Agreeing on efficient emissions reduction*

Olivier Bos† Béatrice Roussillon‡ Paul Schweinzer§

Abstract

We propose a simple mechanism providing incentives to reduce harmful emissions to their efficient level without infracting upon productive efficiency. The mechanism employs a contest creating incentives among participating nations to simultaneously exert efficient productive and efficient abatement efforts. Participation in the most stylised form of the scheme is voluntary and individually rational; all rules are mutually agreeable and are unanimously adopted if proposed. The scheme balances its budget and requires no principal. In a perhaps more realistic stochastic output version which could potentially inform policy decisions, we show that the transfers required by the efficient mechanism create a mutual insurance motive which may serve as effective rationale for the (gradual) formation of International Environmental Agreements. (JEL *C7, D7, H4, Q5.* Keywords: *Climate policy, Contests, Agreements.*)

1 Introduction

The disappointing series of failures to reach agreement among the 194 members of the United Nations Framework Convention on Climate Change in Copenhagen (2009), Cancún (2010), Durban (2011), Doha (2012), and Warsaw (2013) highlights the international impasse in preventing further global warming. Yet action seems to be called for: recent research reports shrinking ice mass balance from both Greenland and Antarctica with a projected sea-level rise of one to two meters by 2100.[1] Since an estimated 180 million people live currently in locations less than one meter above sea level, the impact of this change on the world economy will be substantial.[2] This paper studies and answers two questions arising in this context: i) How can incentives be provided to reduce harmful

*Thanks for comments to Heski Bar-Isaac, Valentina Bosetti, Reyer Gerlagh, Alex Gershkov, Thomas Giebe, Robert O. Keohane, Dan Kovenock, Jianpei Li, Philipp Wichardt and seminar participants at the University of Manchester, Paris School of Economics, University of Milan, University of Copenhagen, University of Munich, Collège de France and CESifo, Munich. This paper supersedes two previously independent working papers, *Efficient emissions reduction*, and *Risk pooling in redistributive agreements*. Schweinzer appreciates the generous hospitality of CESifo and his co-authors' employers. Financial support from the University of York Research and Impact Support Fund is gratefully acknowledged. †LEM-Department of Economics, University Panthéon-Assas, 5/7 Avenue Vavin, 75006 Paris, France, olivier.bos@u-paris2.fr. ‡Université Pierre Mendès France, UMR 1215 GAEL and INRA, UMR 1215 GAEL, 38040 Grenoble Cedex 9, France, beatrice.roussillon@upmf-grenoble.fr. §Department of Economics, University of York, Heslington, York YO10 5DD, United Kingdom, paul.schweinzer@york.ac.uk. (30-Jan-2014)

[1] See, for instance, Dasgupta et al. (2009) or Allison et al. (2009). Mitrovica, Gomez, and Clark (2009) predict less uniform sea level changes with a rise of up to 6.3 meters at some coastal sites in the northern hemisphere upon a total collapse of the West Antarctic Ice Sheet.

[2] The original estimate of 180 million is from Nicholls (1995); Nicholls et al. (2008) estimate the economic effects of climate change on coastal cities and ports. A recent analysis of migration induced through climate change is, for instance, Kniveton, Smith, and Black (2012).

emissions to their socially efficient levels while not infracting upon productive efficiency? ii) How can international agreement on the parameters of this mechanism be found?

In our simple model environment, there are two ways to reduce emissions: by producing less or by abating more. The mechanism we propose plays on these two aspects in order to achieve efficiency in both: a stylised contest— based on a relative ranking of all nations' abatement efforts—rewards the countries with the highest abatement efforts with some share of joint agreement output. In a nutshell, marginal production is 'taxed' to fund a prize pool, and marginal abatement increases the probability of winning a share of this pool. By designing the contest appropriately, both equilibrium incentives can be set efficiently at the margin. The precision with which the contest ranking is correct, i.e., the precision of mutual abatement monitoring, is one of the design parameters of our proposed mechanism.

In equilibrium, this efficiency inducing contest takes the form of a redistributive mechanism which returns some share of the collected tax pool in the form of prizes to the participants. As it turns out, for stochastic individual GDP, the variance of the redistributed income is lower than that of individual income. The redistributive contest mechanism can therefore fulfill aspects of a mutual insurance agreement which can entice a country with a sufficiently strong dislike of income fluctuations to join an agreement on which, in the absence of this income smoothing argument, it would prefer to free ride.

The main emissions type we have in mind for our model is greenhouse gases. These are widely seen as the main contributing factor to global warming. Emitted by one country as inherent part of the productive process, they are distributed around the globe regardless of where they were produced and, as such, present an externality. A reduction of emissions benefits all countries but the costs of such reductions are carried individually. This generates a classic free-rider problem in which each country would like the threat of global warming removed but none is ready to pay the cost.[3] We think of the abatement efforts as the difference between Business-As-Usual (BAU) investments and green investments. BAU investments are the investments that firms would have made without considering its impacts on the environment. Green investments have the same goal as BAU investments but, in addition, purposefully intend to reduce the CO_2 emissions generated during production.

The environmental literature employs three main approaches to overcome the inefficient emissions abatement problem: command-and-control regulation, quantity-oriented market approaches, and tax-or-pricing regimes. The approach adopted by the 187 signatories of the Kyoto protocol is the quantity-oriented market approach targeting a reduction based on developed countries' emissions in 1990.[4] The treaty, however, failed to obtain ratification by major players including, most prominently, the United States. Moreover, the concern was expressed that developing countries might have ratified the treaty without the intention of keeping emissions in check. This mars the current emissions reduction reality with the dual frustrations of insufficient participation and diluted objectives.

[3] One may advocate the view that some countries could climatically benefit from warming. Russian President Vladimir Putin, for instance, is reported to have said that climate change might be good for his country, as people would no longer need to buy fur coats (Reuters, 2-April-07). The impact on the world economy and consequences in terms of migration, however, make us pessimistic about the likelihood of emerging net beneficiaries.

[4] For details see, for instance, Barrett (1998) and Nordhaus (2006).

Nordhaus (2006) advocates the implementation of market-based instruments and, more specifically, a world harmonised tax on each ton of CO_2 emissions, the revenue from which may be used at will by each national government. This proposal presents many advantages over the status quo; it can achieve efficiency at the world level, it is simple to implement, and it is a well-known instrument. With the proposed harmonised tax, however, each country pays the same price/tax for a ton of CO_2 emissions regardless of its revenue and responsibility for climate change. This could be particularly hard on poor countries and reduce their incentives to join any international agreement.[5] Moreover, imposing a new tax (unilaterally) may be politically unattractive while the individualised winning incentives that our contest scheme presents make participation individually attractive. A harmonised tax—which can only work if all countries participate—is based on the 'polluter pays' principle whereas our mechanism is driven by a 'cleaner wins' principle which we believe could improve participation by developing countries. The reason is that there are many 'low hanging fruits' in developing countries representing easy abatement possibilities and thus good chances to win the competition we suggest.

One may recommend complementing a harmonised carbon tax with subsidising abatement efforts. Such subsidies will, however, typically lead to allocative inefficiencies. As abatement efforts are difficult to measure in absolute terms, countries may be tempted to present productive investment as abatement efforts to get higher subsidies. In contrast, the informational requirements for a contest based on a relative ranking may be easier to satisfy than those for a cardinal scale. Moreover, a contest prize can be fixed independently of the competitors' abatement effort levels while fixed subsidies based on piece-rates depend on absolute levels. As our ranking is relative, it would not be affected by an overall abatement measurement inflation.[6] Thus, a ranking allows to use indirect abatement measurements which could be less easily manipulated than other, cardinal, measures entering tax calculations.

This paper makes two key contributions. First, we show that a contest can implement efficiency in a specific environmental model along both productive and abatement effort dimensions. Second, we demonstrate that (gradual) agreement formation is possible in this model if nations are sufficiently averse to income-variance. Both results are theoretical. In addition, we provide a benchmark result showing which model-resources it takes to implement the first-best solution under the following objectives: absence of a principal, efficiency (i.e., no distortions of the welfare maximising allocation) in both effort dimensions, and budget-balancing. Even if the highly stylised mechanism we discuss may seem unrealistic and difficult to implement directly, we hope that our analysis can deliver new and significant insights on the available options, on the cost of abatement, and on the implementation of agreement mechanisms.

[5] Nordhaus argues that countries should not have to pay the tax until they achieve a certain level of revenue or otherwise should receive subsidies. This, however, leaves no true participation incentive to poor countries.

[6] For instance, our ranking could be based on the difference between a country's BAU forecast for CO_2 emissions and the country's realised emissions with the highest difference country winning. In such a situation, even if countries should symmetrically increase BAU emission forecasts, the relative odds remain the same.

Related literature

Efficiency. The idea behind our main efficiency result relates to Gershkov, Li, and Schweinzer (2009) who analyse the efficient effort choice in team-partnership problems.[7] They construct a contest mechanism where total partnership output is used as a remuneration pool for the partners. The intuition is that, through a relative ranking of efforts, the highest-effort partner expects the biggest share of joint output. Therefore, incentives are generated to exert high efforts which offset the free-riding problem inherent in the partnership. In contrast to our paper, their partnership setup has only one effort dimension and, consequently, there is only one free-riding problem. The emissions reduction problem analysed in the present paper has two such problems: overproduction of harmful emissions and underprovision of reductive or abatement efforts. Moreover, our specific environmental problem setup has little in common with their modelling.

The most directly related studies of the public goods provision problem relating to contests that we are aware of are Morgan (2000) and Giebe and Schweinzer (2013). Morgan (2000) studies a lottery mechanism which uses proceeds obtained from ticket sales for the provision of a public good. Contrasting with our analysis, he is neither concerned with designing a mechanism to implement exact efficiency nor with balancing the mechanism's budget. Using a fixed precision contest, he obtains the result that contests are unable to implement exact efficiency. In a theoretical paper, Giebe and Schweinzer (2013) explore the possibility of the efficient provision of a public good through non-distortionary taxation of a private good which is linked to a lottery. By fine tuning the tax with the lottery, they are able to get efficient consumption levels for both private and public goods. This is in the same spirit as the present analysis, but their individual public good contribution is only a function of the private good consumption and not at the individual's discretion. Moreover, our environmental model needs to balance two dimensions of inefficiency: excessive production and insufficient reduction of emissions. This is impossible in their single-dimensional model where tailoring only private goods consumption can lead to efficiency.

In a paper not relating to contests, Buchholz, Cornes, and Rübbelke (2011) study the existence of equilibria in an aggregative game on which they impose the efficient allocation. In particular, they analyse a matching mechanism in which agents cross-subsidise each other in order to achieve efficiency and study the question which income distributions can be compatible with voluntary and efficient provision of the public good. Contrasting the present analysis, the authors do not explore the design of explicit mechanisms capable of providing incentives for the implementation of this efficient matching equilibrium. Gerber and Wichardt (2009) develop a deposit mechanism that results in the efficient provision of a public good. In their mechanism, countries first commit to paying a fixed fee and then choose their abatement effort. Once a country is committed, its utility maximising strategy is to provide efficient abatement effort. More specifically, a global fund is created with the collected fixed fees and redistributed among countries according to their public good provision

[7] The idea that in many circumstances efficient efforts can be induced by awarding a prize on the basis of a rank order among competitors' efforts is due to Lazear and Rosen (1981). This insight has found numerous applications and extensions, for instance in the work of Green and Stokey (1983), Nalebuff and Stiglitz (1983), Dixit (1987), Moldovanu and Sela (2001), or Siegel (2009). For a detailed survey of the contests literature see the comprehensive Konrad (2008).

levels. If the public good is provided inefficiently, a country is punished by losing the committed funds contained in the global fund. In a two-periods setup, Gersbach and Winkler (2012) design a global refunding scheme and analyse its potential for mitigating climate change. Each country has to pay a fixed fee and determines its own emission tax at each period. Then a participating country is refunded at each period given its relative emission reduction. As we show in the discussion of our benchmark efficiency case, such a tax-based system cannot implement efficiency in general. In an emissions permit trading model, Gersbach and Winkler (2011) develop a model designed to limit free-riding in the form of countries allocating inefficiently many permits. In their model, part of the permits are allocated for free, whereas part is auctioned. The money raised through the auction goes into a global fund and is used to reimburse participating countries according to some previously negotiated share.[8] Although we share important ideas with all three papers, neither models productive efforts and therefore cannot consider the issue of efficiently balancing polluting overproduction with abatement efforts.

Agreements. Our team setup is vindicated by the universally accepted property of international environmental agreements (IEA) to be self-enforcing. Indeed, there is no supranational principal to enforce such arrangements between countries. The consensus in the literature seems to be that there is no consistent theoretical basis on which large voluntary agreements can be formed among independent nation states which do not have to revert to exogenous reward, punishment or exclusion strategies to avoid free-riding on emissions reduction.[9] The main contributions have found that IEA are either unlikely to consist of many participants, or if they do, are similarly unlikely to produce substantial benefits. In particular, Diamantoudi and Sartzetakis (2006) show that no more than four countries will find it profitable to form a coalition regardless of the number of countries participating in the negotiations. Kolstad (2007) demonstrates that the size of stable IEAs decreases as uncertainty grows. Besides, the outcome of non-cooperative coalition formation games depends on the specific membership rules imposed on an IEA. For example, Carraro, Marchiori, and Oreffice (2009) show that the introduction of a minimum participation rule increases the number of signatories.[10] Barrett (2006) studies an alternative to the Kyoto protocol in proposing a system of two treaties, one promoting cooperative 'breakthrough' R&D investments and the other encouraging collective adoption of new technologies emerging from this R&D activity.

