



OPEN

# Combating climate change with matching-commitment agreements

Chai Molina<sup>1,2,3</sup>✉, Erol Akçay<sup>3,8</sup>, Ulf Dieckmann<sup>1,4,8</sup>, Simon A. Levin<sup>1,2,5,6,8</sup> & Elena A. Rovenskaya<sup>1,7,8</sup>

Countries generally agree that global greenhouse gas emissions are too high, but prefer other countries reduce emissions rather than reducing their own. The Paris Agreement is intended to solve this collective action problem, but is likely insufficient. One proposed solution is a matching-commitment agreement, through which countries can change each other's incentives by committing to conditional emissions reductions, before countries decide on their unconditional reductions. Here, we study matching-commitment agreements between two heterogeneous countries. We find that such agreements (1) incentivize both countries to make matching commitments that in turn incentivize efficient emissions reductions, (2) reduce emissions from those expected without an agreement, and (3) increase both countries' welfare. Matching-commitment agreements are attractive because they do not require a central enforcing authority and only require countries to fulfil their promises; countries are left to choose their conditional and unconditional emissions reductions according to their own interests.

Anthropogenic emissions of greenhouse gases (GHGs) have been identified as comprising a key factor in climate change<sup>1</sup>. In the Paris Agreement, countries overwhelmingly agree to limiting “the increase in the global average temperature to well below 2 °C [...] and pursuing efforts to limit the temperature increase to 1.5 °C above pre-industrial levels”<sup>2</sup>. To do so, they must curb GHG emissions; yet according to recent assessments of present and future global GHG emissions, the Earth is still expected to warm by more than 2 °C<sup>3–6</sup>.

Achieving the necessary reduction in GHG emissions is difficult, primarily because it is an instance of Hardin's classic “tragedy of the commons”<sup>7</sup>: each country benefits from reduced *global* emissions, but curbing *its own* emissions carries an economic cost. Countries thus prefer to free-ride on the emissions abatement (i.e., reduction of emissions) of others over reducing their own emissions<sup>8,9</sup>. The implied tragedy is that if countries pursue their own interests in this way, everyone will be worse off than if they had cooperated and acted in their collective interest.

To manage shared resources (or to resolve other collective action problems) within a single country, rules are typically agreed upon and enforced, making cooperation the rational choice by shaping individual incentives. In contrast, the Paris Agreement details countries' voluntarily pledged emission reductions in the hope of averting a “climate tragedy”, but this agreement will likely fall short of its goal. For example, even if countries reduced their emissions by the amounts they pledged, the Earth would still be an estimated 3.2 °C warmer by 2100 (despite the agreement's goal of limiting global warming to no more than 1.5 °C)<sup>10</sup>. The cause for this discrepancy between the Paris Agreement's goal and individual countries' abatement targets is that, because countries are sovereign, their participation must be voluntary. Thus, countries chose their own abatement targets, and had little incentive to make very ambitious pledges. Moreover, the abatement targets that countries have pledged are not legally binding, so countries are not incentivized to fulfill them; consequently some countries' current policies are insufficient to fulfill even their modest pledges<sup>10</sup>. So, while, in principle, “prompt and substantial reductions in greenhouse gas

<sup>1</sup>International Institute for Applied Systems Analysis, Laxenburg, A-2361, Austria. <sup>2</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, 08544 NJ, USA. <sup>3</sup>Department of Biology, University of Pennsylvania, Philadelphia, PA, 19104, USA. <sup>4</sup>Department of Evolutionary Studies of Biosystems, The Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa, 240-0193, Japan. <sup>5</sup>Beijer Institute of Ecological Economics, SE-104 05, Stockholm, Sweden. <sup>6</sup>Resources for the Future, Washington, DC, 20036, USA. <sup>7</sup>Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Leninskiye Gory 1-52, Moscow, 119234, Russia. <sup>8</sup>These authors contributed equally: Erol Akçay, Ulf Dieckmann, Simon A. Levin and Elena A. Rovenskaya. ✉e-mail: [chai.molina@gmail.com](mailto:chai.molina@gmail.com)

emissions on a global scale” could limit warming to meet the 1.5 °C target set in Paris, at 2015 rates of emissions the global carbon budget for this goal will be exhausted already in 2021<sup>11</sup>.

Seeking possible mechanisms for lowering global emissions, many researchers have turned to game theory (see ref. <sup>12</sup> for a review). Most of the resulting proposals involve establishing a central authority that can punish non-participation and noncompliance or otherwise redistribute money or resources (either directly or by approving sanctions imposed by countries)<sup>8,13–16</sup>. But enforcement is difficult to achieve in practice, because, as pointed out by Nordhaus<sup>8</sup>, it conflicts with the principles of countries’ sovereignty, equality and right to manage internal affairs without intervention, first established in the 1648 Treaty of Westphalia. Consequently, Nordhaus has suggested that international trade law be amended to allow sanctions against non-participants, while prohibiting retaliation by the latter<sup>8</sup>. This would require the comprehensive renegotiation of trade laws, which will be a difficult and lengthy process. Moreover, Barrett also stresses the importance of enforcement, but warns that even if countries agree to a treaty with strong enforcement mechanisms, they will likely do so at the expense of lowering their abatement targets “to ensure that punishments are not imposed”<sup>14</sup>. Hence, solutions that encourage countries to cooperate without encroaching on their sovereignty would be desirable, since the global carbon budget for 1.5 °C warming is quickly running out.

Barrett<sup>17,18</sup> was perhaps the first to propose applying Guttman’s<sup>19,20</sup> idea of matching contributions to lower GHG emissions and help mitigate climate change. More recently, Boadway *et al.*<sup>21</sup> applied Guttman’s<sup>19,20</sup> matching scheme to Finus’s<sup>22</sup> Global Emission Game: every country’s emissions abatement is divided into an unconditional and a conditional abatement. Countries simultaneously declare **matching factors**, which define how much conditional abatement they will contribute for a unit of unconditional abatement contributed by each other country. (The literature typically refers to a country’s unconditional abatement as its flat—or direct—contribution, to its conditional abatement as its indirect contribution, and to the matching factors as matching rates. We have chosen our terminology in the hope that it is more transparent.) Assuming countries can commit to the matching factors they declare (see the section “Commitment” for a justification of this assumption), this program allows countries to alter the incentives of other countries without invoking a central authority.

Boadway *et al.*’s<sup>21</sup> inspirational analysis considers an arbitrary number of heterogeneous countries and arrives at an optimistic result. They show that a matching-commitment agreement can lead to new equilibrium emission levels that are locally Pareto efficient and unique and reduce emissions relative to the status quo. However, Boadway *et al.* consider the status quo emissions in the absence of an agreement to be exogenously determined, and their results require that at least one country has an incentive to unilaterally reduce emissions from this status quo even in the absence of an agreement. A more realistic scenario would be that countries follow their Nash equilibrium emissions strategy in the absence of an agreement, so that no country has an incentive to either increase or decrease their emissions<sup>6,23</sup>. In that case, Boadway *et al.*’s optimistic results disappear. In other words, in Boadway *et al.*’s analysis, without an agreement, at least one country’s emissions are irrationally high, even for its pure self-interest; matching-commitment agreements reduce emissions because they allow countries to respond to their incentives and thus lower their emissions. Thus, we do not know whether matching-commitment agreements can work to reduce emissions if countries are able to freely select emissions levels that align with their incentives in the absence of an agreement. Furthermore, Boadway *et al.* do not consider the question of global efficiency, i.e., whether the agreement improves global welfare, the payoff of all countries, or indeed any single country’s payoff, relative to the outcome in the absence of an agreement (i.e., either relative to a Nash equilibrium emissions profile, or to the non-Nash baseline Boadway *et al.* assume is selected in this case).

