtain similar results assuming members' flat contributions are constrained to be non-negative.

Although the parameters underlying our simulations were calibrated based on recent empirical data from widely recognized climate modelling sources, the absolute results of the different agreement simulations are likely far from representative of how these games would play out in practice and should therefore be interpreted with caution. Since our objective was to demonstrate how the matching games and the standard game of coalition formation compare in terms of their ability to close the gap between the Nash baseline and the globally optimal abatement and payoff levels, we did not aspire to accurately reflect the complexities underlying real climate policy negotiations.

The most notable simplification we adopt in configuring the simulation model is the reduction of the dynamic regional abatement benefits and costs to static representations thereof. This was done to increase the model's conceptual fit with the one-shot matching game for which we obtain analytical solutions, but it implies that the abatement strategies and payoffs calculated in our simulations do not accurately reflect the outcomes one would obtain when analyzing the games in a more realistic, dynamic setting. Moreover, the functional forms we adopted for players' abatement benefit and cost curves are at best crude reflections of the real relationships they represent. While the linearity of the benefits functions facilitates our analysis as it ensures dominant strategies for all non-signatories, regions' real abatement benefits curves are likely to be more complex and to give rise to issues of carbon leakage. It is also likely that the costs of negative emissions technologies are comparable across different regions, implying that the simulation results need refinement when players commit to abatement contributions that exceed their BAU emissions baselines (as is the case in the grand coalition and variant 2 outcomes).

More importantly yet, the theoretical models of the matching mechanisms analyzed also have their limitations. First, they assume that all players have perfect information about their own and the other players' abatement benefit and cost functions. This requires certainty over the different players' future economic growth and GHG emissions pathways, the relationship between global emissions and future climate damages incurred by the respective players, as well as the cost of available abatement technologies. It goes without saying that such certainty is unattainable in practice. Second, the models presuppose that all players who choose to join a matching coalition can commit to the matching contributions required by the agreement. When players can defect on their conditional abatements, the freeriding problem reemerges and coalition members may have to resort to sanctioning mechanisms to uphold the agreement. Moreover, when punishment is costly, a second-order freeriding problem arises where players prefer others to do the punishing (Molina et al. 2020). The third matching game variant further requires that the matching players commit to non-negative flat abatements, which introduces an additional risk of defection. Third, any process that involves negotiation is vulnerable to the exploitation of power asymmetries. Our matching models assume that the matching rates negotiated by the coalition members maximize the coalitions' collective payoffs. In practice, it likely cannot be guaranteed that this is the case when coalition members have conflicting interests and unequal powers in the negotiation process. Similarly, the coalition members may have varying preferences for the different possible matching mechanisms. In our simulations, CHN is the only coalition member that reaps a larger payoff from the variant 3 agreement than from variant 2. Given that the variant 2 agreement brings about the largest collective benefit, a negotiation of the matching mechanism to be adopted may have to involve compensations for the players with incentives to vote for a collectively inferior but privately beneficial mechanism. Lastly, Liu (2019) expresses the concern that "matching mechanisms may be too sophisticated for practical implementation". Even if there was a way to resolve the issues of incomplete information, incredible commitment and power asymmetries, it could still be too challenging for players to understand and anticipate all other players' equilibrium contributions and set their optimal strategies accordingly. While our simulation deals with a manageable number of players, IEAs like the Paris Agreement involve significantly larger numbers of countries, making the implementation of a matching mechanism considerably more complex.

Nonetheless, matching agreements also provide clear advantages over other types of agreements. Most importantly perhaps, the players joining a matching agreement are only required to commit to conditional contributions, as opposed to making unconditional pledges as done under the Paris Agreement. This allows the players to retain more of their sovereignty and reduces the threat of incredible commitments (Molina et al. 2020). If, for instance, the players that announced large positive flat abatements reneged on their commitments, the other coalition members could punish the defectors by reducing their conditional abatements accordingly, leaving the whole coalition with lower total abatement levels. This type of punishment does not impose additional costs on the punishers, as it suffices that the latter follow their equilibrium abatement strategies. Another advantage of this type of agreement, as Molina et al. (2020) observe, is that it does not rely on any direct transfers between

players. While many design proposals for potential IEAs involve players subsidizing each other's contributions via monetary or in-kind transfers, the implicit subsidization mechanism of matching agreements facilitates matters because it is more likely to be implemented by players between which diplomatic relations are tense. The type of matching mechanism studied by Boadway et al. (2011) and Molina et al. (2020), in which players set their own matching rates towards each of the other players non-cooperatively, provides even greater degrees of player sovereignty and robustness to commitment issues and power abuse in the sense that it circumvents the negotiation stage and allows players to influence the size of their matching contributions. However, this comes at the expense of an increased complexity of the matching game, which makes it more challenging to analyze analytically and to implement in practice.