On the subject of risk-sharing, which is the main underlying reason for our players to join an agreement, Wilson (1968) discusses the optimal behaviour of a group of individuals, called a syndicate, who must make a common decision under uncertainty that results in a payoff to be shared jointly among them. As in our model, the strictly risk-averse players face individual payoff uncertainty which they pool in the syndicate. The paper derives an optimal investment strategy for the syndicate and is then mainly interested in the derivation of Pareto optimal sharing rules for

[8] In comparison with a permit trading system, our mechanism has the fundamental advantage that an output shock affecting all nations does not change contest efforts at all while it would lead to excess demand or supply in the market for permits. This is in line with both theoretical considerations (see, e.g., Green and Stokey (1983)) and recent experience on the depressed European emissions permit market.

[9] See Barrett (1994), Barrett (2003), Aldy and Stavins (2007), and Guesnerie and Tulkens (2009) for the main results and further discussion.

[10] Chander and Tulkens (2009) show that, typically, the involved contracts are not renegotiation proof.

the syndicate members characterised in terms of, e.g., risk tolerance.[11] Deaton (1991) models the optimal intertemporal consumption behaviour of consumers whose labour income is stochastic over time. He shows that individual savings can act like a buffer stock, protecting individual consumption against shocks. The same idea has been used to model consumption-savings choices and insurance motives across consumers' life-cycles. The same variance-compression idea is present in Green (1987) who discusses tax-financed unemployment insurance and Arnott and Stiglitz (1991) who model voluntary, non-market insurance between households under moral hazard.

Recent contributions to the literature on IEA-membership dynamics include Breton, Sbragia, and Zaccour (2010) and Harstad (2010). Particularly referring to climate change agreements, the latter shows how short-term agreements may have adverse effects on countries' investments in green technology. Indeed, as Buchholz and Konrad (1994) and Beccherle and Tirole (2011) point out, anticipating negotiations can decrease the level of R&D and green investments.[12]

Plan of paper. Following the model definition in section 2, we present the paper's main idea through an illustrative example on which we later expand as new results are introduced. Although stylised, this simple example conveys much of the intuition of the general efficiency results presented in section 3. The model is closed using a simple, simultaneous agreement game in subsection 3.3 based on deterministic output. A much more general (and realistic) participation argument is developed in section 4 for stochastic output. Several model extensions, further examples and an alternative family of success functions are explored in appendix A. All proofs and a further asymmetric extension are in appendix B and supplementary results for the stochastic case can be found in appendix C.

2 The symmetric model

There is a set $\mathcal{N}$ of $n \geq 2$ risk-neutral players. These players are symmetric in the basic model.[13] Each player $i \in \mathcal{N}$ exerts efforts along two dimensions: productive effort $e_i \in [0, \infty]$ and reductive (abatement) effort $f_i \in [0, \infty]$. The reductive efforts need not, in principle, be verifiable.[14] We denote the full effort vectors by $\mathbf{e} = e_1, \dots, e_n$ and $\mathbf{f} = f_1, \dots, f_n$, respectively. The effort costs $c_e(e_i)$ and $c_f(f_i)$ are assumed to be strictly convex and zero for zero effort. Productive efforts generate strictly concave individual gains of $y(e_i)$ and cause strictly convex global emissions of $m(\max\{0, \sum_h e_h - \sum_h f_h\})$—only depending on the difference between global productive and

[11] This idea found multiple notable extensions and applications, for instance, in the work of Wildasin (1995), Banks, Blundell, and Brugiavini (2001), or Demange (2008).

[12] Many environmental papers employ contests to model lobbying activities; see, for example, Hurley and Shogren (1997), Heayes (1997), or, more recently, Kotchen and Salant (2011) and the references therein. The literature on environmental contest modelling of abatement incentives is, nevertheless, small. The only paper that we are aware of Dijkstra (2007). He is interested in the time (in)consistency of environmental policy under imperfect government commitment and is not concerned with implementing efficiency.

[13] Our main results apply to the simple symmetric setting. Subsection 3.2 generalises the model to the asymmetric case; its workings are illustrated in several examples in subsection 3.2 and appendix A.1.

[14] We view the ranking introduced below as generated by some automaton or monitoring device (see also footnote 17). While agreement on this machine and verifiability of its readings are indispensable, the underlying efforts themselves need neither be observable nor verifiable. If we were to add a zero-mean noise term to output (without changing anything else) productive efforts could not be deduced from output either.

abatement efforts—of which player i suffers a known share s_i.[15] Emissions are seen as an externality, a by-product (or factor) of production.[16] We normalise $\sum_h s_h = 1$ which introduces a public bad team problem and summarise a players utility in the absence of any incentive mechanisms as:

$$y(e_i) - s_i m \left(\sum_{j \in \mathcal{N}} (e_j - f_j) \right) - c_e(e_i) - c_f(f_i). \tag{1}$$

As means to alleviate this problem we introduce an incentive system based on a ranking of individual reductive efforts and award the top-ranked players prizes. The total prize pool P, from which these prizes are taken is defined as the sum of fraction $(1-\alpha)$ of each participant's individual income or output $y(e_i)$, i.e., $P = \sum_i (1-\alpha) y(e_i)$. Thus, the incentive mechanism redistributes income and its budget balances by definition. The incentive mechanism awards $\beta^1 P$ to the winner, $\beta^2 P$ to the player coming second, and so on, with $\sum_h \beta^h = 1$.

We assume that some noisy (partial) ranking of the players' reductive efforts is observable and verifiable. We interpret this ranking technology as arising from the agreement's monitoring of mutual abatement efforts. It is part of the mechanism the players need to agree on and gives player i's probability $p_i^h(\mathbf{f})$ of being awarded prize h as a function of the imperfectly monitored reductive efforts of all participants. We assume that $p_i(\mathbf{f}) = (p_i^1(\mathbf{f}), \dots, p_i^n(\mathbf{f}))$ is strictly increasing in f_i, strictly decreasing in all other arguments, equal to $^1/_n$ for identical arguments, twice continuously differentiable, and zero for $f_i = 0$ with at least one $f_{j \neq i} > 0$, $j \in \mathcal{N}$.[17] Ties are broken randomly.

We use the above introduced interpretation of $p_i^h(\mathbf{f})$ as a probability of winning the full prize because this is the standard reading used in contest models. This is entirely equivalent, however, to the interpretation of $p_i^h(\mathbf{f})$ as the *share* of the tax pool P that is allocated to player i in dependence of all players abatement efforts. Since this interpretation does not require the transfer of exorbitantly large payments to a 'winning' player, it might well be the more practically fruitful way of interpreting our model. Formally, as pointed out above, the two interpretations are equivalent.

Given a ranking $p(\mathbf{f}) = (p_1(\mathbf{f}), \dots, p_n(\mathbf{f}))$, a (subgame perfect) equilibrium in this contest game consists of two elements: an identity independent pair (α, β) specifying the tax and tournament prizes and a pair of efforts (e, f) determining output and winning probabilities. Since we are implementing

[15] Requiring non-negative differences in the damage function $m(\cdot)$ ensures that reductive efforts cannot substitute productive efforts. Since this requirement is fulfilled for most of our analysis, we redefine $m := m(\max\{0, \sum_h e_h - \sum_h f_h\})$ and only make the non-negative argument explicit when necessary. (The implied non-differentiabilities play no role in our model.)

[16] In general, we can think of $c_f(f_i)$ as the cost of moving from the status quo to the targeted level of greenhouse gas (GHG) emissions. Stern (2006) argues similarly that the cost of stabilising GHG emission will be at about 1% of GDP per year compared to Business-As-Usual: abatement is seen as a cost to the productive process. In the non-separable case, productive abatement has some concave benefit $\tilde{y}(f)$ which we omit from our model. But ignoring this additional benefit makes our problem harder to solve and therefore is just a simplification.

[17] Since this contest success function is general, the reductive efforts determining the contest outcome can be easily normalised with respect to, for instance, the individual (perceived) emission consumption share s_i. As usual, this ranking technology can be interpreted as monitoring technology, i.e., the slope of the function can be determined, e.g., by the frequency of inspections or the design of surveillance equipment. From a design point of view, the underlying assumption is that higher monitoring precision comes at a higher cost; infinite precision is not attainable. The inclusion of some monitoring cost ω financed out of the prize pool which is then split βP, $1 - \beta - \omega$, ω) is straightforward and does not qualitatively change any of our results.

efficiency we are looking for a symmetric equilibrium in pure strategies.[18]

2.1 Timing and participation

This subsection defines a simple, symmetric, and simultaneous agreement formation game in order to derive the efficient mechanism in a simple setting. Note that a more general participation game supporting gradual agreement formation is defined in section 4 on the basis of a mutual insurance idea resulting from stochastic output.

Since the players' expected payoffs are symmetric in the basic model, we can think of a simple proposal game in which the identity independent design parameters $\langle \alpha, \beta; p(\mathbf{f}) \rangle$ are proposed by one randomly chosen player and the game is played if and only if all others simultaneously agree to the proposed parameters. The equilibrium concept used in such a game is subgame-perfect equilibrium which our solution satisfies. In order to discourage strategic disagreement, our design is slightly more involved: we propose a two-stage mechanism at the first stage of which an arbitrary player (called player 1) is randomly chosen to propose the two balanced budget contracts $C = \langle \alpha, \beta; p(\mathbf{f}) \rangle$ and $C' = \langle \alpha', \beta'; p(\mathbf{f})' \rangle$. The first contract C is invoked if all players agree to participate in the agreement. It implements efficient efforts in subgame perfect equilibrium. The second contract C' is invoked by the agreeing players if at least one player fails to participate and implements inefficient efforts which successfully deter non-agreement.[19]

More precisely, at the first stage of the game, if all players accept C, then the contest specified by C is set up, players commit their output shares $(1 - \alpha)y(e_i)$ and the game proceeds to the next stage. If at least one player rejects C, the agreeing players form a residual agreement, implement C' and again proceed to the second stage. If less than two players agree to setting up the mechanism (C, C'), then the game ends and each player obtains their individual utility without agreement. At the second stage, conditional on the formation of an agreement, players choose their efforts simultaneously to maximise own expected utilities. The noisy ranking of reductive efforts specifies a winner, second, etc, final output realises, and the prize pool is redistributed to the winner, second, etc, according to the contract specified by C, or C', respectively.

Non-participation in the agreement can be discouraged by either the simple simultaneous agreement game described above, or the threat of agreement members to implement C'. It is easy to see that such a sufficiently strong punishment contract C' always exists: setting $C' = \langle \alpha' = 1, \cdot; \cdot \rangle$ replicates the pre-agreement scenario in which all players are worse off than with an agreement.[20] This extreme form of punishment, however, will typically not be necessary. As illustrated in subsection 3.3, a second-best contract C' will generally be able to implement higher levels of abatement

[18] The efficient allocation is symmetric because of the assumed concavity of production and cost convexity. Especially in the more complicated model variants discussed in the extensions, there may well be other (mixed) equilibria, perhaps of an asymmetric nature, which we disregard for the present analysis. The reason is that they can never implement the efficient allocation.

[19] Formally, this second problem is equivalent to allowing a signatory to exit the agreement (i.e., renege on his commitments) after the agreement is formed. As pointed out by Chander and Tulkens (2009), this contract will typically not be renegotiation proof and commitment to C' is crucial. We discuss other enforcement measures in appendix A.2 and a general agreement model based on stochastic output in section 4.

[20] For a detailed study of how punishments can be used to force agreement see Chander and Tulkens (1995).

than those materialising absent an IEA.[21]

2.2 First-best benchmark

Much of the economics behind our results can be understood from the simple symmetric two players case on which the main body of the paper rests. For this two-player setup, we label players as i, j with $i = 1, 2$ and $j = 3 - i$. We define the efficient levels of both productive and reductive efforts (e^*, f^*) as those maximising social welfare absent of incentive aspects

$$\begin{array}{rcl} \underset{(e,f)}{max}\ u(\mathbf{e},\mathbf{f}) & = & 2y(e) - m\,(2e-2f) - 2c_e(e) - 2c_f(f) \\ \frac{\partial u}{\partial e}, \frac{\partial u}{\partial f} & \Leftrightarrow & \begin{cases} y'(e^*) = m'(2e^*-2f^*) + c_e'(e^*), \\ m'(2e^*-2f^*) = c_f'(f^*). \end{cases} \end{array} \tag{2}$$

(The expressions following the curly bracket are the necessary first-order conditions for optimality resulting from the concavity/convexity assumptions made.) In the absence of an incentive scheme, a player $i = 1, 2$ individually maximises

$$\begin{array}{rcl} \underset{(e_i,f_i)}{max}\ u_i(e_i, f_i) & = & y(e_i) - s_i m(e_i + e_j - f_i - f_j) - c_e(e_i) - c_f(f_i) \\ \frac{\partial u}{\partial e_i}, \frac{\partial u}{\partial f_i} & \Leftrightarrow & \begin{cases} y'(e) = s_i m'(2e-2f) + c_e'(e), \\ s_i m'(2e-2f) = c_f'(f). \end{cases} \end{array} \tag{3}$$

in which s_i is player i's local share of global emissions. We write $e = e_i = e_j$, $f = f_i = f_j$ after maximisation. Since we normalise $s_i + s_j = 1$, the individual first-order conditions in (3) cannot both equal those in (2). In order to overcome this inefficiency in *both* dimensions, we introduce an endogenised rank-order emissions reduction reward scheme, i.e., a contest. We ask each participating nation to commit to contributing a share $(1-\alpha)$ of their individual output $y(e_i)$ to the mechanism and therefore form a pool of prize money of size $P = (1-\alpha)\,(y(e_i) + y(e_j))$. In a contest specifying player i's winning probability as $p_i(\mathbf{f})$ based on both players' reductive efforts, we want to assign βP to the winner and $(1-\beta)P$ to the player coming second. Notice that such a mechanism redistributes income. The individual problem under our incentive mechanism is therefore[22]

$$\underset{(e_i,f_i)}{max}\quad \underbrace{\alpha y(e_i)}_{\text{retained output}} + \underbrace{p_i(\mathbf{f})\beta P}_{\text{first prize}} + \underbrace{(1-p_i(\mathbf{f}))(1-\beta)P}_{\text{second prize}} - \underbrace{s_i m(e_i + e_j - f_i - f_j)}_{\text{damage from emissions}} - \underbrace{(c_e(e_i) + c_f(f_i))}_{\text{effort costs}}.$$

We define individual rationality as the requirement that the utility from participating in this mechanism for appropriately chosen $\langle \alpha, \beta; p(\mathbf{f}) \rangle$ exceeds $i)$ the utility from non-formation of the agreement (3), $ii)$ of free-riding on the others' reductive efforts *within* the agreement and $iii)$ on free-riding on the others' reductive efforts *outside* the agreement. In the first case, no agreement exists at all while in the third case, an agreement outsider benefits from the reductive efforts of the agreement members. The second case concerns an agreement member's inefficient effort provision with committed output share.