Against this background, this paper addresses two main questions: First, can a matching-commitment agreement incentivize countries to reduce their emissions relative to a baseline scenario without an agreement, and in which countries’ choices align with their incentives? Second, are countries better off with such an agreement than without? We build on Boadway *et al.*’s<sup>21</sup> formulation, but require the baseline scenario—that is, countries’ emissions in the absence of an agreement for reducing them—to be a Nash equilibrium determined endogenously by the countries’ strategic decisions. For arbitrarily many countries that may differ in their economic and climate-related properties, and that choose their emissions independently, we find conditions guaranteeing that such a Nash equilibrium emissions profile exists and is unique—i.e., the baseline scenario is well-defined. Next, we analyze the effects of a matching-commitment agreement on the emissions of two such countries. We show that a matching-commitment agreement generates a new, unique equilibrium emissions profile. (More precisely, there are multiple subgame-perfect Nash equilibria—see “Model and methods” and ref. <sup>24</sup>—which all yield identical emissions profiles.) At this equilibrium emissions profile, both countries’ emissions are lower and their payoffs are higher than at baseline, i.e., at the equilibrium without a matching-commitment agreement. Lastly, we show that the equilibrium emissions profile with matching-commitment agreements is locally Pareto efficient.

While Boadway *et al.*’s<sup>21</sup> results broke new ground and motivated all that we have done, there are subtle problems with their arguments for both the existence and the uniqueness of the equilibrium, and hence the claims for the general case do not hold up (see Remark B.15 and Appendix C.11.6 for details). Indeed, even the two-country case is challenging for providing a rigorous proof of existence and uniqueness, which requires a different mathematical approach than that of Boadway *et al.* We introduce that method and proof in this paper for matching-commitment agreements between two countries, and show rigorously that the equilibrium (Eqs. (5) and (6)) was, in fact, correctly identified by Boadway *et al.* The *n*-country case remains an open problem.

## Model and methods

**Basic climate game (BCG).** We start by describing a model for the public goods problem of global GHG emissions between two countries (or unions thereof, e.g., the European Union), building on Boadway *et al.*’s<sup>21</sup> model, which in turn is similar to one described by Finus (ref. <sup>22</sup>, Ch. 9) and an example of Folmer and von Mouche’s<sup>25</sup> transboundary pollution games. As is often the case for game-theoretical models of international

cooperation, we focus on two countries, to retain mathematical tractability (see, e.g., refs. 26–32; we compare our analysis with the literature on  $n$ -player matching mechanisms in the section “Two vs. many countries”).

We consider two countries deciding on their emissions levels,  $e_i \in \mathbb{R}$  ( $i = 1, 2$ ). We denote the **emissions profile**, i.e., the vector of countries’ emissions, by  $e = (e_1, e_2) \in \mathbb{R}^2$ , and the **total emissions** by  $e = e_1 + e_2$ . Note that we allow countries to have negative net GHG emissions (e.g., by capturing carbon from the atmosphere), since this is assumed in many theoretical mitigation scenarios that meet the goals of the Paris Agreement<sup>33</sup>. We denote by  $B_i(e_i)$  and  $D_i(e)$  the benefit and damage, respectively, to country  $i$  ( $i = 1, 2$ ) when its own emissions level is  $e_i$  and the total emissions are  $e = e_1 + e_2$ . The benefit and damage functions are assumed to be increasing and twice differentiable, as well as decelerating and accelerating, respectively (see “Increasing damages from climate change”). Country  $i$ ’s total payoff as a function of the emissions profile  $e$  is

$$\prod_i(e) = B_i(e_i) - D_i(e). \quad (1)$$

Countries seek to maximize their own payoffs, and are indifferent to the other country’s payoff.

In our baseline scenario, which we call the **basic climate game (BCG)**, representing the absence of any international agreement to curb emissions, countries independently and simultaneously choose their own emissions levels. When countries play the BCG and there is a unique Nash equilibrium, we view it as the baseline scenario, as it represents the predicted emissions profile for rational countries in the absence of any agreement to limit emissions. Unfortunately, and not surprisingly, the Nash equilibrium of the BCG does not maximize **global welfare**, i.e., the total payoff of both countries,  $\Pi = \Pi_1 + \Pi_2$ , and is hence not socially optimal. Indeed, we show in Appendix A.3 that the total emissions at a **social optimum (SO)**, i.e., at an emissions profile that maximizes global welfare, must be lower than the total baseline emissions (Appendix A.3 discusses when an SO exists, which is in general not guaranteed by our assumptions). Moreover, since the baseline emissions profile is not Pareto-efficient, there are emissions profiles at which both countries’ payoffs are higher than at baseline (Appendix A.4). Thus, an international environmental agreement that results in lower emissions and higher national payoffs than baseline is desirable (but is still likely to be second-best from a global perspective).

**Matching climate game (MCG).** One suggested approach to establishing an accord that incentivizes a reduction in GHG emissions is to use a matching-commitment scheme<sup>17,21</sup>. In this approach, country  $i$  may reduce its emissions relative to its baseline emissions level  $\bar{e}_i$  by an abatement level  $A_i \geq 0$  ( $i = 1, 2$ ), so that its actual emissions level is

$$e_i = \bar{e}_i - A_i. \quad (2)$$

Country  $i$ ’s abatement  $A_i$  is the sum of its unconditional abatement  $a_i$ , and the conditional abatement that it performs in response to the other country’s unconditional abatement,

$$A_i = a_i + m_i a_j, \quad (3)$$

where  $m_i$  is the proportionality factor at which country  $i$  matches country  $j$ ’s unconditional abatement and  $i, j = 1, 2$ ,  $i \neq j$  (that is,  $j$  is the *nonfocal* country). The countries then play the following game, henceforth referred to as the **matching climate game (MCG)** (see also Definition A.9):

- Stage I: Countries independently and simultaneously choose their (non-negative) matching factors,  $m_i \geq 0$ , to which they are subsequently committed.
- Stage II: Countries independently and simultaneously choose their unconditional abatement levels,  $a_i$ , with full knowledge of the matching factors chosen in stage I.

Denoting the total baseline emissions and the total abatement by  $\bar{e} = \bar{e}_1 + \bar{e}_2$  and  $A = \bar{e} - e$ , respectively, country  $i$ ’s payoff becomes the nonlinear function

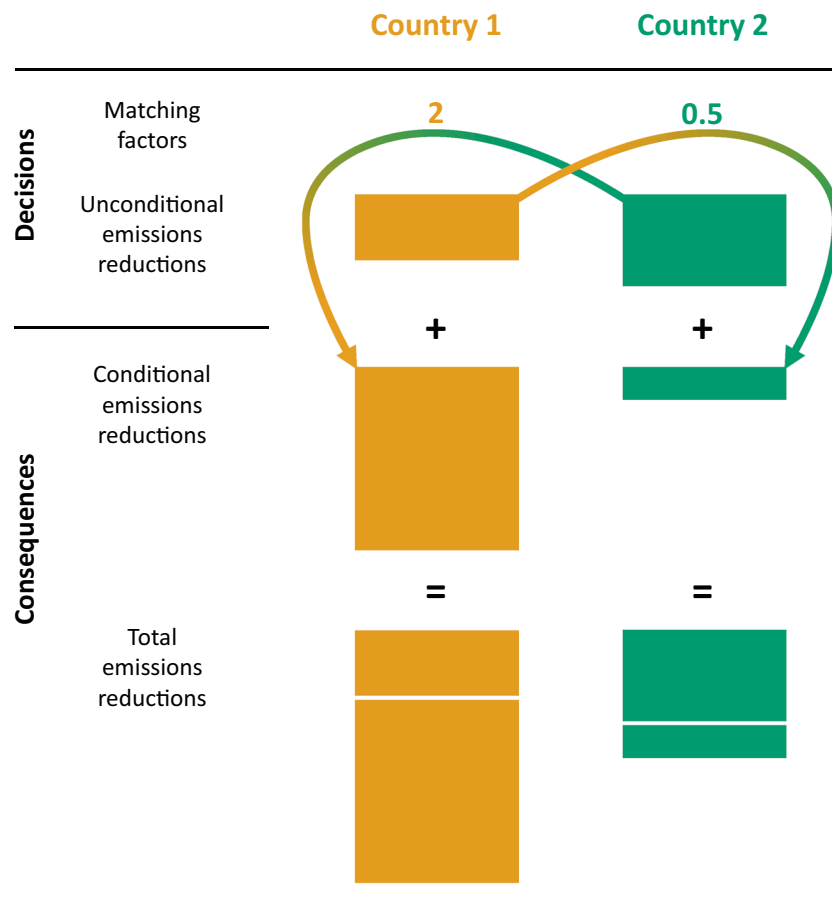
$$\prod_i(e) = B_i(\bar{e}_i - A_i) - D_i(\bar{e} - A) \quad \text{for } i = 1, 2. \quad (4)$$

Note that the baseline emissions levels  $\bar{e}_1$  and  $\bar{e}_2$  must be specified to determine the country payoffs in the MCG. Below, we will take these to be the Nash equilibrium emissions of the BCG (which, under the conditions described in the section “Assumptions on the costs and benefits of emissions”, exists and is unique). These baseline emissions are thus endogenously determined in the sense that they are the rational levels of emissions in the absence of any climate agreement.