7 Conclusion

This paper investigates the effects of matching mechanisms on coalition stability, abatement contributions and payoffs in IEAs among heterogeneous players. It does so to explore whether matching agreements can help overcome the freeriding problem that standard environmental agreements commonly struggle with. The three variants of matching agreements we analyze feature uniform, cooperatively determined matching rates by which coalition members match the flat contributions of (i) the other coalition members, (ii) all other regions, or (iii) only non-members, while non-members do not commit to any matching and maximize their individual payoffs. As opposed to most other studies on matching agreements, we not only develop theoretical game models, but also analyze them numerically by means of simulations with an empirically calibrated model.

We find that the first matching agreement variant we analyzed performs worse in terms of stability than the standard agreement. While the latter produces a stable two-player coalition, the variant 1 matching agreement fails to produce any coalitions for which the participating players have incentives to commit to a membership. The second variant produces a stable grand coalition and significantly increases the abatement and payoff levels beyond the non-cooperative Nash baseline. Partial coalitions are unstable in this game. The third variant produces a two-player coalition similar to the standard coalition formation game, but with different members and higher abatement and payoff levels due to the matching mechanism.

In terms of the agreements' ability to bridge the gaps between the global Nash baseline abatement and payoff levels and those of the globally optimal grand coalition outcome, the variant 2 matching agreement also fares better than the variant 3 matching agreement and the standard agreement in our simulations. In particular, the variant 2 matching agreement narrows the global abatement and payoff gaps by 57.5 and 52.7%, respectively. The variant 3 matching agreement only manages to narrow the abatement and payoff gaps by 16.9 and 28.7%. This is still superior to the results of the stable two-player coalition produced by the standard game of coalition formation, which bridges the global abatement and payoff gaps by 5.9 and 7.3%, respectively.

Hence, matching agreements as analyzed in this study may be able to offer a better alternative to standard agreements, provided that the coalition matches not just their fellow members but also the singletons' contributions. The matching commitment incentivizes all matched players to contribute more to global abatement than they otherwise would. We leave it up to future research to determine whether these results hold in dynamic, multi-period games and when considering more complex payoff specifications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10726-024-09878-w.

Acknowledgements We thank Miyuki Nagashima and Rob Dellink for indispensable support. Helpful comments from two anonymous reviewers and EAERE 2022 conference participants are gratefully acknowledged.

Funding Open access funding provided by Copenhagen University.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Barrett S (1990) The problem of global environmental protection. Oxf Rev Econ Policy 6(1):68–79
Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894
Boadway R, Song Z, Tremblay J-F (2007) Commitment and matching contributions to public goods. J Public Econ 91(9):1664–1683
Boadway R, Song Z, Tremblay J-F (2011) The efficiency of voluntary pollution abatement when countries can commit. Eur J Polit Econ 27(2):352–368
Buchholz W, Cornes R, Peters W, Rübbelke D (2015) Pareto improvement through unilateral matching of public good contributions: the role of commitment. Econ Lett 132:9–12
Buchholz W, Cornes R, Rübbelke D (2011) Interior matching equilibria in a public good economy: an aggregative game approach. J Public Econ 95(7–8):639–645
Buchholz W, Cornes R, Rübbelke D (2012) Matching as a cure for underprovision of voluntary public good supply. Econ Lett 117(3):727–729
Buchholz W, Cornes R, Rübbelke D (2014) Potentially harmful international cooperation on global public good provision. Economica 81(322):205–223
Buchholz W, Liu W (2020) Global public goods and unilateral matching mechanisms. J Public Econ Theory 22(2):338–354
Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52(3):309–328
Day T, Röser F, Kurdziel M (2016) Conditionality of intended nationally determined contributions (INDCs). Technical report, New Climate Institute. Accessed: 07 Dec 2023