[21] Designing C' just sufficiently bad to serve as a deterrent resembles the idea of γ-core stability in Chander (2007). An alternative way of deterring this kind of free-riding on the agreement is to grant most favoured 'green' trading terms only to participating nations. This idea is further explored in appendix A.2. For a simulation of agreement stability using plausible data based on an integrated assessment model see Bosetti et al. (2012).

[22] In the two-players setting, note that $p_i(\mathbf{f}) = 1 - p_j(\mathbf{f})$, $\beta^1 = \beta$ and $\beta^2 = 1 - \beta$.

2.3 The main example

In this subsection we illustrate our main efficiency result through example. In particular, we show in a simple setup that it is possible to reach the efficient allocation among symmetric players who consent to the agreement parameters. We continue to build on this example throughout the paper as we introduce further results.

Example 1: We use a simple, symmetric example with quadratic costs and square root production function to demonstrate the basic idea of our mechanism.[23] In this setup, a benevolent planner maximising the sum of social utility net of total cost (2) maximises the objective

$$\underset{(e,f)}{max}\ 2e^{1/2} - (2e-2f)^2 - 2(e^2+f^2) \ \Leftrightarrow \begin{cases} e^* \approx 0.2823, \\ f^* \approx 0.1882. \end{cases} \tag{4}$$

The corresponding individual problem (in the absence of an incentive mechanism) leads to inefficient provision of efforts because

$$\underset{(e_i,f_i)}{max}\ e_i^{1/2} - s_i(e_i+e_j-f_i-f_j)^2 - (e_i^2+f_i^2) \quad \Leftrightarrow \begin{cases} e \approx 0.3029 > e^*, \\ f \approx 0.1514 < f^*, \end{cases} \tag{5}$$

for symmetric damage shares $s_1 = s_2 = 1/2$. Notice that, with respect to the efficient efforts, the combined externality and free-riding inherent in the problem imply that players both produce too much and abate too little.

For our incentive agreement we assume in the present example that the probability of winning the reduction award is given by the (generalised) Tullock success function specifying a player's probability of winning as a function of that player's effort over the total sum of efforts.[24] The prize pool which we collect for incentive purposes is $P = (1-\alpha)(e_i^{1/2} + e_j^{1/2})$. Then, an individual's problem under the incentive scheme is

$$\underset{(e_i,f_i)}{max}\ \alpha e_i^{\frac{1}{2}} + \frac{f_i^r}{f_i^r+f_j^r}\beta P + \frac{f_j^r}{f_i^r+f_j^r}(1-\beta)P - s_i(e_i+e_j-f_i-f_j)^2 - (e_i^2+f_i^2) \tag{6}$$

for some exponent $r > 0$ specifying the precision with which the ranking selects the highest reduction effort nation among the set of competitors.[25] We interpret this exponent as the accuracy with which the agreement monitors the emissions reduction efforts of its members. Whenever we consider the Tullock example case in the following, we write the corresponding contract as $\langle \alpha, \beta; r \rangle$ instead of the general ranking based contract $\langle \alpha, \beta; p(\mathbf{f}) \rangle$.

[23] We will return to this example setup throughout the paper to illustrate further ideas and extensions.

[24] Under a Tullock contest success function, the contestant with the highest effort does not necessarily win the prize. Hence, the resulting ranking has occasionally been referred to as 'non-fully discriminatory,' 'non-deterministic,' 'noisy,' or 'fuzzy.' Our interpretation is that the ranking is inexact in the sense that the monitoring technology it is based on is not perfect. The Tullock (or Logit) form has been axiomatised by Skaperdas (1996) and follows naturally from micro-foundations à la Fu and Lu (2012) or Jia (2008).

[25] The particular monitoring technology is not very important as we generalise over the set of applicable success functions in appendix A.3. What is important is that the success function incorporates enough randomness in its outcome. If the ranking is too precise (as is the case with the all-pay auction—which can be viewed as the $r = \infty$ limit-case of the Tullock function) then equilibria in pure strategies typically fail to exist. This would be problematic as our contest strives to implement the efficient pure effort choices.

Upon maximisation, this gives the two simultaneous first-order conditions

$$16e_i = 8f_i + \frac{1+\alpha}{\sqrt{e_i}},\ 2e_i = 4f_i + \frac{\sqrt{e_i}r(\alpha-1)(2\beta-1)}{2f_i}. \tag{7}$$

We again consider the simplest case in which symmetric nations are identical (as we did before for the planner) and set $e = e_1 = e_2$, $f = f_1 = f_2$, with $s_i = {}^1/_2$. We then force the resulting efforts in line with the efficient efforts by imposing $e = e^*$ and $f = f^*$ from (4) and solve (7) for the efficiency inducing design parameters $\langle \alpha, \beta; r \rangle$

$$\alpha^* = \frac{3}{5},\ \beta^* = \frac{1}{2} + \frac{1}{6r}. \tag{8}$$

As β^* depends on the precision of the monitoring technology r, the rewards scheme—and in particular the relative size of the prizes paid to the winner and loser given by β—can be designed as seen fit and compatible with the chosen monitoring technology.[26] The mechanism satisfies $\beta \in [\frac{1}{2}, 1]$ if $r \geq {}^1/_3$, implying that the losing nation needs never pay more than the committed share $1 - \alpha$. Figure 1

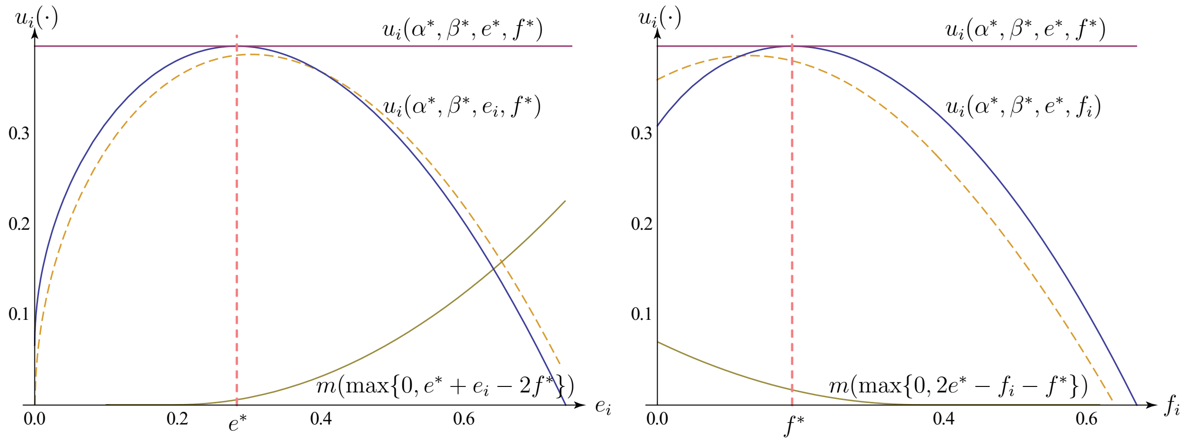


Figure 1: The top, horizontal line is the equilibrium utility from $(\alpha^*, \beta^*, e^*, f^*)$ implementing efficient efforts e^* and f^*. The curves below show the utility from unilaterally deviating in either effort dimension. Notice the positive utility from free-riding at zero efforts. The dashed curves give the (outside) utility from no agreement formation exhibiting both overproduction in e_i and underprovision of abatement f_i relative to the socially efficient levels.

shows that participating in the contest gives higher utility than staying out and free-riding on the other's effort. It confirms $(\alpha^*, \beta^*, e^*, f^*)$ as equilibrium in pure strategies with full participation (on an appropriately chosen plot-range outside of which utility is negative).◁

The economics of this example is simple: An increase in productive efforts e_i causes individual output $y(e_i)$—and, hence, the prize pool P—as well as global pollution $m(\sum_h e_h - \sum_h f_h)$ to rise. Of these, the player retains shares α and s_i, respectively. An increase in reductive efforts f_i enlarges the player's chance to win the prize share β in the reduction contest (while decreasing the competitors' chances) and simultaneously decreases global pollution. Trading off α against β allows us to fine-tune efforts to their efficient levels.

[26] There are well-known existence issues with symmetric pure-strategy equilibrium with $r > 2$ in rent-seeking contests (see, e.g., Schweinzer and Segev (2012)) but, as shown in proposition 2, these do not apply with the same severity to our problem where costs are convex and the prize pool is endogenous.

Example 2: To get a feeling for the magnitudes implied by our example mechanism we plug in the 2011 global GDP of \$80tr, or, among two identical players, \$40tr GDP per player. Thus, our proposed mechanism collects $P = (2/5)80 = \$24.4$tr or \$16tr from each player. The following table lists the redistribution implied by the efficient mechanism. Depending on the precision of monitoring r, the mechanism parameters and transfers are

r	β	$1-\beta$	1st	2nd	transfer	%
1	$^2/_3$	$^1/_3$	\$21.3tr	\$10.7tr	±\$5.3tr	±13.3%
2	$^7/_{12}$	$^5/_{12}$	\$18.7tr	\$13.2tr	±\$2.7tr	±6.7%
.	...		...		...	
5	$^8/_{15}$	$^7/_{15}$	\$17.0tr	\$15.0tr	±\$1.0tr	±2.7%
.	...		...		...	
11	$^{17}/_{33}$	$^{16}/_{33}$	\$16.5tr	\$15.5tr	±\$0.5tr	±1.2%

where the rightmost column's percentages are taken from the total symmetric \$40tr GDP. Our sufficient existence condition derived in proposition 2 guarantees existence up to and including $r = 5$. Actual existence, however, is only lost for $r > 11$. At this monitoring level, a transfer of 1.2% of GDP is sufficient to implement both efficient abatement and efficient production.◁

As pointed out in the model section above, we can alternatively employ the interpretation of $p_i(\mathbf{f})$ as the endogenous share of the tax pool that is allocated to player i in dependence of all players abatement efforts. Under this interpretation, there are no 'winning' or 'losing' players and each player gets the share specified by $p_i(\mathbf{f})$ bounded from above by βP and from below by $(1-\beta)P$. (There is no need for more complicated multi-prize configurations.) Under this interpretation, while the incentives incorporated in the mechanism ensure efficient efforts along both dimensions, all symmetric equilibrium transfers cancel out.

3 The deterministic output model

3.1 Equilibrium characterisation and existence

We begin by characterising the design parameters which induce efficiency in our model. Recall that, under the contest scheme, an individual participant $i = 1, 2$ chooses a pair of efforts (e_i, f_i) to

$$\max_{(e_i,f_i)} \alpha y(e_i) + p(\mathbf{f})\beta P + (1-p(\mathbf{f}))(1-\beta)P - s_i m(e_i + e_j - f_i - f_j) - c_e(e_i) - c_f(f_i) \quad (9)$$

in which $p(\mathbf{f})$ is the probability of coming first in a ranking of reductive efforts f and the prize pool is $P = (1-\alpha)(y(e_1) + y(e_2))$. We require that $y' > 0$, $y'' < 0$, $m' > 0$, $m'' > 0$, and $c'_{1,2} > 0$, $c''_{1,2} > 0$. We moreover assume that $m(\cdot)$ only depends on the difference of total productive minus reductive efforts. Taking derivatives with respect to both effort types, we obtain the simultaneous pair of first-order conditions defining individually optimal efforts (e_i, f_i) as

$$\begin{aligned} c'_e(e_i) + s_i m'(e_i + e_j - f_i - f_j) &= (1 - \beta + \alpha\beta + (1-\alpha)(2\beta - 1)p(f_i, f_j))y'(e_i), \\ c'_f(f_i) + (\alpha - 1)(2\beta - 1)(y(e_i) + y(e_j))p'(f_i, f_j) &= s_i m'(e_i + e_j - f_i - f_j). \end{aligned} \quad (10)$$

Assuming tentatively that a symmetric equilibrium $e = e_i = e_j$, $f = f_i = f_j$, $s_i = {}^1\!/_2$ exists (until existence is demonstrated in proposition 2) this simplifies to

$$\begin{aligned} 2c_e'(e) + m'(2e-2f) &= (\alpha+1)y'(e), \\ 2c_f'(f) - m'(2e-2f) &= 4(1-\alpha)(2\beta-1)p'(f,f)y(e). \end{aligned} \tag{11}$$

Equating these efforts to the efficient efforts e^*, f^* resulting from the solution to the social planner's problem in (2), we obtain

$$4p'(\mathbf{f}^*)(2\beta-1) \;=\frac{y'(e^*)}{y(e^*)} \Leftrightarrow \begin{cases} c_e'(e^*) = \alpha y'(e^*), \\ c_f'(f^*) = 4(1-\alpha)(2\beta-1)p'(\mathbf{f}^*)y(e^*) \end{cases} \tag{12}$$

in which $\mathbf{f}^* = (f^*, f^*)$. We know from (2) that there exists an $\alpha \in [0,1]$ to satisfy the first equation. Substituting this α into the second equation determines $\beta \in [{}^1\!/_2, 1]$ for a suitably chosen ranking $p(\cdot)$. Without further restrictions on the design parameters—and in particular the slope of the ranking technology $p(\cdot)$ in equilibrium—the set of necessary conditions in (12) can always be satisfied. Taking equilibrium existence as given (until we verify it in proposition 2), the following proposition establishes the precise criteria on the parameters for both productive and reductive efficiency to obtain simultaneously for any number of players $n \geq 2$. In all following results we employ the simple prize structure $\beta = \left(\beta^1, \frac{1-\beta^1}{n-1}, \ldots, \frac{1-\beta^1}{n-1}\right)$ assigning a single winning prize and another prize to all losers. This is not necessary but simplifies the exposition considerably.