The process by which national abatements arise from the unconditional abatements and matching factors is visualized in Figs. 1 and 2.

**Assumptions on the costs and benefits of emissions.** In this paper, we aim to compare the equilibrium outcomes of the basic climate game (BCG) and the matching climate game (MCG). Since such a comparison is predicated on the existence of well-defined equilibrium outcomes for these two games, additional assumptions are needed to guarantee this. For example, in the BCG, it may be advantageous for countries to emit or abate without bound, in which case no Nash equilibrium exists. In the MCG, matching may incentivize a country to abate without bound, in which case the best-response functions for stage II (given matching factors chosen in stage I) need not even be well-defined.

In Appendix A.1, we identify a sufficient condition on the benefit and damage functions that guarantees the existence of a unique Nash equilibrium for the BCG; we call this condition **bounded global emissions (BGE)**

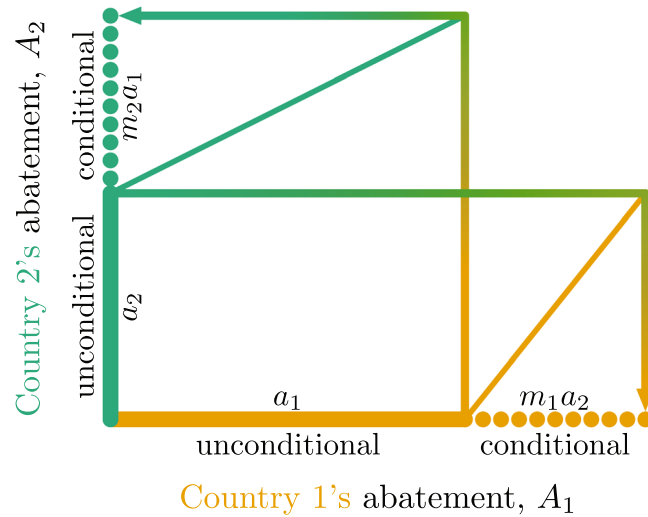


**Figure 1.** Graphical representation of the matching climate game (MCG). In stage I, countries decide on their matching factors. In stage II, countries choose their unconditional abatements, taking into consideration the matching factors chosen in stage I. These decisions then determine the conditional abatements countries perform over and above their unconditional abatements: each country's conditional emissions reduction is the product of its own matching factor, and the *other* country's unconditional abatement.

(Definition A.2), because, roughly, it ensures that all rational emissions profiles lie inside a bounded set. In Appendix A.6, we show that under the assumptions of the BCG, the MCG's stage-II best-response functions are well-defined if and only if (iff) for any country  $i$ , in the limit of negative infinite emissions, either the marginal benefits become infinite or the marginal damages vanish; we refer to this condition as **bounded abatement with matching (BAM)**, because it guarantees that no matching factors can incentivize infinite abatement. Henceforth we assume the benefit and damage functions are such that BAM holds and that global emissions are bounded. It turns out that (at least for the case of two countries) the BAM condition is sufficient to ensure a unique equilibrium outcome for the MCG, in the sense described in our "Results" section. (Note that Boadway *et al.*<sup>21</sup> do not assume the BAM condition, and their analysis does not address the possibility of the MCG's stage-II best-response functions being undefined. Consequently, the MCG's stage-II Nash equilibria in their analysis may not exist.)

While the model described above and the results that follow involve two countries only, note that for any number of countries,  $n \geq 2$ , the BGE and BAM assumptions (resp.) ensure the existence and uniqueness of a Nash equilibrium for the BCG, and that the MCG's stage-II best-response functions are well-defined. Regardless of the number of countries, these properties are necessary to sensibly ask whether and when the MCG yields equilibria that improve on the baseline emissions profile.

**Solution concepts.** The relevant solution concept for the basic climate game (BCG) is the Nash equilibrium, i.e., an emissions profile from which no country has an incentive to deviate unilaterally (see ref.<sup>24</sup>, Ch. 4). The matching climate game (MCG), on the other hand, is a two-stage game in which a strategy for country  $i$  ( $i = 1, 2$ ) specifies its choice of a matching factor,  $m_i = \mu_i$  in stage I, and of an unconditional abatement level for any combination of matching factors chosen by both countries in stage II,  $a_i(m_1, m_2)$  ( $i = 1, 2$ ). A natural choice of solution concept for the MCG is therefore the **subgame-perfect equilibrium (SPE)** (see Ch. 7 of ref.<sup>24</sup>), which yields Nash equilibria at every subgame (the MCG has infinitely many subgames: in addition to the MCG itself, for any choice of matching factors in stage I, the game played in stage II is a subgame of the MCG). For the MCG, an SPE must then satisfy two conditions: first, it must constitute a Nash equilibrium for the full MCG, and second, for any pair



**Figure 2.** Graphical representation of how two countries’ decisions on their matching factors  $m_i$  and unconditional abatements  $a_i$  determine their total abatements  $A_i$  in the matching climate game (MCG). The two countries’ decisions are shown as thick continuous lines, while the consequences of these decisions are shown as thick dotted lines, with thin continuous lines serving as visual aids. Each country’s total abatement, indicated for countries 1 and 2 by the lengths of the thick horizontal and vertical lines, respectively, is partitioned into two components: an unconditional abatement and a conditional abatement. In stage I, each country chooses its matching factor, indicated by the slopes of the two diagonal lines: the ratio of country  $i$ ’s conditional abatement and country  $j$ ’s ( $j \neq i$ ) unconditional abatement is country  $i$ ’s matching factor  $m_i$ , so the slopes of the top-left and bottom-right diagonal lines are  $m_2$  and  $1/m_1$ , respectively. In stage II, each country chooses its unconditional abatement ( $a_i$ ), indicated by the lengths of the horizontal and vertical segments of the thick continuous lines. These decisions then determine each country’s conditional abatement, indicated by the lengths of the dotted lines. The lines with graduated colours represent how one country’s unconditional abatement is matched by the other country.

of matching factors  $(m_1, m_2)$ , the abatements  $(a_1(m_1, m_2), a_2(m_1, m_2))$  must constitute a Nash equilibrium of stage II. For any choice of matching factors played in stage I, an SPE selects abatements that are Nash equilibria for stage II, so this solution concept avoids Nash equilibria (for the full MCG game) that are sustained by incredible threats or other irrational behaviour in stage II. We focus on SPEs precisely because they avoid such irrational behaviour, and our main aim is to identify and compare SPEs of the MCG with the baseline emissions profile.