Dellink R, Chateau J, Lanzi E, Magné B (2017) Long-term economic growth projections in the shared socioeconomic pathways. Glob Environ Chang 42:200–214
Dellink RB, Altamirano-Cabrera JC, Finus M, van Ierland EC, Ruijs A, Weikard HP (2004) Technical background paper of the STACO model (version 1.0)
Dellink RB, de Bruin KC, Nagashima M, van Ierland EC, Urbina-Alonso Y, Weikard HP, Yu S (2015) STACO technical document 3: model description and calibration of STACO 3
Finus M (2001) Game theory and international environmental cooperation, new horizons in environmental economics. Edward Elgar Publishing Ltd, Cheltenham
Finus M, Altamirano-Cabrera J-C, Van Ierland EC (2005) The effect of membership rules and voting schemes on the success of international climate agreements. Public Choice 125(1):95–127
Finus M, Ierland Ev, Dellink R (2006) Stability of climate coalitions in a cartel formation game. Econ Gov 7(3):271–291
Fujita T (2013) A self-enforcing international environmental agreement on matching rates: Can it bring about an efficient and equitable outcome? Strategic Behav Environ 3(4):329–345
Guttman JM (1978) Understanding collective action: matching behavior. Am Econ Rev 68(2):251–255
Guttman JM, Schnytzer A (1992) A solution of the externality problem using strategic matching. Soc Choice Welf 9(1):73–88
Hagen A, von Mouche P, Weikard H-P (2020) The two-stage game approach to coalition formation: Where we stand and ways to go. Games 11(1):3
Hoel M (1992) International environment conventions: the case of uniform reductions of emissions. Environ Resource Econ 2(2):141–159
IIASA (2018) Shared socioeconomic pathways database - version 2.0. https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage &page=10#v2. Accessed: 28 Dec 2020
Kawamata K, Horita M (2014) Applying matching strategies in climate change negotiations. Group Decis Negot 23(3):401–419
Lessmann K, Kornek U, Bosetti V, Dellink R, Emmerling J, Eyckmans J, Nagashima M, Weikard H-P, Yang Z (2015) The stability and effectiveness of climate coalitions: a comparative analysis of multiple integrated assessment models. Environ Resource Econ 62:811–836
Liu W (2018) Global public goods and coalition formation under matching mechanisms. J Public Econ Theory 20(3):325–355
Liu W (2019) Participation constraints of matching mechanisms. J Public Econ Theory 21(3):488–511
Molina C, Akçay E, Dieckmann U, Levin SA, Rovenskaya EA (2020) Combating climate change with matching-commitment agreements. Sci Rep 10(1):1–12
Morris J, Paltsev S, Reilly J (2012) Marginal abatement costs and marginal welfare costs for greenhouse gas emissions reductions: results from the eppa model. Environ Model Assess 17(4):325–336
Mäler K-G (1989) The acid rain game. Stud Environ Sci 36:231–252
NOAA (2020) Noaa's annual greenhouse gas index. https://www.esrl.noaa.gov/gmd/aggi/
Rennert K, Prest BC, Pizer WA, Newell RG, Anthoff D, Kingdon C, Rennels L, Cooke R, Raftery AE, Ševčíková H et al (2022) The social cost of carbon: advances in long-term probabilistic projections of population, GDP, emissions, and discount rates. Brook Pap Econ Act 2021(2):223–305
Riahi K, Van Vuuren DP, Kriegler E, Edmonds J, O'neill BC, Fujimori S, Bauer N, Calvin K, Calvin K, Dellink R, Fricko O et al (2017) The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. Glob Environ Chang 42:153–168
Rübbelke D (2006) Analysis of an international environmental matching agreement. Environ Econ Policy Stud 8(1):1–31
Samir K, Lutz W (2017) The human core of the shared socioeconomic pathways: population scenarios by age, sex and level of education for all countries to 2100. Glob Environ Chang 42:181–192
Tol RS (2009) The economic effects of climate change. J Econ Perspect 23(2):29–51
UNFCCC (2012) Appendix i-quantified economy-wide emissions targets for 2020. http://unfccc.int/meetings/copenhagen_dec_2009/items/5264.php. Accessed 07 Dec 2023
UNFCCC (2015) Synthesis report on the aggregate effect of the intended nationally determined contributions. https://unfccc.int/resource/docs/2015/cop21/eng/07.pdf. Accessed: 2023-12-07
UNFCCC (2023) Nationally determined contributions under the Paris agreement. https://unfccc.int/sites/default/files/resource/cma2023_12.pdf. Accessed: 2023-12-07
Weikard H-P (2009) Cartel stability under an optimal sharing rule. Manchester School 77(5):575–593
WMO (2020) State of the global climate 2020. https://storymaps.arcgis.com/stories/6942683c7ed54e51b433bbc0c50fbdea

Wood PJ, Jotzo F (2015) The stability of mechanisms for matching abatement commitments when not all countries commit. Unpublished manuscript