Proposition 1. *For appropriately chosen $\langle \alpha, \beta; p(\mathbf{f}) \rangle$ and $P = (1-\alpha)\sum_j y(e_j)$, player $i \in \mathcal{N}$ chooses efficient productive as well as reductive efforts (e^*, f^*) in*

$$\max_{(e_i, f_i)} \; \alpha y(e_i) + \sum_h \left(\beta^h p_i^h(\mathbf{f})P\right) - s_i m\Big(\sum_h (e_h - f_h)\Big) - c_e(e_i) - c_f(f_i). \tag{13}$$

The proof can be found in appendix B and follows the same intuition as described in the example. It is straightforward to show that proposing $C = \langle \alpha, \beta; p(\mathbf{f}) \rangle$ at the first stage of the game maximises player 1's expected utility, given that players' efforts are functions of the proposed mechanism $(e(\alpha, \beta, p), f(\alpha, \beta, p))$. This is unsurprising as payoffs are symmetric and what maximises welfare must also maximise the proposer's utility.[27] This implies immediately that a simple symmetric model where every player decides simultaneously on whether to set up an agreement or not results in both unanimous agreement and full participation in symmetric, subgame perfect equilibrium.

A consequence of this first result is that full efficiency in the symmetric n-player model can be obtained with just two different prizes: one for the winner and another for everyone else. As one only needs to check for a winning 'abatement-champion,' such a scheme is easy to monitor. Since the general objective (13) is not necessarily well-behaved without further assumptions on $p(\cdot)$, we proceed to show that equilibria exist for the subclass of problems governed by the Tullock success function $p_i(f) = f_i^r/(f_1^r + \cdots + f_n^r)$. Hence, for the following proposition we concern ourselves with mechanisms of the form $\langle \alpha, \beta; r \rangle$, where the designer chooses the parameter r (interpreted as monitoring intensity) as seen fit. Depending on this exponent r, the Tullock function may be first convex and then concave. So, again, the underlying optimisation problem is non-concave.

[27] The design of the proposal stage is less straightforward in the asymmetric case.

Proposition 2. *Consider a mechanism $\langle \alpha, \beta; r\rangle$. Under the Tullock success function, if c_f is sufficiently convex, a symmetric pure-strategy equilibrium exists in which production and abatement are efficient.*

The proof of this proposition establishes a sufficient condition for quasi-concavity of the players' objective. In particular, the sufficient threshold (68), given in appendix B, ensures the existence of a symmetric pure-strategy equilibrium for contests $\langle \alpha, \beta; r\rangle$ governed by the Tullock success function by specifying an upper bound on admissible r. If this condition is respected, an equilibrium which implements the efficient efforts characterised in proposition 1 is certain to exist. Since r can be chosen by the designer, this condition can in principle always be satisfied. If, however, for some chosen environment, the effort cost of abatement are insufficiently convex or, equivalently, the chosen monitoring precision r (or the equivalent slope of $p(\mathbf{f})$) is too high, then pure-strategy equilibrium fails to exist. In that case, giving up the simple 'flat-loser' prize structure $\beta = \left(\beta^1, \frac{1-\beta^1}{n-1}, \dots, \frac{1-\beta^1}{n-1}\right)$ in favour of a structure which awards multiple first prizes $\beta' = (\beta^1 = \beta^2 \geq \dots \geq \beta^n)$ eases the existence problem at the expense of the implemented abatement efforts.[28]

3.2 Asymmetries

In order to show that our efficiency result is not an artifact of our symmetry assumptions, this subsection explores cost asymmetries among players. The main argument is presented for any number of players $n \geq 2$ and identity dependent shares (α_i, β_i). An illustrative example follows and further discussions of the asymmetric case can be found in appendix A.1.

Let $i \in \mathcal{N}$ and $n \geq 2$. We illustrate that, for appropriately chosen $\langle \alpha_i, \beta_i; p(\mathbf{f})\rangle_i^n$, prize pool $P = \sum_{j=1}^n (1-\alpha_j) y_j(e_j)$, and prize structure $\left(\beta_i, \frac{1-\beta_i}{n-1}, \dots, \frac{1-\beta_i}{n-1}\right)$, efficient solutions exist to player i's asymmetric problem

$$\underset{(e_i, f_i)}{max}\ \alpha_i y_i(e_i) + p_i^1(\mathbf{f})\beta_i P + \sum_{i \neq j} p_j^1(\mathbf{f}) \left(\frac{1-\beta_j}{n-1}\right) P - s_i m \left(\sum_{i=1}^n e_i - f_i\right) - c_{i,e}(e_i) - c_{i,f}(f_i). \quad (14)$$

Analogous to (2), let player i's asymmetric efficient efforts be given by

$$y_i'(e_i^*) = m'(G) + c_{e_i}'(e_i^*) \text{ and } m'(G) = c_{f_i}'(f_i^*) \quad (15)$$

in which $G = \sum_{j=1}^n (e_j - f_j)$. Let the payment shares α_i and winning shares be identity-dependent, i.e., a winning player i gets share β_i and a winning j gets share β_j of the total prize pool $P = \sum_{j=1}^n (1-\alpha_j) y_j(e_j)$. Thus, taking all e_j^*, f_j^*, $j \neq i$, as given, player i maximises (14). Taking derivatives with respect to e_i, f_i and inserting (15), determines player i's best response through[29]

$$\alpha_i = \frac{y_i'(e_i)(1-H) - (1-s_i)m'(G)}{y_i'(e_i)(1-H)}, \text{ where } H = \beta_i p_i^1(\mathbf{f}) + \frac{\sum_{j\neq i}(1-\beta_j)p_j^1(\mathbf{f})}{(n-1)},$$
$$\beta_i = \frac{(n-1)\left((1-s_i)m'(G)\right) - \sum_{j\neq i}(1-\beta_j)p_{j(f_i)}^1(\mathbf{f})P}{p_{i(f_i)}^1(\mathbf{f})(n-1)P} \quad (16)$$

[28] It is easy to see why this is the case: an equilibrium where every player gets the same prize must exist (with zero abatement efforts). By continuity, equilibria (with small reductive efforts) will exist under a prize structure which gives the same prize to every player except the one coming last. For details see Schweinzer and Segev (2012).
[29] The expressions (16) can be simplified further but then get excessively lengthy.

and $p^1_{i(f_i)}(\mathbf{f})$ denotes $\frac{\partial}{\partial f_i} p^1_i(\mathbf{f})$. (16) corresponds to (62) and elicits asymmetric efficient efforts (e^*_i, f^*_i). Without putting any restrictions on the numbers α_i and β_i, a (numerical) solution to (16) can always be found. Since the same is true for the best responses of player i's opponents, we confirm that a solution to the complete system $(\alpha_1, \ldots, \alpha_n; \beta_1, \ldots, \beta_n)$ exists.[30]

In the introduction, we stress that diluted objectives among development nations may be problematic for an IEA. In order to address this problem, our asymmetric extension models developing countries as exhibiting a smaller cost of emissions reduction than their more highly developed counterparts. In our model, this increases a developing country's chance of winning the reduction contest. Similarly, in equilibrium, developed countries feature higher productive efforts and thus contribute a bigger share of the tournament prize pool than less productive countries.[31]

Example 3: As an illustration of this asymmetric modelling idea we extend the example from section 2.3 to three players and parameterise player $i = 1, 2, 3$'s cost with respect to productive and reductive efforts by the pair of scalars (γ_i, δ_i).[32] Following the general strategy outlined above, we use identity dependent shares (α_i, β_i) and prize structures $\left(\beta_i, \frac{1-\beta_i}{n-1}, \ldots, \frac{1-\beta_i}{n-1}\right)$. The planner's objective is

$$\max_{(e_1,f_1,e_2,f_2,e_3,f_3)} \sum_i e_i^{1/2} - \left(\sum_i (e_i - f_i)\right)^2 - \sum_i \gamma_i e_i^2 - \sum_i \delta_i f_i^2. \tag{17}$$

For total prize pool $P = \sum_i (1-\alpha_i) e_i^{1/2}$, individuals $i = 1, 2, 3$ choose (e_i, f_i) to maximise

$$\begin{aligned}
&\alpha_1 e_1^{\frac{1}{2}} + \tfrac{f_1^r}{\sum_h f_h^r}\beta_1 P + \left(1 - \tfrac{f_1^r}{\sum_h f_h^r}\right)\left(\tfrac{f_2^r}{f_2^r+f_3^r}\tfrac{1-\beta_2}{2} + \tfrac{f_3^r}{f_2^r+f_3^r}\tfrac{1-\beta_3}{2}\right) P - s_1 \left(\textstyle\sum_h (e_h - f_h)\right)^2 - \gamma_1 e_1^2 - \delta_1 f_1^2,\\
&\alpha_2 e_2^{\frac{1}{2}} + \tfrac{f_2^r}{\sum_h f_h^r}\beta_2 P + \left(1 - \tfrac{f_2^r}{\sum_h f_h^r}\right)\left(\tfrac{f_1^r}{f_1^r+f_3^r}\tfrac{1-\beta_1}{2} + \tfrac{f_3^r}{f_1^r+f_3^r}\tfrac{1-\beta_3}{2}\right) P - s_2 \left(\textstyle\sum_h (e_h - f_h)\right)^2 - \gamma_2 e_2^2 - \delta_2 f_2^2,\\
&\alpha_3 e_3^{\frac{1}{2}} + \tfrac{f_3^r}{\sum_h f_h^r}\beta_3 P + \left(1 - \tfrac{f_3^r}{\sum_h f_h^r}\right)\left(\tfrac{f_1^r}{f_1^r+f_2^r}\tfrac{1-\beta_1}{2} + \tfrac{f_2^r}{f_1^r+f_2^r}\tfrac{1-\beta_2}{2}\right) P - s_3 \left(\textstyle\sum_h (e_h - f_h)\right)^2 - \gamma_3 e_3^2 - \delta_3 f_3^2.
\end{aligned}$$

The result for $(\gamma_1 = 1, \delta_1 = 1, \gamma_2 = {}^2/_3, \delta_2 = {}^3/_4, \gamma_3 = {}^1/_3, \delta_3 = {}^1/_2)$ and $s_i = {}^1/_3$ leads to the efficiency inducing asymmetric shares of[33]

$$\begin{aligned}
&\alpha_1 = 0.619, \beta_1 = 0.511, \alpha_2 = 0.553, \beta_2 = 0.550, \alpha_3 = 0.313, \beta_3 = 0.712,\\
&e^*_1 = 0.274, f^*_1 = 0.204, e^*_2 = 0.339, f^*_2 = 0.272, e^*_3 = 0.476, f^*_3 = 0.408. \triangleleft
\end{aligned}$$

We now extend the above basic asymmetric example to an asymmetric contest based on *relative* reductive efforts per-unit-GDP. This alternative to the standard model incorporates an equally 'fair'

[30] A formal argument is outside the scope of this paper. A proof, however, can be built along similar lines to those of proposition 4 in Giebe and Schweinzer (2013).

[31] Thus, in our asymmetric model, the larger part of emissions reductions are implemented where they are the cheapest, i.e., in emerging economies.

[32] The two players asymmetric case in the example setup is rather special since $\beta_1 + (1 - \beta_2) = 1$ implies $\beta_1 = \beta_2$. Hence, in order to lend itself to a solution, the two players case requires success function slopes of precisely

$$\frac{\partial}{\partial f_i} p_i(\mathbf{f}) = \frac{(1 - s_i) c'_f(f_i)}{(\beta_1 + \beta_2 - 1) P}.$$

Cases with higher numbers of players than two do not imply equal prize shares and do not induce complications.