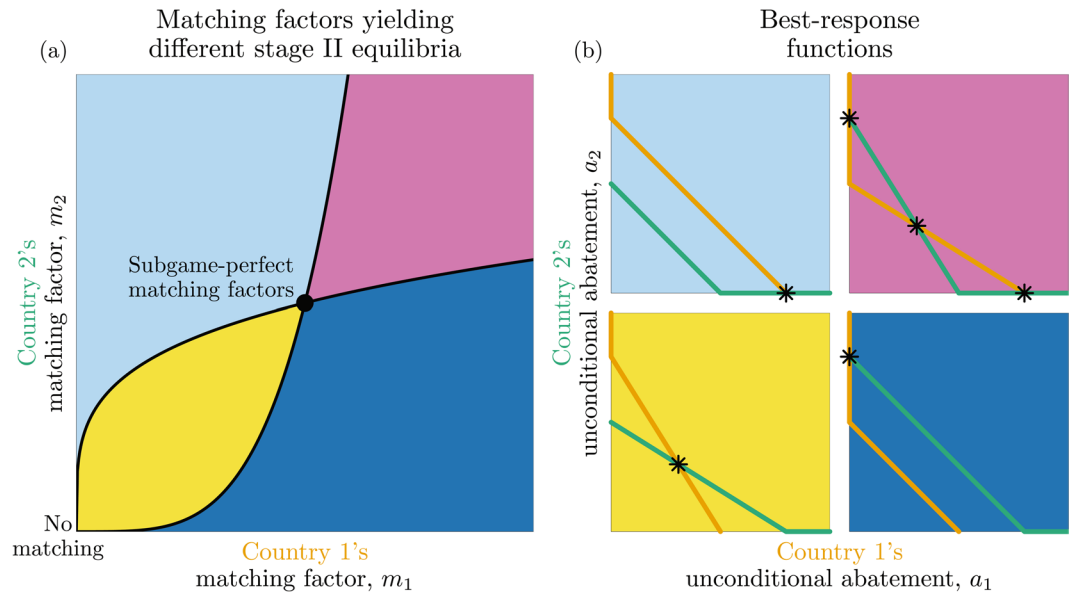
### Results

**The matching climate game has a unique equilibrium emissions profile.** We first use backward induction to find subgame-perfect equilibria (SPEs) for the matching climate game (MCG; see the “Model and methods” section and Definition A.9) between two countries: We assume that given any choice of matching factors in stage I, countries may expect one another to choose unconditional abatement levels that are Nash equilibria given their choices of matching factors,  $(m_1, m_2)$ ; we then find which strategies will be played in stage I. A more detailed account of the process follows below, and the full details are given in Appendix B.

In Appendix B.1, we find the Nash equilibrium levels of unconditional abatement in stage II, given a choice of matching factors in stage I, which can be geometrically characterized as intersections of the stage-II best-response functions,  $\mathcal{R}_1$  and  $\mathcal{R}_2$ . The results of our analysis are summarized in Fig. 3. In particular, we find that there are four regions in the set of non-negative matching factors, in which the stage-II Nash equilibria are qualitatively different. These regions are delimited by the two **stage-II delimiter curves**,  $\phi_1$  and  $\phi_2$  (shown in Fig. 3a as the concave and convex curves, respectively), defined as follows: for any  $m_j \geq 0$ ,  $m_i = \phi_i(m_j)$  is the matching factor at which the intercepts of the two countries’ stage-II best-response functions with the  $a_i$ -axis are identical (Appendix C.7). Between  $\phi_i$  and the  $a_j$ -axis, country  $i$ ’s stage-II best-response curve intercepts the  $a_i$ -axis higher than does  $j$ ’s.

The functions  $\phi_1$  and  $\phi_2$  intersect at two pairs of matching factors: (1)  $m_1 = m_2 = 0$ , for which the unique stage-II Nash equilibrium is no abatement (baseline), and (2) a pair of matching factors  $(m_1^c, m_2^c)$  that are reciprocal (satisfy  $m_1^c \cdot m_2^c = 1$ ; since our model is a one-shot game, there is no future time at which countries have occasion to reciprocate in the sense of responding to one another’s actions), defined by Eqs. (5) and (6) below, for which the best-response functions overlap for positive unconditional abatements, yielding a continuum of stage-II Nash equilibria.

Furthermore, we find that for pairs of matching factors other than  $(m_1^c, m_2^c)$ , the stage-II best-response functions cross at most once in the interior of the quadrant of positive unconditional abatements, i.e.,  $\{(a_1, a_2) | a_1 > 0, a_2 > 0\}$ . This implies that for  $(m_1, m_2) \neq (m_1^c, m_2^c)$  either one best-response function is above the other (upper-left and lower-right panels of Fig. 3b), or they cross at some point in the interior of that



**Figure 3.** (a) Schematic overview of outcomes in stage II of the matching climate game (MCG) for pairs of matching factors chosen in stage I; different colours correspond to qualitatively different stage-II Nash equilibria. The coloured areas are separated by the stage-II delimiter curves ( $\phi_1$  and  $\phi_2$ , shown as the concave and convex black curves, respectively), and the filled circle marks the pair of matching factors yielding pairs of matching factors in the correspondingly coloured area in panel (a), with asterisks indicating stage-II Nash equilibria. (b) Each panel shows a qualitatively different configuration of the best-response functions for pairs of matching factors in the correspondingly coloured area in panel (a), with asterisks indicating stage-II Nash equilibria. When the subgame-perfect matching factors are played in stage I, the stage-II best-response functions overlap for positive unconditional abatements. Note that, in general, the best-response functions need not be piecewise linear.

quadrant (upper-right and lower-left panels of Fig. 3b). Consequently, there are either one, two, or three stage-II Nash equilibria (Fig. 3b), and these cases can be distinguished by comparing the stage-II best-response functions' intercepts with the  $a_1$ - and  $a_2$ -axes:

- On or between  $\phi_1$  (resp.  $\phi_2$ ) and the  $a_2$ -axis ( $a_1$ -axis), there is a stage-II Nash equilibrium at which only country 1 (2) abates unconditionally.
- Between the curves  $\phi_1$  and  $\phi_2$ , there is a stage-II Nash equilibrium at which both countries abate unconditionally.

Next, for any pair of non-negative matching factors,  $(m_1, m_2) \in \mathbb{R}_{\geq 0}^2$ , the payoffs  $\Pi_i(m_1, m_2)$  ( $i = 1, 2$ ) can be determined assuming that a Nash equilibrium is played in stage II; we ascribe these payoffs to pairs of matching factors and obtain payoff functions  $\Pi_i(m_1, m_2)$  that we then use to find matching factors that are Nash equilibria of stage I (assuming Nash equilibrium unconditional abatements are played in stage II). These Nash equilibrium matching factors, along with the choice of Nash equilibria played in stage II that was used to calculate the stage I payoff functions, constitute SPEs.

However, the multiplicity of stage-II Nash equilibria for some pairs of matching factors introduces a complication, because the payoff to country  $i$  given  $(m_1, m_2)$  may then not be uniquely defined even under the assumption that only Nash equilibria are played in stage II (see top-right corner of Fig. 3b); one must then analyze an infinite family of continuous-strategy games, each with payoff functions  $\Pi_i(m_1, m_2)$  ( $i = 1, 2$ ) determined by a particular choice of stage-II equilibria (for all pairs of matching factors for which multiple stage-II equilibria exist).

Despite this complication, it turns out that there is a unique SPE emissions profile. In particular, analyzing all possible pairs of payoff functions  $\Pi_i(m_1, m_2)$  ( $i = 1, 2$ ) obtained assuming equilibrium stage-II play (see Theorem B.19 in Appendix B.2), we show that although the MCG has infinitely many SPEs, the unique pair of matching factors  $(m_1^c, m_2^c)$  is played at all SPEs. Moreover, the equilibrium abatements  $A_i$  (and, therefore the payoffs  $\Pi_i$ ) for the MCG are uniquely defined in the sense that they do not depend on which SPE is chosen. Specifically, with  $m$  and  $a$  denoting the unique positive solutions of

$$B'_1(\bar{e}_1 - a) = (1 + m)D'_1(\bar{e} - (1 + m)a), \tag{5}$$

$$mB'_2(\bar{e}_2 - ma) = (1 + m)D'_2(\bar{e} - (1 + m)a), \tag{6}$$

the equilibrium matching factors are  $(m_i^c, m_j^c) = (1/m, m)$ , and the emissions profile at any SPE is

$$(e_1, e_2) = (\bar{e}_1 - a, \bar{e}_2 - ma). \quad (7)$$

### Both countries are better off playing the matching climate game than the basic climate game.

Comparing the subgame-perfect emissions profile of the matching climate game (MCG) given by Eq. (7) with the baseline emissions profile ( $\bar{e}$ , the Nash equilibrium of the basic climate game, abbreviated BCG), we find:

- At the equilibrium emissions profile of the MCG, both countries' emissions are lower, but their payoffs are higher, than at the baseline equilibrium (i.e., the subgame-perfect emissions profile is a Pareto improvement on the Nash equilibrium of the BCG).
- The equilibrium emissions profile of the MCG is locally Pareto efficient, i.e., a small deviation from the MCG's equilibrium abatement levels cannot increase both countries' payoffs.
- The MCG's equilibrium emissions profile is socially optimal iff the two countries' marginal benefits at the equilibrium emissions profile are equal, which need not generally be the case; it then follows that from the perspective of a social planner, the country having lower marginal benefits at the subgame-perfect emissions profile emits too much, while the one with higher marginal benefits emits too little (Lemma A.8).

### Discussion

Matching-commitment agreements have been garnering increasing interest as a way to construct international agreements to lower global greenhouse gas (GHG) emissions without requiring extensive international monitoring and enforcement. Our main contribution in this paper is to show the theoretical viability of such an approach assuming that countries behave strategically in the absence of an agreement. We show that matching-commitment agreements can indeed result in increased abatement and welfare for the countries involved, relative to a baseline of strategic behavior in the absence of agreements. Our results suggest that matching-commitment agreements can be a promising option for addressing other public goods problems when enforcement is difficult.

To conclude, in the next few subsections we position our results in a broader context. First, we highlight some appealing features of the matching climate game (MCG). Next, we touch on related models from the literature on matching-commitment agreements. We then discuss the main simplifying assumptions made in our model, and mention how our assumptions regarding the damage and benefit functions could be relaxed. Lastly, we discuss how our methods can be applied to the provision of public goods in general.

**Advantages over other climate-agreement schemes.** In addition to improving on the baseline scenario, the matching climate game (MCG) has a number of attractive features in comparison with other mechanisms (including other matching mechanisms) suggested to address international public goods problems such as reducing global GHG emissions, which we describe below.

The MCG asks countries to relinquish very little of their sovereignty and entails no negotiation. First, a country that agrees to participate in the MCG may still choose not to match, and hence does not commit to any policy change, other than refraining from increasing its emissions in response to other countries' emissions reductions, a phenomenon known as carbon leakage. (Similar to Barrett's<sup>9</sup> approach, including a clause stipulating that the matching agreement does not come into force unless all countries are signatories can help assuage countries' fears that their environmental policies would be undone by defecting countries, and hence stabilize participation in the treaty and avoid leakage.) Second, unlike many other climate-agreement schemes (e.g., refs. <sup>8,13</sup>), signatories are not asked to expose themselves to the possibility of economic sanctions. Third, countries decide on their own policies for matching factors and unconditional abatements, rather than submitting to the authority of an institution or coalition of countries (see, e.g., ref. <sup>34</sup> for such a model without matching; see refs. <sup>35,36</sup>, respectively, for matching mechanisms in which a coalition of signatories decides on a common matching factor at which all members will match the total unconditional abatement of all countries, or of other coalition members). Fourth, countries are only asked to commit to conditional actions (as opposed to the Paris Agreement's pledges), and to refrain from emitting more than they would in the absence of an agreement. In other words, the only encroachments on national sovereignty that the MCG makes is requiring countries to keep their word and fulfill a conditional commitment that they choose to make.

The fact that the MCG does not employ costly punishments is also appealing, because it sidesteps the second-order free-riding problem: when punishment is costly, players are tempted to let others do the punishing<sup>37,38</sup>. While in the MCG countries can implicitly punish other countries for insufficiently high matching commitments by reducing their unconditional abatements, at the subgame-perfect emissions profile doing so is not costly in the sense that the countries choose Nash equilibrium unconditional abatements (given the matching factors chosen in stage I).

In an analysis of a similar matching-commitment agreement between  $n$  identical countries, Rübbelke<sup>39</sup> points out that, because countries commit only to matching factors, they can still adjust their unconditional abatements if they discover (after having made their matching commitments) that they underestimated the benefits of abatement. This flexibility allows countries to at least partially (and sometimes fully) correct their emissions abatements based on updated estimates of benefits, avoiding the need for costly renegotiations. Although we do not repeat Rübbelke's analysis for our model, the flexibility in the choice of unconditional abatements in the MCG is likely to give rise to a similar ability to compensate for updated estimates of the benefits of abatement.

Lastly, many suggestions for climate agreements involve international monetary (or in-kind) transfers through which countries pay to subsidize the abatement efforts of other countries<sup>13,26,34,40–42</sup>. The MCG similarly subsidizes abatements in that positive matching factors effectively reduce the marginal costs of abatements. However, in the

MCG, countries' subsidies are implemented not by transferring resources, but through additional abatements provided conditionally. This type of subsidy is essentially unilateral and seems more likely to be implemented than one that involves monetary or in-kind transfers, especially in the case of countries between which diplomatic relations are tense or nonexistent.

**Related literature.** To our knowledge, Rübbelke's<sup>39</sup> was the first to analyze a matching-commitment agreement to reduce GHG emissions, in which each country declares a single matching factor at which it will match any other country's abatement. As is common in the literature on matching-commitment agreements, his analysis is restricted to identical countries and symmetric equilibria, whereas we allow for heterogeneous countries and asymmetric equilibria. The style of our model is also different from Rübbelke's<sup>39</sup> in that—similarly to Boadway *et al.*<sup>21</sup>—we use Finus's<sup>22</sup> well-established emissions game to specify payoffs from emissions profiles (ref. <sup>39</sup> instead builds on the standard model of public goods, see, e.g., ref. <sup>43</sup>). Our analysis is closest in spirit to Boadway *et al.*'s<sup>21</sup>, with the most important difference being this: Boadway *et al.*'s main result concerns the efficacy of the matching climate game (MCG) in emissions reductions relative to an exogenously-determined baseline scenario in which at least one country would benefit from unilaterally reducing its emissions—in which case emissions would be reduced even without an agreement, if this country simply followed its incentives. By contrast, our main contribution here is to show that the MCG incentivizes countries to lower emissions relative to an endogenously-determined baseline scenario from which they have no incentive to deviate.

Matching mechanisms are most commonly studied in the general context of public goods or externality problems, and in that context our analysis is closest to Guttman and Schnytzer<sup>44</sup>. However, those authors (a) show that a subgame-perfect equilibrium (SPE) exists for positive externalities, and show that their approach may not work for negative externalities (such as GHG emissions, analyzed here), and (b) do not establish that matching commitments yield unique equilibrium actions (i.e., an emissions profile in a climate agreement), whereas we are able to do so. Note also that in the context of public goods provision, Buchholz *et al.*<sup>45,46</sup> have shown that when players are heterogeneous, matching often leads to inefficient equilibria at which not all players contribute. Our analysis shows that this problem is avoided in our approach, regardless of any possible inequalities between the countries.

Lastly, we mention that another growing branch of the research into social dilemmas in general, and climate governance in particular, involves the application of methods from statistical physics (see ref. <sup>47</sup> for a review, or ref. <sup>48</sup> for a notable example). Broadly, this literature takes a dynamic, evolutionary game-theoretic perspective, in which the independent agents (e.g., people, but also possibly countries) interact many times, and over time alter their strategies based on the payoffs they and others receive (but do not strategically analyze the game they are facing). To our knowledge, these methods have not yet been applied to matching-commitment agreements, and are a promising avenue for circumventing the technical difficulty of analyzing matching-commitment agreements among many countries in the “classical” game-theoretic setting.

**Main simplifying assumptions.** Our analysis makes use of a number of simplifying assumptions about climate agreements, the most important of which are discussed below.