[33] In this first example, we use countries of same 'size' s_i. This is done to concentrate on the effect of cost asymmetries. The example of appendix A.1 focuses on the effect of size asymmetries, that is, different s_i.

measure of abatement effort for all asymmetric contestants. Our asymmetric abatement contest takes now the form

$$u_i(\mathbf{e},\mathbf{f};\boldsymbol{\varepsilon}) = \overbrace{\alpha_i y(e_i,\varepsilon_i)}^{\text{output}} + \overbrace{p_i^1(\mathbf{e},\mathbf{f})\beta_i P}^{\text{winning}} + \overbrace{\sum_{j\neq i\in\mathcal{N}} p_j^1(\mathbf{e},\mathbf{f}))\frac{1-\beta_j}{n-1}P}^{\text{losing}} - \overbrace{s_i m(\sum_h (e_h - f_h))}^{\text{damage}} - \underbrace{c_i(e_i,f_i)}_{\text{cost}}, \text{ for the redistribution pool } P = \sum_{i=1}^{n}(1-\alpha_i)y(e_i,\varepsilon_i), \tag{18}$$

in which the winning probability $p_i^1(\mathbf{e},\mathbf{f})$ is a two-dimensional version of the generalised Tullock success function. In particular, the winning probability $p_i^1(\mathbf{e},\mathbf{f})$ is now based on the ratio x of the two strategic variables: reductive efforts over a function of productive efforts[34]

$$p_i^1(\mathbf{e},\mathbf{f}) = \frac{x_i^r}{x_1^r + \cdots + x_n^r}, \; x_i = \frac{f_i}{y(e_i)}. \tag{19}$$

The probabilities $p_{j\neq i}^1(\mathbf{e},\mathbf{f})$ are defined in the same way as player j's probability of winning in a contest not involving player i. We again use identity dependent tax rates α_i and winning shares β_i. For simplicity, we only discriminate between winners and losers, i.e., if player i wins, then we award $\beta_i P$ to the winner and $\frac{1-\beta_i}{n-1}P$ to each of the losing players.

The contest success function employed in this subsection is therefore the following multi-dimensional version of the generalised Tullock success function[35]

$$p_i^1(\mathbf{e},\mathbf{f}) = \frac{(f_i/y(e_i))^r}{(f_1/y(e_1))^r + \cdots + (f_n/y(e_n))^r} = \frac{x_i^r}{x_1^r + \cdots + x_n^r}, \; r > 0. \tag{21}$$

The general characterisation of this contest is irrelevant for the purposes of illustrating our variance compression argument and can be obtained from the authors upon request.

Example 4: In our standard square-root/quadratic example environment, efficient efforts (e^*, f^*) are independent of the contest technology and thus given by the asymmetric but otherwise unchanged three-players version of (4) as[36]

$$\begin{array}{l}(\mathbf{e}^*,\mathbf{f}^*) \in \arg \underset{(e,f)}{max}\; u(\mathbf{e},\mathbf{f},\boldsymbol{\varepsilon}) = \\ y(e_1+\varepsilon_1) + y(e_2+\varepsilon_2) + y(e_3+\varepsilon_3) - \\ (e_1+e_2+e_3-f_1-f_2-f_3)^2 - \\ \gamma_1(e_1^2+f_1^2) - \gamma_2(e_2^2+f_2^2) - \gamma_3(e_3^2+f_3^2)\end{array} \Leftrightarrow \begin{cases} e_1^* = 0.513, f_1^* = 0.535, \\ e_2^* = 0.359, f_2^* = 0.267, \\ e_3^* = 0.288, f_3^* = 0.178. \end{cases} \tag{22}$$

[34] For completeness, we define that $p_i(\mathbf{e},\mathbf{f}) = 1$ if $y(e_i) = 0$, $f_i > 0$ and all $y(e_{-i}) > 0$. Similarly, we let $p_i(\mathbf{e},\mathbf{f}) = 1/m$ if $m = |y(e_j) = 0|_{j\in\mathcal{N}}$.

[35] To exclude unbounded ratios x_i, we assume that $p_i^1(\mathbf{e},\mathbf{f}) = 0$ if $e_i = 0$. This discontinuity at zero plays no role in the examples discussed in this setup. The idea can be easily generalised to more than two dimensions. A simple way of achieving this is to use

$$\tilde{x}_i = \frac{f_i}{y_i^1(e_i^1) + \cdots + y_i^m(e_i^m)} \tag{20}$$

in which $e_i^1, \ldots, e_i^m$ is player i's m-dimensional 'normalisation' effort transformed, if necessary, by the functions $y_i^h(\cdot)$, $h = 1, \ldots, m$. To the best of our knowledge, this 'normalised,' relative efforts 'per-unit-GDP' formulation of the Tullock success function is original to this paper.

[36] Since the two-players asymmetric example with $\beta_1 + (1-\beta_2) = 1 \Leftrightarrow \beta_1 = \beta_2$ is still 'too symmetric' to illustrate the workings of the asymmetric mechanism we switch into a three players example in this subsection.

Again for simplicity, we choose linear cost asymmetries $\gamma_3 = 1$, $\gamma_2 = 2/3$, $\gamma_1 = 1/3$ to differentiate players but asymmetric productive capability could be modelled in exactly the same way. Finally, we use relative pollution shares of $s_1 = 4/12$, $s_2 = 3/12$, $s_3 = 5/12$. For simplicity, we only present numerical results in this subsection.

For the three-players asymmetric contest, the redistribution pool is $P = \sum_{i=1}^{3}(1-\alpha_i)(\sqrt{e_i}+\varepsilon_i)$. This setup results in the three players' objectives

$$\begin{aligned}
u_1(\mathbf{e},\mathbf{f};\boldsymbol{\varepsilon}) &= \alpha_1 y(e_1,\varepsilon_1) + p_1^1(\mathbf{e},\mathbf{f})\beta_1 P + (1-p_1^1(\mathbf{e},\mathbf{f}))\left(\tfrac{x_2^r}{x_2^r+x_3^r}\tfrac{1-\beta_2}{2} + \tfrac{x_3^r}{x_2^r+x_3^r}\tfrac{1-\beta_3}{2}\right)P - \\
&\quad s_1(e_1+e_2+e_3-f_1-f_2-f_3)^2 - \gamma_1\left(e_1^2+f_1^2\right), \\
u_2(\mathbf{e},\mathbf{f};\boldsymbol{\varepsilon}) &= \alpha_2 y(e_2,\varepsilon_2) + p_2^1(\mathbf{e},\mathbf{f})\beta_2 P + (1-p_2^1(\mathbf{e},\mathbf{f}))\left(\tfrac{x_1^r}{x_1^r+x_3^r}\tfrac{1-\beta_1}{2} + \tfrac{x_3^r}{x_1^r+x_3^r}\tfrac{1-\beta_3}{2}\right)P - \\
&\quad s_2(e_1+e_2+e_3-f_1-f_2-f_3)^2 - \gamma_2\left(e_2^2+f_2^2\right), \\
u_3(\mathbf{e},\mathbf{f};\boldsymbol{\varepsilon}) &= \alpha_3 y(e_3,\varepsilon_3) + p_3^1(\mathbf{e},\mathbf{f})\beta_3 P + (1-p_3^1(\mathbf{e},\mathbf{f}))\left(\tfrac{x_1^r}{x_1^r+x_2^r}\tfrac{1-\beta_1}{2} + \tfrac{x_2^r}{x_1^r+x_2^r}\tfrac{1-\beta_2}{2}\right)P - \\
&\quad s_3(e_1+e_2+e_3-f_1-f_2-f_3)^2 - \gamma_3\left(e_3^2+f_3^2\right).
\end{aligned} \tag{23}$$

The corresponding six first-order conditions induce the set of efficient efforts (22) as individual global maxima for the following set of design parameters $\langle \alpha, \beta, r^* = 2\rangle$[37]

$$\begin{aligned}
&\alpha_1^* = 0.610, \quad \alpha_2^* = 0.740, \quad \alpha_3^* = 0.810, \\
&\beta_1^* = 0.767, \quad \beta_2^* = 0.518, \quad \beta_3^* = 0.460.
\end{aligned} \tag{25}$$

The below figure confirms that these identify an equilibrium for the game where a ratio of asymmetric strategic variables is chosen.◁

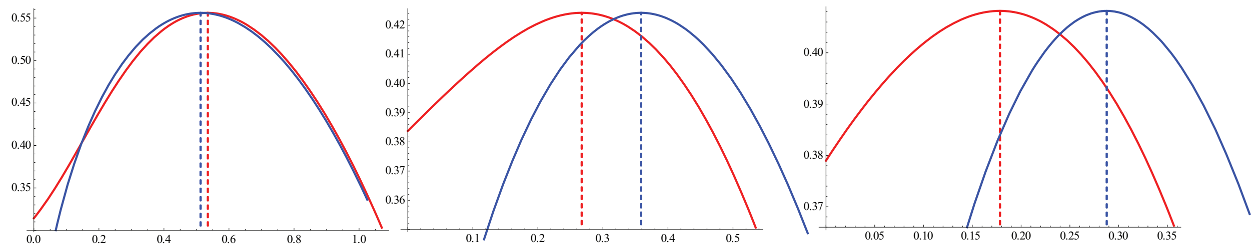


Figure 2: Efficient effort pairs as global maximisers of (23) for player 1's (left), player 2's (centre) and player 3's unilateral deviation utility (right). Red are abatement efforts, blue productive efforts. (The ranges shown contain all maxima.)

Although we demonstrate most of our results in a simplified symmetric setup, the above two examples suggest that at least some of the attractive properties of the mechanism we introduce carry over to the asymmetric case. Finally, the question whether and in how far the efficient asymmetric abatement levels can be used directly as inputs into the contest success function or not is, of course,

[37] In the setup of subsection 4.2, the efficiency inducing parameter set of the symmetric version of this 'normalised' contest based on $x_i = f_i/y(e_i)$ is given by

$$\alpha^* = \frac{11}{15},\ \beta^* = \frac{1}{2} + \frac{1}{4r}. \tag{24}$$

This is more favourable to the participants than the corresponding parameters under the standard contest (8).

politically charged. We therefore restrict ourselves to pointing out that any standard normalisation of input efforts is feasible. For instance, it is perfectly possible to normalise inputs such that each country which exerts it's efficient abatement level has the same equilibrium chance of winning the first prize $^1/_n$. An interesting implementation of this normalisation idea is to assign a vector of individual weights $\psi = (\psi_1, \ldots, \psi_n)$, with each $\psi_i > 0$, to the heterogenous contestants' reductive efforts turning the basic success function into

$$p_i(f, \psi) \ = \frac{\psi_i f_i^r}{\sum_j \psi_j f_j^r}, \text{ where } \frac{\psi_i}{\delta_i^r} = \frac{\psi_j}{\delta_j^r} \ \forall j \neq i \tag{26}$$

and thus 'levelling the playing field.'[38] Following the assignment of ψ, one can proceed with the analysis of section 3 without further change. A strength of a contest-based mechanism is thus its ability to adopt different abatement-cost distribution rules. Our concept of reductive effort can easily incorporate fairness considerations by encompassing, for instance, population, geographic size, or GDP. The latter is used in our asymmetric example 4.

3.3 Agreement formation

In the symmetric and simultaneous case, the argument for full participation is almost trivial since every player's utility is identical and there is no improvement for any player if the agreement is not unanimously formed. Equilibrium existence then implies that free-riding on the reductive effort is not attractive once a nation is committed to the agreement.[39] However, as the number of participants in the mechanism n goes up, the utility from free-riding on an existing agreement increases as the disutility from pollution $m(\sum_h (e_h - f_h))$ approaches the efficient level. Hence, if gradual agreement formation is allowed, then the only leverage left in the efficient contract C is the contest on the pre-committed output share of $(1-\alpha)$—which is generally not sufficient to deter free-riding once an agreement is in place. The alternative contract C' is, however, capable of eradicating all gains from free-riding on the agreement by—in its most extreme form—replicating non-agreement pollution levels.

Proposition 3. *Participation in the symmetric mechanism specifying the pair of contracts* $C = \langle \alpha^*, \beta^*; p^*(\mathbf{f}) \rangle$ *determined through* (13) *and* $C' = \langle \alpha' = 1, \beta' = {}^1/_2; \cdot \rangle$ *is individually rational in the sense that the utility from free-riding efforts* e^s, f^s *on* C' *cannot exceed the utility obtained when agreeing to* C

$$y(e_i^s) - s_i m(e_i^s + (n-1)e'(\alpha', \beta', \cdot) - f_i^s - (n-1)f'(\alpha', \beta', \cdot)) - c_e(e_i^s) - c_f(f_i^s) \leq u_i(e^*, f^*). \tag{27}$$

This result is intuitive as the agreement parameters are identity independent and the efficient allocation maximises welfare. The agreement will therefore always be formed. As participation in

[38] This idea of creating a more symmetric contest through the appropriate choice of ψ was studied recently, among others, by Franke, Kanzow, Leininger, and Schwartz (2011), and Franke (2012).

[39] Once a nation has committed her share of output $(1-\alpha)y(e_i)$ to the agreement, the only possibility for free-riding is on her reductive efforts—which we show in proposition 1 to be suboptimal. A deposit-based mechanism à la Gerber and Wichardt (2009) would rectify any (in our paper unmodeled) 'paying-up' problems among committed agreement members.

the agreement is individually rational, free-riding is fully deterred. Off the equilibrium path, the second-best contract C'—which is implemented if at least one player fails to participate—may still allow substantial emissions reductions. The following example shows that the agreement's raison d'être needs not necessarily be surrendered to holdup attempts.