*One-shot game.* Similar to, e.g., ref. <sup>49</sup>, our simple formulation of the choice of GHG emissions levels in a one-shot game forces countries to make all environmental policy decisions at one point in time. It thus removes the possibility of countries responding to either observed environmental changes or one another's behaviour. In Barrett's<sup>49</sup> words, it “can be interpreted as compressing perhaps a century of decision-making into a single period,” and the damage and benefit in question “should be interpreted as capturing the full consequences, including into the distant future, of decisions taken this century.” Some of the unforeseen environmental consequences of these emissions will only be experienced in the far future, so no observations of such consequences are available to affect policy in the near future. However, the assumption that countries cannot observe and respond to each other's behaviour is more restrictive. Even if the full consequences of emissions are not immediately apparent, countries can still assess other countries' emissions, and would likely vary their own emissions in response. We therefore regard this model as a first step and proof-of-concept for the use of matching-commitment agreements to lower GHG emissions.

There is, however, some experimental evidence that matching-commitment agreements can increase the provision of public goods. In particular, Guttman<sup>50</sup> and Bracht *et al.*<sup>51</sup> observe that matching-commitment mechanisms similar to that employed in our matching climate game (MCG) yield significant improvements in public good provision in repeated public goods games played in groups, and between pairs, respectively (see refs. <sup>52,53</sup> for reviews of the experimental literature on matching).

*Commitment.* Because our model is a one-shot game, it only addresses the issue of commitment in a limited fashion: even if country 2 could renege on the matching factor it committed to in stage I, it has no incentive to do so, and in particular, it does better playing a cooperative subgame-perfect equilibrium (SPE) than it does by reverting to its baseline emissions, i.e., choosing  $m_2 = 0$  and  $a_2 = 0$ . This is because, even if country 1 continues to play  $m_1^c$ , it can still respond to country 2's defection by decreasing its unconditional abatement,  $a_1(m_1^c, m_2)$ . But, if both countries are free to renege on their matching commitments, stage I is merely “cheap talk” and the original basic climate game (BCG; Definition A.1)—in which the baseline emissions profile is the only equilibrium—is recovered.

In the language of Barrett<sup>14</sup>, the ability of countries to commit assumes the “enforcement problem” is solved. As Barrett<sup>49</sup> notes, this is justified by the international-law custom of treaties being binding, *pacta sunt servanda*. This custom is codified in the Vienna Convention on the Law of Treaties: “Every treaty in force is binding upon the parties to it and must be performed by them in good faith” (ref. <sup>54</sup>, p. 339). Thus, if countries can agree that whatever matching commitments they themselves select be *legally binding*, they are more likely to keep their



word. Since countries are also free to refrain from any matching (i.e., matching factors may be zero), and since the matching commitments are conditional on other countries' actions, such commitments are weaker—and hence safer from being exploited, and thus more likely to be kept—than commitments to specific emissions abatements.

Two recent examples suggest that legally binding commitments to match emissions reductions would likely be an improvement on the status quo, even without an explicit enforcement mechanism in place. First, some parties to both the Kyoto Protocol and the Paris Agreement rejected legally binding targets<sup>55,56</sup>; this suggests that countries ascribe more weight to commitments that are explicitly legally binding, possibly due to the risk of reputation loss<sup>57</sup>. Second, although the Paris Agreement's emissions targets are not binding, other aspects of the agreement are, and even the withdrawal of the US from the Paris Agreement will conform to the agreement's rules for withdrawal<sup>58</sup>.

*Two vs. many countries.* While it is widely acknowledged that an effective agreement must include as parties the two largest emitters, the US and China, their combined emissions amount to about 45% of global emissions, so a climate agreement between only these countries will likely be insufficient<sup>3</sup>. We leave the analysis of matching-commitment agreements among many countries for future work, but note that a number of studies suggest that such agreements among many countries can also be effective in reducing emissions. For instance, encouraging analytical results include those of Fujita<sup>35</sup> and Rübhelke<sup>39</sup> (described in the sections “Advantages over other climate-agreement schemes” and “Related literature”). Using the STACO model<sup>40</sup> to estimate countries' damages and benefits from emissions, Kawamata and Horita<sup>59</sup> simulate a matching scheme similar to the matching climate game (MCG) for six countries and regions and show that emissions are substantially reduced. Lastly, in Guttman's<sup>50</sup> experimental study, matching commitments also increased public good provision in repeated games in groups of 3–6 players.

In the context of multilateral matching-commitment agreements, it will likely be imperative that *all* countries be signatories. This is because although matching-commitment agreements do not strictly require signatories to perform any abatement, they do require that countries refrain from emitting more than they would in the absence of an agreement. If some countries opt out of the agreement, and signatories collectively abate their total emissions, then non-signatories will have an incentive to increase their emissions from their no-agreement baseline emissions (this is the carbon leakage mentioned in the section “Advantages over other climate-agreement schemes”).

*Increasing damages from climate change.* Ricke *et al.*<sup>60</sup> recently estimated the country-level social cost of carbon (CSCC)—a measure of the long-run economic cost of a tonne of CO<sub>2</sub> emitted in 2020. The CSCC is analogous to the marginal damage suffered by each country,  $D'$ ; it quantifies the marginal cost of emissions by comparing a reference emissions pathway (describing how emissions are distributed over a length of time) to a pathway that begins with an *additional* pulse of one tonne of CO<sub>2</sub> emissions, but is otherwise identical to the reference pathway. Ricke *et al.*<sup>60</sup> repeat these calculations for a variety of different socioeconomic scenarios (that result in different emissions pathways) and for different empirical estimates of the economic impacts of warming. (Note that warming, and consequently damage due to climate change, does not depend strongly on the emissions pathway and is well described by the cumulative emissions<sup>61,62</sup>). Thus, our formulation assumes that damages result from global cumulative emissions over a long time-horizon, ignoring the possible impact of the emissions pathway).

While one of Ricke *et al.*'s<sup>60</sup> model specifications results in a negative CSCC (i.e., a climate *benefit*) for some countries, they note that “[t]hese results are among the most sensitive in the analysis, as under the BHM long-run and DJO damage model specifications all countries have positive CSCC”, where BHM and DJO refer to the damage-function estimates of Burke *et al.*<sup>63</sup> and Dell *et al.*<sup>64</sup>, respectively. Moreover, they find that “under the long-run impact model specifications [...] countries like Russia, Canada, Germany and France that have negative CSCC under the reference case switch to having among the highest positive CSCCs”. Another analysis by Kompas *et al.*<sup>65</sup> applies a “computable general equilibrium” approach and finds that all of the 139 countries included in their analysis benefit from complying with the Paris Agreement, providing additional support for our assumption that the marginal damages of emissions are positive.

**Relaxation of assumptions on the damages and benefits of emissions.** The assumption that global emissions are bounded (Definition A.2) can be relaxed: its function is to ensure that the basic climate game (BCG) has a unique Nash equilibrium, but all that is required for our results to hold is that the BCG has a Nash equilibrium, and that it is chosen as the baseline emissions profile. In contrast, the assumption of bounded abatement with matching (BAM) is both necessary and sufficient for the matching climate game's (MCG) stage-II best-response functions to be well-defined, and hence cannot be relaxed.

**Application to the provision of general public goods.** While the matching climate game (MCG) and our analysis are framed in terms of climate agreements, they can be applied equally well to other public goods problems. The public good provided in our model is the *abatement* of GHG emissions, and so reduced emissions correspond to increased abatement. This in turn implies that our assumption that the benefits and damages of emissions are decelerating and accelerating, respectively, translate into requiring that the costs and benefits of providing the public good being accelerating and decelerating, respectively. Our analysis then shows that the MCG may also be useful in promoting the provision of public goods satisfying these assumptions, in situations in which enforcing cooperation is difficult<sup>46,53,66</sup>.

Received: 19 February 2020; Accepted: 24 March 2020;

Published online: 19 June 2020

## References

- Pachauri, R. K., Meyer, L. & Core Writing Team (eds.) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR\\_AR5\\_FINAL\\_full\\_wcover.pdf](http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full_wcover.pdf) (IPCC, Geneva, Switzerland, 2014).
- United Nations. *Paris Agreement*. [https://treaties.un.org/pages/ViewDetails.aspx?src=TREATYmtdsg\\_no=XXVII-7-dchapter=27clang=en](https://treaties.un.org/pages/ViewDetails.aspx?src=TREATYmtdsg_no=XXVII-7-dchapter=27clang=en) (2015).
- Schreurs, M. A. The Paris climate agreement and the three largest emitters: China, the United States, and the European Union. *Politics Gov.* **4**, 219–223, <https://doi.org/10.17645/pag.v4i3.666> (2016).
- Brown, P. T. & Caldeira, K. Greater future global warming inferred from earth's recent energy budget. *Nature* **552**, 45–50, <https://doi.org/10.1038/nature24672> (2017).
- Millar, R. J. *et al.* Emission budgets and pathways consistent with limiting warming to 1.5 °C. *Nat. Geosci.* **10**, 741–747, <https://doi.org/10.1038/ngeo3031> (2017).
- Chen, C. & Zeckhauser, R. Collective action in an asymmetric world. *J. Public Econ.* **158**, 103–112, <http://www.sciencedirect.com/science/article/pii/S0047272717302116> (2018).
- Hardin, G. The tragedy of the commons. *Science* **13**, 1243–1248, <https://doi.org/10.1126/science.162.3859.1243> (1968).
- Nordhaus, W. Climate clubs: Overcoming free-riding in international climate policy. *Am. Econ. Rev.* **105**, 1339–1370, <http://www.aeaweb.org/articles?id=10.1257/aer.15000001> (2015).
- Barrett, S. Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation. *Proc. Natl. Acad. Sci. USA* **113**, 14515–14522, <https://doi.org/10.1073/pnas.1604989113> (2016).
- UNEP. *The Emissions Gap Report 2017*. <https://www.unenvironment.org/resources/emissions-gap-report> (United Nations Environment Programme (UNEP), Nairobi, 2017).
- Tokarska, K. B. & Gillett, N. P. Cumulative carbon emissions budgets consistent with 1.5 °C global warming. *Nat. Clim. Change* **8**, 296–299, <https://doi.org/10.1038/s41558-018-0118-9> (2018).
- Wood, P. J. Climate change and game theory. *Ann. N. Y. Acad. Sci.* **1219**, 153–170, <https://doi.org/10.1111/j.1749-6632.2010.05891.x> (2011).
- Falkinger, J., Hackl, F. & Pruckner, G. J. A fair mechanism for efficient reduction of global CO<sub>2</sub>-emissions. *FinanzArchiv/Public Finance Analysis* **53**, 308–331, <http://www.jstor.org/stable/40912732> (1996).
- Barrett, S. Climate treaties and the imperative of enforcement. *Oxf. Rev. Econ. Policy* **24**, 239–258, <https://doi.org/10.1093/oxrep/grn015> (2008).
- Gerber, A. & Wichardt, P. C. Providing public goods in the absence of strong institutions. *J. Public Econ.* **93**, 429–439, <https://doi.org/10.1016/j.jpubeco.2008.10.006> (2009).
- Bos, O., Roussillon, B. & Schweinzer, P. Agreeing on efficient emissions reduction. *Scand. J. Econ.* **118**, 785–815, <https://doi.org/10.1111/sjoe.12164> (2016).
- Barrett, S. The problem of global environmental protection. *Oxf. Rev. Econ. Policy* **6**, 68–79, <https://doi.org/10.1093/oxrep/6.1.68> (1990).
- Barrett, S. *Convention on Climate Change: Economic Aspects of Negotiations*. [https://openlibrary.org/books/OL1348009M/Convention\\_on\\_climate\\_change](https://openlibrary.org/books/OL1348009M/Convention_on_climate_change) (Organization for Economic Cooperation & Development, Paris, 1992).
- Guttman, J. M. Understanding collective action: matching behavior. *Am. Econ. Rev.* **68**, 251–255, <http://www.jstor.org/stable/1816699> (1978).
- Guttman, J. M. A non-Cournot model of voluntary collective action. *Economica* **54**, 1–19, <http://www.jstor.org/stable/2554340> (1987).
- Boadway, R., Song, Z. & Tremblay, J.-F. The efficiency of voluntary pollution abatement when countries can commit. *Eur. J. Political Econ.* **27**, 352–368, <https://doi.org/10.1016/j.ejpoleco.2010.10.003> (2011).
- Finus, M. *Game Theory and International Environmental Cooperation*, New Horizons in Environmental Economics. <https://books.google.at/books?id=a8u0AAAIAAJ> (Edward Elgar, 2001).
- Löschel, A. & Rübbelke, D. T. G. On the voluntary provision of international public goods. *Economica* **81**, 195–204, <https://doi.org/10.1111/ecca.12081> (2014).
- Maschler, M., Zamir, S. & Solan, E. *Game Theory*. <https://books.google.at/books?id=lqwzqgvhwXsC> (Cambridge University Press, 2013).
- Folmer, H. & von Mouche, P. The acid rain game: a formal and mathematically rigorous analysis. In Kriström, B., Dasgupta, P. & Löfgren, K.-G. (eds.) *Economic Theory for the Environment: Essays in Honour of Karl-Göran Mäler*, Contemporary Trends in European Social Sciences. 138–161, <https://books.google.ca/books?id=BuQkAQAAMAAJ> (Edward Elgar Publishing, 2002).
- Buchholz, W. & Konrad, K. A. Strategic transfers and private provision of public goods. *J. Public Econ.* **57**, 489–505, <http://www.sciencedirect.com/science/article/pii/004727279580008W> (1995).
- Panayotou, T., Sachs, T. & Peterson, A. Developing Countries and the Control of Climate Change: A Theoretical Perspective and Policy Implications. CAER II Discussion Paper No. 44, <https://pdfs.semanticscholar.org/4a05/2fa713252eadb020ced77323cbf5c0d7a360.pdf> (Harvard Institute for International Development, 1999).
- Golombek, R. & Hoel, M. Unilateral emission reductions and cross-country technology spillovers. *B. E. J. Econ. Anal. Policy* **4**, <https://doi.org/10.2202/1538-0637.1318> (2004).
- Hoekstra, R. & Janssen, M. A. Environmental responsibility and policy in a two-country dynamic input–output model. *Econ. Systems Res.* **18**, 61–84, <https://doi.org/10.1080/09535310500440894> (2006).
- Di Maria, C. & van der Werf, E. Carbon leakage revisited: unilateral climate policy with directed technical change. *Environ. Resour. Econ.* **39**, 55–74, <https://doi.org/10.1007/s10640-007-9091-x> (2008).
- Baylis, K., Fullerton, D. & Karney, D. H. Negative leakage. *J. Assoc. Environ. Resour. Econ.* **1**, 51–73, <https://doi.org/10.1086/676449> (2014).
- Kim, S. E. & Urpelainen, J. Technology competition and international co-operation: Friends or foes? *Brit. J. Polit. Sci.* **44**, 545–574, <https://doi.org/10.1017/S0007123412000762> (2014).
- van Vuuren, D. P. *et al.* Alternative pathways to the 1.5 °C target reduce the need for negative emission technologies. *Nat. Clim. Change* **8**, 391–397, <https://doi.org/10.1038/s41558-018-0119-8> (2018).
- Carraro, C., Eyckmans, J. & Finus, M. Optimal transfers and participation decisions in international environmental agreements. *Rev. Int. Organ.* **1**, 379–396, <https://doi.org/10.1007/s11558-006-0162-5> (2006).
- Fujita, T. A self-enforcing international environmental agreement on matching rates: Can it bring about an efficient and equitable outcome? *Strateg. Behav. Environ.* **3**, 329–345, <https://doi.org/10.1561/102.00000033> (2013).
- Buchholz, W., Cornes, R. & Rübbelke, D. T. G. Potentially harmful international cooperation on global public good provision. *Economica* **81**, 205–223, <https://doi.org/10.1111/ecca.12074> (2014).
- Oliver, P. Rewards and punishments as selective incentives for collective action: Theoretical investigations. *Am. J. Sociol.* **85**, 1356–1375, <http://www.jstor.org/stable/2778382> (1980).
- Yamagishi, T. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116, <https://doi.org/10.1037/0022-3514.51.1.110> (1986).
- Rübbelke, D. T. G. Analysis of an international environmental matching agreement. *Environ. Econ. Policy Stud.* **8**, 1–31, <https://doi.org/10.1007/BF03353991> (2006).