Example 5: Our argument in proposition 3 uses the maximal threat $C' = \langle \alpha' = 1, \beta'; p(\mathbf{f})' \rangle$ to show that free-riding can always be discouraged. This extreme case, however, renders the agreement wholly ineffective if a punishment becomes necessary. The purpose of the following example is to show that such severe measures are not generally needed. Typically, the punishment of a deserter leaves enough freedom to increase abatement levels over those realising under no agreement. As outlined in the previous sections, contract C implements efficient efforts. Consider now a deviation by some player which triggers $C' = \langle \alpha', \beta'; p(\mathbf{f})' \rangle$. Denote the equilibrium agreement utility attained by adhering to C' by $u'(\cdot)$ and the corresponding equilibrium efforts by $e'(\alpha', \beta', p')$, $f'(\alpha', \beta', p')$. By inflicting sufficient damage through $m(\cdot)$, we need to ascertain that free-riding utility $u_i^s(e_i^s, f_i^s)$—with the agreement members adhering to C'—is smaller than what participation in C gives, i.e.,

$$u_i^s(e_i^s, f_i^s|C') = y(e_i^s) - s_i m(e_i^s + (n-1)e'(\cdot) - f_i^s - (n-1)f'(\cdot)) - c_e(e_i^s) - c_f(f_i^s) \leq u_i(e^*, f^*|C).$$

Since (4) implies that efforts e', f' are monotonic in α', β', payoff $u_i^s(\cdot|C')$ is continuous in α' and β'. Hence there exists an $\alpha' \in (\alpha, 1]$ which ensures the above inequality for suitable β' and $p'(\mathbf{f})$. Consider the case of $n+1$ players in the example setup of section 2.3. Then, full participation efficient efforts are given by

$$e^* = \frac{n+2}{2 \times 2^{1/3}\left((2+n)(3+2n)^2\right)^{1/3}}, \; f^* = \frac{n+1}{2 \times 2^{1/3}\left((2+n)(3+2n)^2\right)^{1/3}} \tag{28}$$

which are implemented by

$$\alpha^* = \frac{4\sqrt{e^*}(2e^* - f^*)(n+1) - 1}{n}, \; \beta^* = \frac{1}{n+1} + \frac{2(e^* - 2f^*)f^*}{\sqrt{e^*}r(\alpha - 1)} \tag{29}$$

for the Tullock success function parameterised by r. This determines $u_i(e^*, f^*|C)$. For the deviation utility $u_i^s(e_i^s, f_i^s|C')$, the deviation efforts e^d, f^d are determined by the first-order conditions

$$2e^s = \frac{1}{2\sqrt{e^s}} + \frac{2(f^s + n(f'(\cdot) - e'(\cdot)) - e^s)}{n+1}, \; f^s = \frac{e^s + n(e'(\cdot) - f'(\cdot))}{n+2} \tag{30}$$

in which $e'(\cdot), f'(\cdot)$ are the agreement equilibrium efforts in the agreement under C'. These functions $e'(\cdot), f'(\cdot)$, and therefore the damage they inflict on the deviator through $m(e^s + ne' - f^s - nf')$, are determined by

$$\alpha' = \frac{n\left(4\sqrt{e'}(e' + e^s - f^s + 2e'n - f'n) - 1\right) - 1}{n^2 - 1},$$

$$\beta' = \frac{\left(e^{2\prime}(4+8n) - 2f'(n-1)(f' + f^s + 2f'n - e^s) + 2e'(2e^s - 2f^s + f'(n-3)n) - \sqrt{e'}(n+1)\right)}{\left(4e^{2\prime}n(1+2n) - 4e'n(-e^s + f^s + fn) - \sqrt{e'}n(1+n)\right)}.$$

In our example setup, it turns out that participation is individually rational for any number of players. The details for the simplest case of three players (two in the agreement, one outside) are

$$e' = 0.302, f' = 0.195, e^s = 0.314, f^s = 0.132, (e^* = 0.273, f^* = 0.205)$$

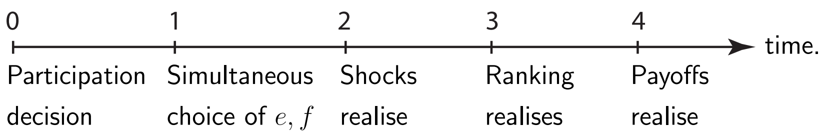
for $C' = \langle \alpha' \approx 0.910, \beta' = 1; r' = 1 \rangle$. If there is no agreement, efforts are $e^d = 0.303, f^d = 0.151$, so the 29% increased abatement efforts achieved by the punishment contract are substantial.◁

Two further (ad-hoc) ways in which agreement participation can be ensured in the deterministic output version of our model, i.e., granting most favoured 'green' trading terms only to participating nations and environmental certification, are illustrated in appendix A.2.

4 The stochastic output model

Since the previous section suggests that we cannot generally secure participation of an outsider in an existing (efficient) agreement without recurring to punishments or other ad-hoc measures, this section exploits the fact that the efficient mechanism introduced in the previous sections has redistributive properties in order to derive a stronger, more methodical argument for agreement formation. In order to capture the mutual insurance idea of redistributive agreements, we make two key modifications to the deterministic framework of the model section 2: we allow for stochastic output y and consider risk-averse decision makers. In particular, we assume that a nation's output process is stochastic, i.e., given by $y(e_i, \varepsilon_i) = \tilde{y}(e_i) + \varepsilon_i$, where $\tilde{y}(e_i)$ is weakly concave and the shocks ε_i are distributed according to the law $\mathcal{L}(\mu = 0, \sigma^2)$ characterized by the mean μ and the finite variance σ^2 of some continuously differentiable distribution F with symmetric probability density F' over (any subset) of $(-\infty, \infty)$.[40] As suggested by, for instance, Chamberlain (1983) or Owen and Rabinovitch (1983), any member of the class of elliptical distributions is eligible for this distribution F.[41]

The full timing of the modified interaction is



4.1 Preferences and information

Expected utility functions are defined over a decision maker's uncertain wealth w. The idea of risk aversion is incorporated in the curvature of the Bernoulli-utility function v over certain payoffs. We assume (in our symmetric model) that v is concave and that the player's embodied attitude towards risk is identical for all players. Recall that the only stochastic influence in our model comes from idiosyncratic shocks to output $y(e_i, \varepsilon_i) = \tilde{y}(e_i) + \varepsilon_i$, where $\tilde{y}(e_i)$ is weakly concave and the shocks ε_i

[40] The additive structure of the shock is crucial for the arguments we develop. We keep the stochastic process generating income uncertainty as general as possible in terms of distribution and variance while the mean is kept at zero for modelling convenience (otherwise the riskless model could not be used as a benchmark). For a combined flow & stock interpretation of income in our one-shot model this may not be entirely unreasonable. A financial crisis, for instance, could be interpreted as a negative shock to this composite variable.

[41] Elliptical distributions are a generalisation of the normal family containing, among others, the Student-t, Logistic, Laplace and symmetric stable distributions. A detailed presentation of these distributions is available in Landsman and Valdez (2003) and Fang, Kotz, and Ng (1987).

are distributed according to $\varepsilon_i \sim \mathcal{L}(\mu = 0, \sigma^2)$ over $(-\infty, \infty)$. We adopt the following formulation for player $i \in \mathcal{N}$

$$\mathbb{E}[u_i(\mathbf{e}, \mathbf{f}; \varepsilon_i)] = \int_{-\infty}^{\infty} v(\underbrace{y(e_i, \varepsilon_i) - s_i m \left(\sum_{i \in \mathcal{N}} (e_i - f_i) \right) - c_e(e_i) - c_f(f_i)}_{\text{risk-neutral wealth}}) dF(\varepsilon_i) \tag{31}$$

with Bernoulli utility function $v(0) = 0$, $v' > 0$, and $v'' \leq 0$. Since the player's choice of efforts is invariant under any increasing, concave transformation of v, we can split the optimisation stage—giving the choice of efforts—from the risk-based participation decision. Hence, methodologically, we start by analysing the above underbraced decision problem in isolation—this step is equivalent our risk-neutral analysis in section 3—and discuss the risk transformation separately.

Example 6: In this example, we study a simple, pure redistributive mechanism based on a harmonised output tax. In the simplest possible case of a pure redistribution agreement (without contest), each of two symmetric individuals' utility is given, for $P = (1 - \alpha)(y(e_1, \varepsilon_1) + y(e_2, \varepsilon_2))$, by

$$\alpha y(e_i, \varepsilon_i) + q_i P - s_i m(e_1 + e_2 - f_1 - f_2) - c_e(e_i) - c_f(f_i). \tag{32}$$

The difference to the case of no redistribution (31) is simply the presence of the predetermined subsidy share q_i of the federal budget P. Maximisation of this objective function leads—for the quadratic costs and square-root production function case of (4)—to the individually optimal effort pair which, for symmetric $s_i = q_i = 1/2$, is given as

$$\hat{e} = \frac{(1+\alpha)^{2/3}}{2^{4/3} \times 3^{2/3}}, \ \hat{f} = \frac{(1+\alpha)^{2/3}}{2 \times 2^{4/3} \times 3^{2/3}} \tag{33}$$

implying that $\hat{e} = e^*$ for $\alpha^* = 4/5$. Since $\hat{e} = 2\hat{f}$, however, no $\alpha \in [0, 1]$ can implement efficient abatement f^* and a simple redistributive policy can therefore never implement efficient abatement.[42]

Remarkably, however, for i.i.d. output shocks ε_i, the variance of the expected share of the pooled and redistributed resource P is given by

$$\mathbb{V}\left[\frac{1}{2}(1-\alpha)\left(y(e_1, \varepsilon_1) + y(e_2, \varepsilon_2)\right)\right] = \frac{(1-\alpha)^2}{2}\sigma^2 < \sigma^2. \tag{34}$$

The redistribution pool has therefore a lower variance than individual output σ^2. Obviously, the resulting variance is lowest if all individual income $y(e_i, \varepsilon_i)$ is pooled and redistributed (i.e., $\alpha = 0$). A similar compression is shown for *all* higher moments of the redistribution pool's distribution in appendix A.4. Thus, although we restrict attention to distributions characterised completely by mean and variance in this paper, our argument should be, in principle, applicable to a much larger class of distributions.[43] $\triangleleft$

[42] This is a general result; for a derivation (in a different context) see, for instance, Giebe and Schweinzer (2013).

[43] We would like to reiterate that the pure redistributive mechanism studied in this subsection cannot achieve efficiency. If, for some reason, efficiency is not an issue or is out of reach for exogenous reasons, then all variance compression results derived in what follows for the contest mechanism are also applicable for the (much simpler) pure redistributive mechanism.

4.2 The efficient stochastic mechanism

We now illustrate the influence of stochastic output of the form $y(e_i, \varepsilon_i) = \tilde{y}(e_i) + \varepsilon_i$, i.i.d. $\varepsilon_i \sim \mathcal{L}(\mu = 0, \sigma^2)$, on the equilibrium of the symmetric contest game.

Example 7: Consider the square-root output $\tilde{y}(e_i) = e_i^{1/2}$, quadratic cost setting of our example from section 2.3 where $n = 2$ and symmetric $s_i = 1/n$. Recall that under the efficient mechanism (8), player i expects *equilibrium* utility[44]

$$u_i(\mathbf{e}^*, \mathbf{f}^*; \boldsymbol{\varepsilon}) = \alpha^* y_i(e^*, \varepsilon_i) + \frac{1}{n}P - \frac{1}{n}(ne^* - nf^*)^2 - (e^*)^2 - (f^*)^2, \tag{35}$$

with $P = n(1-\alpha^*)\sum_{i=1}^n y(e^*, \varepsilon_i)$. The main insurance argument rests on the simple observation that the variance of individual output σ^2 is higher than that of the pooled resource $\tilde{\sigma}^2$ to which all players contribute a share. The intuition is that the risk inherent in individual shocks with zero expectation evens out among several players. This is easiest to see if we assume i.i.d. shocks ε_i for which $\mathbb{V}[\sum_i \varepsilon_i] = \sum_i \mathbb{V}[\varepsilon_i]$. In the symmetric model, players expect a prize P with probability $1/n$ in equilibrium; therefore the variance of this prize expectation is given by

$$\mathbb{V}\left[(1-\alpha^*)\frac{1}{n}\sum_{i=1}^n \left((e_i^*)^{\frac{1}{2}} + \varepsilon_i\right)\right] = (1-\alpha^*)^2 \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{(1-\alpha^*)^2}{n}\sigma^2 < \sigma^2. \tag{36}$$

Therefore, the player expects a payout from the prize pool which is less risky than the own income. Therefore, any convex combination between $\alpha^* y(e^*) + (1-\alpha^*)y(e^*)$ is less risky that the stand-alone GDP. Hence, the higher the redistribution inside the agreement, the more pronounced the co-insurance effect. Using an analogous argument on the remaining term, the equilibrium variance of the complete contest is given by

$$\begin{aligned}\tilde{\sigma}^2 = \mathbb{V}\left[\left(\alpha^* + \tfrac{1-\alpha^*}{n}\right)\left((e_i^*)^{\frac{1}{2}} + \varepsilon_i\right) + \frac{1-\alpha^*}{n}\sum_{j=1,\ j\neq i}^n \left((e_j^*)^{\frac{1}{2}} + \varepsilon_j\right)\right] = \\ \left(\alpha^* + \frac{1-\alpha^*}{n}\right)^2 \sigma^2 + (1-\alpha^*)^2 \frac{n-1}{n^2}\sigma^2 = \frac{(\alpha^*)^2(n-1)+1}{n}\sigma^2 < \sigma^2 \text{ for all } n \geq 2. \triangleleft\end{aligned} \tag{37}$$

Thus, ceteris paribus, nations endowed with variance-averse preferences favour the redistributive contest over standing alone. In the extreme case of $\alpha^* = 0$, where all output is pooled, the pooled variance sinks to $\mathbb{V}[u_i(\mathbf{e}^*, \mathbf{f}^*; \boldsymbol{\varepsilon})] = \sigma^2/n$. Our main participation result relaxes the special assumptions made on output and the independence of shocks used in the above example. All we require for our first result is that not all ε_i, $i \in \mathcal{N}$, exhibit perfect positive correlation.

Proposition 4. *Consider individual output $y(e_i, \varepsilon_i) = \tilde{y}(e_i) + \varepsilon_i$, $i \in \mathcal{N}$, for identically distributed $\varepsilon_i \sim \mathcal{L}(\mu = 0, \sigma^2)$ with covariance σ_{ij} for $(\varepsilon_i, \varepsilon_j)$. Then any balanced budget mechanism which is redistributive, i.e., $\alpha^* < 1$ and symmetric, i.e., assigns equal winning probabilities in symmetric pure strategy equilibrium, has a lower variance than individual output.*

[44] Notice that, since the efficient mechanism uses the ranking only for incentive purposes (based on expectations), the share $(1-\alpha)y(e_i, \varepsilon_i)$ that players commit to stochastic.

The main idea of our participation argument is that this reduction in income risk can be used to motivate the formation of an agreement. Although we study exclusively incentives for joining an international environmental agreement in this paper, the same basic variance compression argument should be applicable to many other international bodies. In order to show the versatility of the idea, we now turn to the analysis of the asymmetric redistributive contest.

The next proposition verifies the variance-compression intuition developed for the symmetric redistributive pool for the asymmetric case under both independent and non-independent shocks. Again, we assume that shocks ε_1 and ε_2 are not perfectly positively correlated.

Proposition 5. *Consider a two-player, balanced budget contest mechanism which is redistributive, i.e., $\alpha_1 < 1$, $\alpha_2 < 1$, and asymmetric, i.e., assigning not necessarily equal winning probabilities $p_1(\mathbf{e}, \mathbf{f}), p_2(\mathbf{e}, \mathbf{f})$ in asymmetric pure strategy equilibrium. The equilibrium payoff from this class of mechanisms has a lower variance than individual output if and only if the following conditions are fulfilled for all $(i, j) \in \{1, 2\}$*

$$\begin{aligned} 1 &> \beta_i^2(1-\alpha_j)^2 + (\beta_i - \alpha_i\beta_i - \alpha_i)^2 + 4\alpha_i\beta_i(1-\alpha_i) \text{ for i.i.d. shocks,} \\ 1 &> \beta_i(3 - 2\alpha_i - \alpha_j)(2\alpha_i + \beta_i - \beta_i\alpha_j) + (\beta_i - \alpha_i\beta_i - \alpha_i)^2 \text{ otherwise.} \end{aligned} \tag{38}$$

The condition for variance compression depends on there being some redistribution in the first place: $\alpha_1^* + \alpha_2^* < 2$. Moreover, since we are only looking at two players, each $\alpha_i^* < 1$ in order to allow for risk pooling. From the point of view of player $i = 1, 2$, her winner's share $\beta_i^* > 1/2$ determines the income transfer in case of winning and $1 - \beta_j^*$, $j = 3 - i$, determines how much income is redistributed in case she loses. For asymmetric winning probabilities $p_i(\mathbf{e}, \mathbf{f})$, the interplay of these variables in (38) determines when risk-pooling is possible.

Equations (38) look more complicated than they are. The reason is that we want to specify parameters (α, β) in sufficient generality to be applicable for any redistribution problem. Indeed, the conditions mean that our efficient mechanism leads to a lower variance than individual output if and only if the pair of parameters (α_i, β_i) for $i = \{1, 2\}$ are small enough. To provide an intuition, remark that if $\alpha_i \to 0$ and $\alpha_j \to 0$, our mechanism is better in terms of variance reduction if and only if $\beta_i, \beta_j \in [0, \sqrt{6}/2]$. Moreover our mechanism is still better if $\beta_i \to 0$ and $\beta_j \to 0$ for all values of α_i, α_j in $[0, 1]$.

4.3 The general participation result

In this subsection, we show that our variance-compression results can be used to argue that the distribution of the shock expected by a player not participating in the redistribute agreement is a mean-preserving spread of the shock expected by agreement members.[45] As a consequence, we can show that a sufficiently variance-averse player will prefer to join the agreement over staying outside.

In the simplest case of a n-member, pure redistributional contest mechanism (superscript rm), a symmetric member's expected equilibrium payoff is

$$\mathbb{E}[u^{\text{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\text{rm}})] = \int_{-\infty}^{+\infty} (\tilde{y}(e^*) + \varepsilon^{\text{rm}} - s_i m(ne^* - nf^*) - c_e(e^*) - c_f(f^*))\, dF(\varepsilon^{\text{rm}}). \tag{39}$$

[45] See Rothschild and Stiglitz (1970) for the idea of mean-preserving spreads as a measure of risk.

Given an existing agreement with $n-1$ participants, the equilibrium payoff of a player $i \in \mathcal{N}$ who is free-riding (superscript fr) on the abatement efforts of the agreement is given by

$$\mathbb{E}[u^{\text{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{fr}})] = \int_{-\infty}^{+\infty} \Big(\tilde{y}(\tilde{e}) + \varepsilon^{\text{fr}} - s_i m(\tilde{e} + (n-1)e^* - \tilde{f} - (n-1)f^*) - c_e(\tilde{e}) - c_f(\tilde{f})\Big)\, dF(\varepsilon^{\text{fr}}) \tag{40}$$

in which $\tilde{\mathbf{e}}, \tilde{\mathbf{f}}$ are equal to $\mathbf{e}^*, \mathbf{f}^*$ with the free-rider's positions replaced by $\tilde{e}, \tilde{f}$. We would like to show that, under a sufficiently concave transformation v of the utility implied by (39) and (40), we can ascertain agreement participation, i.e., that $\mathbb{E}[v(u^{\text{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\text{rm}}))] \geq \mathbb{E}[u^{\text{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{fr}})]$.[46] Notice that, for this purpose, we can ignore the influence of the individually suffered damage share $s_i m(ne^* - nf^*)) < s_i m(\tilde{e} + (n-1)e^* - \tilde{f} - (n-1)f^*)$.

Existing results for the case of $\varepsilon^{\text{rm}} = \varepsilon^{\text{fr}} \equiv 0$ show,[47] however, that the typical case is $\mathbb{E}[u^{\text{rm}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{rm}})] \leq \mathbb{E}[u^{\text{fr}}(\mathbf{e}^*, \mathbf{f}^*); \varepsilon^{\text{fr}}]$ which implies that no agreement is formed because

$$u^{\text{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\text{rm}})\Big|_{\varepsilon^{\text{rm}} \equiv 0} < u^{\text{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{fr}})\Big|_{\varepsilon^{\text{fr}} \equiv 0}. \tag{41}$$

In our stochastic environment, remark that ε^{fr} and ε^{rm} are two random variables which follow two specific—zero mean—distributions, $\varepsilon^{\text{fr}} \sim \mathcal{L}(0, \sigma^2)$ and $\varepsilon^{\text{rm}} \sim \mathcal{L}\left(0, \frac{\alpha^2(n-1)+1}{n}\sigma^2\right)$ with $\sigma^2 > \frac{\alpha^2(n-1)+1}{n}\sigma^2$, i.e., $\mathbb{V}[\varepsilon^{\text{fr}}] > \mathbb{V}[\varepsilon^{\text{rm}}]$. Indeed, in contrary to an isolated free rider whose income shocks are distributed as for independent individuals, agreement members face only mutualised risk in equilibrium. As shown in proposition 4, this *ex-post* variance of the distribution of mutualised risk is lower than individual income variance. Notice further, that the symmetric equilibrium solutions to (39) and (40) are invariant under increasing concave transformations, i.e., that the equilibrium choice of effort does not change if the concerned decision makers change their degree of risk aversion. Hence, there exists a function $v(\cdot)$, with $v'(\cdot) > 0, v''(\cdot) \leq 0$, which leads to $\mathbb{E}[v(u^{\text{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\text{rm}}))] \geq \mathbb{E}[v(u^{\text{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{fr}}))]$ for any positive difference $\mathbb{V}[\varepsilon^{\text{fr}}] - \mathbb{V}[\varepsilon^{\text{rm}}]$. Our main participation result then follows immediately.

Proposition 6. *For every positive difference of the equilibrium variances between the redistributive and the free-riding mechanisms, there is a family of concave functions v which provides a higher payoff to the redistributive mechanism.*

Proof. Assume that (41) holds (otherwise we are done), where $u^{\text{rm}}(\cdot)$ and $u^{\text{fr}}(\cdot)$ are both linear functions in their respective arguments ε^{rm} and ε^{fr}. From theorems 7 and 8 in appendix C, it follows that $\varepsilon^{\text{fr}} \leq_{\text{icv}} \varepsilon^{\text{rm}}$ such as $\mathbb{E}[v(\varepsilon^{\text{fr}})] \leq \mathbb{E}[v(\varepsilon^{\text{rm}})]$ for all increasing and concave functions v. Consequently, there exists a sufficiently concave, increasing function v such that $\mathbb{E}[v(u^{\text{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\text{rm}}))] \geq \mathbb{E}[v(u^{\text{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\text{fr}}))]$. ☐

[46] For the sake of simplicity, we consider here only equilibrium payoffs for risk-neutral agents. Yet, this result still holds under any concave transformation for risk-averse agents. That is a direct implication of our assumptions on preferences.

[47] Solid theoretical arguments against agreement participation in the deterministic case were derived, for instance, by Diamantoudi and Sartzetakis (2006) and Guesnerie and Tulkens (2009).

This proposition shows that, in our stochastic setup, there is a degree of risk aversion which leads to full participation in the symmetric redistribution agreement. An illustration of the intuition is attempted in figure 3.

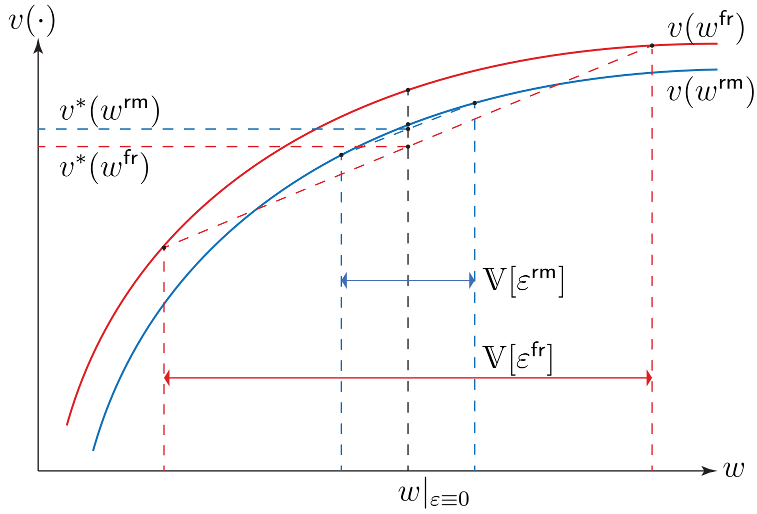


Figure 3: Variance compression leads to participation under sufficient variance aversion.

Example 8: We now extend our simple symmetric example from section 2.3 in order to explicitly make the point that variance aversion simplifies the satisfaction of a player's participation constraint. Reconsider the problem of player $i \in \{1, 2\}$ under Bernoulli utility function $v(\cdot)$ and standard normally distributed shocks $\varepsilon \sim \mathcal{N}(0, \sigma^2 = 1)$. Since players decide their efforts on the basis of expected shocks of zero, the efficient efforts defined in (4) remain unchanged and the efficiency inducing parameters $\alpha^* = 3/5, \beta^* = 7/12, r = 2$ are still given by (8). Hence, a player's expected equilibrium utility from participating in the redistributive mechanism is given by the equivalent to (39) as

$$\mathbb{E}[u^{\mathsf{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\mathsf{rm}})] = \int_{-\infty}^{\infty} v\left((e^*)^{1/2} + \varepsilon^{\mathsf{rm}} - \frac{1}{3}(3e^* - 3f^*)^2 - (e^*)^2 - (f^*)^2\right) dF(\varepsilon^{\mathsf{rm}}). \quad (42)$$

Similarly, a player's equilibrium utility from free riding on the residual two-player agreement follows from (40) as

$$\mathbb{E}[u^{\mathsf{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\mathsf{fr}})] = \int_{-\infty}^{\infty} v\left(\tilde{e}^{1/2} + \varepsilon^{\mathsf{fr}} - \frac{1}{3}(\tilde{e} + 2e^* - \tilde{f} - 2f^*)^2 - \tilde{e}^2 - \tilde{f}^2\right) dF(\varepsilon^{\mathsf{fr}}). \quad (43)$$

Solving these problems under linear $v(\cdot)$ for their symmetric equilibrium utilities gives for the agreement

$$e^* = 0.273,\ f^* = 0.205,\ \alpha^* = 0.455,\ \beta^* = 0.574 \ \Rightarrow\ \mathbb{E}[u^{\mathsf{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\mathsf{rm}})] = 0.392 \quad (44)$$

and, for the free-riding player,[48]

$$\tilde{e} = 0.317,\ \tilde{f} = 0.126 \ \Rightarrow\ \mathbb{E}[u^{\mathsf{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\mathsf{fr}})] = 0.399. \quad (45)$$

[48] The two-player agreement efforts and utilities required for the calculation of the free-riding utility are the same as in the example of subsection 4.2.

Therefore, a risk neutral player will not join the two-player agreement. Calculating the same expected equilibrium utilities under a concave Bernoulli utility function $v(w) = w^{3/4}$, however, gives

$$\mathbb{E}[v(u^{\mathsf{rm}}(\mathbf{e}^*, \mathbf{f}^*; \varepsilon^{\mathsf{rm}}))] \; = 0.495 > \mathbb{E}[v(u^{\mathsf{fr}}(\tilde{\mathbf{e}}, \tilde{\mathbf{f}}; \varepsilon^{\mathsf{fr}}))] = 0.435 \tag{46}$$

and therefore full participation in the three players agreement.◁

Section 4 presents a redistributive agreement model in which an individual member's variance of income is compressed through joining the agreement. Under the assumptions we make, variance aversion implies risk aversion and, hence, risk averse players prefer less to more income variance. The variance compression is reached on the part of income which is pooled and redistributed among members: the risk sharing is an effect of the reduced risk of the pooled and redistributed income. We use this property of redistributive agreements in a contest model which implements efficient efforts choices among nations facing multiple external effects. Since this balanced budget contest must redistribute wealth in order to implement efficient efforts—through awarding member states first, second, etc prizes—the efficient contest compresses the income risk of member states.