40. Finus, M., van Ierland, E. & Dellink, R. Stability of climate coalitions in a cartel formation game. *Econ. Gov.* **7**, 271–291, <https://doi.org/10.1007/s10101-005-0009-1> (2006).
41. Altmeppen-Bartscher, M., Rübbelke, D. T. G. & Sheshinski, E. Environmental protection and the private provision of international public goods. *Economica* **77**, 775–784, <https://doi.org/10.1111/j.1468-0335.2009.00796.x> (2010).
42. Buchholz, W., Falkinger, J. & Rübbelke, D. T. G. Non-governmental public norm enforcement in large societies as a two-stage game of voluntary public good provision. *J. Public Econ. Theory* **16**, 899–916, <https://doi.org/10.1111/jpet.12084> (2014).
43. Varian, H. R. *Intermediate Microeconomics: A Modern Approach: Ninth International Student Edition*. <https://books.google.ca/books?id=A0feCgAAQBAJ> (WW Norton & Company, 2014).
44. Guttman, J. M. & Schnytzer, A. A solution of the externality problem using strategic matching. *Soc. Choice Welf.* **9**, 73–88, <https://doi.org/10.1007/BF00177674> (1992).
45. Buchholz, W., Cornes, R. & Rübbelke, D. T. G. Interior matching equilibria in a public good economy: An aggregative game approach. *J. Public Econ.* **95**, 639–645, <https://doi.org/10.1016/j.jpubeco.2010.12.010> (2011).
46. Buchholz, W., Cornes, R. & Rübbelke, D. T. G. Matching as a cure for underprovision of voluntary public good supply. *Econ. Lett.* **117**, 727–729, <http://www.sciencedirect.com/science/article/pii/S0165176511006185> (2012).
47. Perc, M. *et al.* Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51, <https://doi.org/10.1016/j.physrep.2017.05.004> (2017).
48. Pacheco, J. M., Vasconcelos, V. V. & Santos, F. C. Climate change governance, cooperation and self-organization. *Phys. Life Rev.* **11**, 573–586, <https://doi.org/10.1016/j.plrev.2014.02.003> (2014).
49. Barrett, S. Climate treaties and approaching catastrophes. *J. Environ. Econ. Manag.* **66**, 235–250, <https://doi.org/10.1016/j.jeem.2012.12.004> (2013).
50. Guttman, J. M. Matching behavior and collective action: Some experimental evidence. *J. Econ. Behav. Organ.* **7**, 171–198, [https://doi.org/10.1016/0167-2681\(86\)90004-1](https://doi.org/10.1016/0167-2681(86)90004-1) (1986).
51. Bracht, J., Figuières, C. & Ratto, M. Relative performance of two simple incentive mechanisms in a public goods experiment. *J. Public Econ.* **92**, 54–90 <http://www.sciencedirect.com/science/article/pii/S0047272707000618> (2008).
52. Epperson, R. & Reif, C. Matching schemes and public goods: A review. *ZEW Discussion Papers 17-070*, ZEW - Zentrum für Europäische Wirtschaftsforschung/Center for European Economic Research, <http://ftp.zew.de/pub/zew-docs/dp/dp17070.pdf> (2017).
53. Reif, C., Rübbelke, D. T. G. & Löschel, A. Improving voluntary public good provision through a non-governmental, endogenous matching mechanism: Experimental evidence. *Environ. Resour. Econ.* **67**, 559–589, <https://doi.org/10.1007/s10640-017-0126-7> (2017).
54. United Nations. *Vienna Convention on the Law of Treaties*. [https://treaties.un.org/Pages/ViewDetailsIII.aspx?src=TREATYmtdsg\\_no=XXIII-1chapter=23Temp=mtsg3clang=\\_en](https://treaties.un.org/Pages/ViewDetailsIII.aspx?src=TREATYmtdsg_no=XXIII-1chapter=23Temp=mtsg3clang=_en) (1969).
55. Bolin, B. The Kyoto negotiations on climate change: A science perspective. *Science* **279**, 330–331, <https://doi.org/10.1126/science.279.5349.330> (1998).
56. Dimitrov, R. S. The Paris agreement on climate change: Behind closed doors. *Glob. Environ. Politics* **16**, 1–11, [https://doi.org/10.1162/glep\\_a\\_00361](https://doi.org/10.1162/glep_a_00361) (2016).
57. Simmons, B. Treaty compliance and violation. *Annu. Rev. Political Sci.* **13**, 273–296, <https://doi.org/10.1146/annurev.polisci.12.040907.132713> (2010).
58. Harvey, C. Withdrawing from the Paris deal takes four years. Our next president could join again in 30 days. [https://www.washingtonpost.com/news/energy-environment/wp/2017/06/05/withdrawing-from-the-paris-deal-takes-four-years-our-next-president-could-join-again-in-30-days/?utm\\_term=.b48447778582](https://www.washingtonpost.com/news/energy-environment/wp/2017/06/05/withdrawing-from-the-paris-deal-takes-four-years-our-next-president-could-join-again-in-30-days/?utm_term=.b48447778582) (2017).
59. Kawamata, K. & Horita, M. Applying matching strategies in climate change negotiations. *Group Decis. Negot.* **23**, 401–419, <https://doi.org/10.1007/s10726-013-9354-6> (2014).
60. Ricke, K., Drouet, L., Caldeira, K. & Tavoni, M. Country-level social cost of carbon. *Nat. Clim. Chang.* **8**, 895–900, <https://doi.org/10.1038/s41558-018-0282-y> (2018).
61. Allen, M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166, <https://doi.org/10.1038/nature08019> (2009).
62. Matthews, H. D., Solomon, S. & Pierrehumbert, R. Cumulative carbon as a policy framework for achieving climate stabilization. *Phil. Trans. R. Soc. A* **370**, 4365–4379, <https://doi.org/10.1098/rsta.2012.0064> (2012).
63. Burke, M. & Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239, <https://doi.org/10.1038/nature15725> (2015).
64. Dell, M., Jones, B. F. & Olken, B. A. Temperature shocks and economic growth: Evidence from the last half century. *Am. Econ. J. Macroecon.* **4**, 66–95, <https://doi.org/10.1257/mac.4.3.66> (2012).
65. Kompas, T., Pham, V. H. & Che, T. N. The effects of climate change on GDP by country and the global economic gains from complying with the Paris climate accord. *Earth's Future* **6**, 1153–1173, <https://doi.org/10.1029/2018EF000922> (2018).
66. Buchholz, W. & Liu, W. Global public goods and unilateral matching mechanisms. *J. Public Econ. Theory* **22**, 338–354, <https://doi.org/10.1111/jpet.12300> (2018).

## Acknowledgements

The preparation of this paper was supported by grants from the Warren Center for Network and Data Sciences at the University of Pennsylvania to C.M., from the US Army Research Office (W911NF-12-R-0012-03) to E.A., from the Defense Advanced Research Projects Agency NGS2 program (grant no. D17AC00005) to E.A. and S.A.L. and from the National Science Foundation (grant no. GEO-1211972), the Nordforsk-funded project Green Growth Based on Marine Resources: Ecological and Socio-Economic Constraints (GreenMAR) and the US Army Research Office (grant no. W911NF1810325) to S.A.L. We thank Robin Boadway for his valuable comments on an earlier draft.

## Author contributions

C.M. devised the project, analyzed the model and drafted the manuscript in collaboration with E.A., U.D., S.A.L. and E.A.R. All authors interacted in model development, provided input and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63446-1>.

**Correspondence** and requests for materials should be addressed to C.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020