Although the model we present is highly stylised and our distributional assumptions are made for reasons of modelling simplicity rather than realism, the insurance property can be derived as an entirely general feature of redistributional mechanisms. This insurance aspect is a formidable reason to join international agreements which, so far, seems to have been overlooked in the agreement formation literature. Since, moreover, our variance compression results are already obtained for small agreements, this paper presents a theoretical rationale to start an agreement among a handful of progressive nations, let them reap the benefits of risk sharing, and only gradually enlarge the successful international agreement.

5 Concluding remarks

This paper gives theoretical reasons for independent nations which are sufficiently averse to income shocks to agree on a simple, redistributive contest, organised among themselves, which can implement both efficient productive and pollution abatement efforts. Nevertheless, the desirable characterisation of many properties of this mechanism must be left for future work: Which share of global (per capita) GDP would have to be redistributed—in reality—to the country with the highest emissions reduction in order to implement our results? Is the resulting wealth redistribution one we would like to see? Can the mechanism's design parameters be effectively negotiated? Answers to all these questions have significant policy implications, are to a large extent empirical and are at least partly determined by politics. At any rate we do not feel qualified to answer these questions now.

There is, however, a set of immediate challenges to the mechanism we propose which we can respond to now and would like to address in the remainder of this concluding discussion. i) Measurement of output. At the national level, y corresponds to GDP. But GDP measurement relies on approximation, and GDP estimates are often revised. If we are seriously contemplating large international cash transfers that depend on national output figures, the accuracy, and manipulability, of those measures is a concern. While it is not sufficient for the measure of y to be 'right on average' in order to achieve efficiency, it is also true that our measure of efficiency (welfare maximisation) relies on the same measurement imperfections. So our mechanism is as good as a *complete information* mechanism can be in this setup. Obviously, introducing private information would improve the realism of our setup greatly—but since we do not have a model implementing efficiency and ensuring participation even under full information, we are reluctant to attempt a solution of the incomplete information case directly.

ii) Non-manipulability of the 'abatement effort monitoring device.' Similarly, the manipulability of any monitoring device must be an issue for our mechanism. It is remarkable, however, that monitoring of reductive efforts f as required by our model needs not be perfect—on the contrary, the precision of the detector is a design element of the agreement we propose. Imperfect discrimination is one of the main features of the contest technology we employ. Hence, in our mechanism, imperfect measurement of f becomes a 'feature' and not a 'bug.'

iii) Commitment to share-of-output payments. By becoming an agreement member, a country pledges a certain share of national output. Committing to this pledge is a political issue which is equivalent to leaving the agreement. Since the paper discusses several scenarios in which joining the agreement is rational, all of these are also effective in avoiding agreement desertion in the form of withdrawal of pledged resources. We would also like to point out that there are several existing international agreements and unions among (formerly) independent states which redistribute pledged resources in a way not too distant from our model.

iv) Separation between productive and reductive efforts. As we model efforts as either productive or reductive, a technology in our model cannot, apparently, be both productive *and* reductive. This is only a model simplification: adding another concave production function based on reductive efforts to the maximisation problem would make our sufficient existence condition *easier* to satisfy. Hence, the omission of a productive aspect of abatement is only a simplification.

v) The treatment of countries as single, profit-maximising decision makers. While this is a standard modeling assumption, the micro-politics of decision making on production (or abatement) levels may well be much more challenging than suggested by our simple model. Who pays or receives the marginal benefit of transfers through these contests? Who actually owns the output and therefore funds the prize pool? Although our model cannot address these questions in its present form, our main ideas could be equally well applied on the state or municipal levels where the micro-actors would be much easier to identify. In a similar vein, another immediate application possibility is to 'smaller' abatement competitions at separate industry levels of the participating countries.

vi) Unrealistically large transfers. While we present the incentives in our mechanism in terms of winning probabilities, we would like to stress that an equivalent interpretation in terms of actual effort-dependent shares of the prize pot is possible without changing our results. The obvious advantage of such a mechanism is that, in symmetric pure-strategy equilibrium, all payments cancel each other out and no net-transfers are necessary. The downside is that the statistical (cost) advantage of requiring only ordinal information to determine a 'winner' are lost.

We neither belittle nor shrug off any of these important problems an actual agreement needs to solve. To a large extent, however, we feel that *any* mechanism attempting to solve the emissions problem will have to face a variant of these problems. The present paper attempts to name and discuss these challenges—and provides first results showing that a mechanism along the lines we indicate can *in theory* correct nations' combined incentives to emit too much while abating too little.

Appendix A: Extensions and robustness

A.1 Asymmetric pollution shares

In this subsection, we extend the example setup of section 2.3 with unequal relative damage shares $s_i \in (0,1)$, $i = 1,2$. Since shares sum to 1, both efficient effort types are still given by (4). Player $i's$ problem is unchanged and imposing efficiency (4) we obtain the shares

$$\alpha_i^* = \frac{1}{5}(1+4s_i),\ \beta^* = \frac{1}{2} + \frac{1}{6r}. \tag{47}$$

Notice that only α_i^* turns out to depend on the player's identity (class), the efficiency-inducing prize structure β is identical to the symmetric case. As to be expected, the share $(1-\alpha_i^*)$ of output which has to be committed to the contest gets arbitrarily small when the public bad problem disappears as s_i approaches 1. On the other extreme, a player who does not suffer from the effects of global warming at all must be asked to commit close to $4/5$ of her output to the contest in order to induce efficient efforts on her behalf. A numerical example taking relative damage shares of $s_1 = {}^1/_4$, $s_2 = {}^3/_4$ requires $\alpha_1 = 0.4$ and $\alpha_2 = 0.8$ in order to implement efficiency.

In the more general case of $n > 2$ players with damage shares parameterised by $s_i = \frac{2i}{n+n^2}$, $i = 1,2,\ldots,n$, with $\sum_{i=1}^{n} s_i = 1$, efficient efforts are given by

$$4e(1+n) = \frac{1}{\sqrt{e}} + 4fn, en = f(1+n) \ \Leftrightarrow \begin{cases} e^* = \dfrac{1+n}{2\times 2^{1/3}\left((1+n)(1+2n)^2\right)^{1/3}}, \\ f^* = \dfrac{n}{2\times 2^{1/3}\left((1+n)(1+2n)^2\right)^{1/3}}. \end{cases} \tag{48}$$

The n-player individual asymmetric problem in the example setup under the two-part price structure $\left(\beta, \frac{1-\beta}{n-1}, \ldots, \frac{1-\beta}{n-1}\right)$ employed previously is

$$\begin{aligned} \underset{(e_i,f_i)}{max}\ \alpha_i e_i^{1/2} + \frac{f_i^r}{f_i^r + (n-1)(f^*)^r}\beta P + \left(1 - \frac{f_i^r}{f_i^r + (n-1)(f^*)^r}\right)\left(\frac{1-\beta}{n-1}\right)P- \\ s_i(e_i + (n-1)e^* - f_i - (n-1)f^*)^2 - (e_i^2 + f_i^2) \end{aligned} \tag{49}$$

which is solved by the intimidating but straightforward efficient parameters

$$\begin{aligned} \alpha_i^* &= \frac{\sqrt{\frac{1+n}{((1+n)(1+2n)^2)^{1/3}}}(1+n+ns_i)n - \left((1+n)(1+2n)^2\right)^{1/3}}{(n-1)\left((1+n)(1+2n)^2\right)^{1/3}}, \\ \beta^* &= \frac{(n-1)n(1+2n)^2\left(\frac{1+n}{\left((1+n)(1+2n)^2\right)^{1/3}}\right)^{3/2} + n\left(n^2-1\right)(1+n+ns_i) + 2(1+n)^2 r(2+n(3+s_i))}{2n(1+n)^2 r(2+n(3+s_i))} \end{aligned} \tag{50}$$

where, again, β is not identity dependent. A numerical example for $n = 187$, $r = 3$ gives for 'type' $s_i = 1/n$ (i.e., $i = 94$) an efficiency-inducing $\alpha_i^* = 0.50133$ and redistribution vector of

$\left(\beta^* = 0.17024, \frac{1-\beta^*}{n-1} = 0.00446, \ldots, \frac{1-\beta^*}{n-1} = 0.00446\right)$ which compares to the flat $1/n = 0.00534$. Under the contest, type $s_i = 1/n$ gives up roughly 50% of her output but gets back 41.6% even if losing the contest. She gets almost 16 times her output if she wins. Notice that, if equilibrium existence allows, then the two can be further equalised by employing a more precise ranking and thereby increasing r.

A.2 Enforcing participation in the deterministic output model

The purpose of this subsection is to show that a simple way to deter free-riding of individual nations even in the deterministic output version of our model is to grant most favoured 'green' trading terms only to participating nations. Similarly, environmental certification conditional on treaty commitment can be a powerful complementary tool to enforce participation in the IEA.

The expansion of equilibrium abatement efforts from their no agreement level f^0 to the level within the agreement f^* creates 'green' products; the higher abatement efforts, the greener is productive output. The idea is to label free-riding countries' products and thereby creating (political) incentives to respect commitments to the IEA. Consequently, firms' lobbying against environmental regulation may result in that country's desertion followed by the IEA labelling its goods. We thus propose a negative label which signals a product lacking the 'green' environmental standards enforced by the IEA.[49]

This idea can be formalised in our setup by decreasing the value of a deserting country's output. Loosely speaking, we require that—once certified as environmentally unfriendly—a consumer's willingness to pay for labelled products decreases.[50] Therefore, the revenue generated from the production of labelled products sinks as these products suffer a decrease in price of $x(\mathbf{f}) \in [0, 1]$. This fraction corresponds to the deserter's deviation from agreed abatement levels. Denoting the outside equilibrium efforts of a single deserter by (e^d, f^d), desertion utility is

$$u_i^d(e_d, f_d) = \underbrace{\left(\frac{f^* - f_i^0}{f^*}\right)}_{=x\in[0,1]} y(e_i^d) - s_i m(e_i^d + (n-1)e^* - f_i^d - (n-1)f^*) - c_e(e_i^d) - c_f(f_i^d) \quad (51)$$

for $i \in \mathcal{N}$ and $n > 2$ (since there must be at least two players left in the agreement after i deserts). A sufficiently large fraction x will successfully deter free-riding on reductive efforts and can be seen as alternative to the blunt global threat represented by contract C'.[51]

A further step in this direction is the formation of an exclusive trade agreement. If a deserter can be excluded from the fraction of trade corresponding to the necessary abatement investments within

[49] There are many green labelling examples: UK supermarket chain Tesco has recently introduced a promotional campaign on carbon labels. US Walmart and French Casino have similar ambitions. Examples of negative labelling campaigns are the mandatory GMO labelling implemented in Europe, and "we're Greenpeace, and we want a fresh green Apple" targeted at US computer maker Apple. Grankvist, Dahistrand, and Biel (2004) argue that negative labelling may have a higher consumption impact than positive labelling. Engel (2004) underlines the necessity to inform consumers, especially when a firm is found cheating on its environmental claims.

[50] Models on certification and standard settings have been studied intensely; see for instance Lerner and Tirole (2006) or Harbaugh, Maxwell, and Roussillon (2011)

[51] To some extent, the particular choice of $x(\mathbf{f})$ is arbitrary. Any function of abatement efforts implementing sufficient deterrence (such as the function used for the exclusive trade example below) could be used instead.

the agreement, then individual desertion can, again, be discouraged. As above, green production—generated by the expanded abatement effort—is traded among countries and produces wealth. Our model does not take into account the international trading aspects of production and thus there is no direct way of measuring the involved consequences to individual wealth.[52] A simple (ad-hoc) way of nevertheless capturing the idea of enforcement through an exclusive trading agreement is to restrict trade on (and therefore capitalising on) the fraction of productive output corresponding to the reductive investments within the treaty to agreement members. Then player i's desertion utility in (51) can be reduced by using

$$x(\mathbf{f}) = \left(1 - \frac{f^* - f_i^0}{e^*}\right) \tag{52}$$

in which reductive effort f_i^0 is the equilibrium abatement level without agreement. As indicated above, to some extent this choice is arbitrary.[53] The intuition for (52) is that an agreement defector j can free-ride on the reductive efforts of agreement members (through a cleaner environment) but is punished by restricted access to the agreement market consisting of the tradables $\sum_{h\neq j} y(e_h)$.

Example 9: Returning to the simple quadratic cost, square-root production example of section 2.3, this implies that it is individually rational to participate in the agreement if

$$u_i(e^*, f^*) \geq x(\mathbf{f}) e_i^{\frac{1}{2}} - \frac{1}{n}\left(e_i + (n-1)e^* - f_i - (n-1)f^*\right)^2 - e_i^2 - f_i^2. \tag{53}$$

Solving the deserter's maximisation problem (on the rhs) for the exclusive trade agreement under (52) leads to the first-order conditions

$$2f^* + \frac{e^* - f^* + f_i^0}{2e^*\sqrt{e_i^d}} = \frac{2(e_i^d + f^* - f_i^d + e^*(n-1) + e_i^d n)}{n}, \quad f_i^d = \frac{e_i^d + (e^* - f^*)(n-1)}{n+1} \tag{54}$$

in which (e^*, f^*) are the equilibrium effort levels provided inside the agreement. Plotting the utilities from the desertion efforts (e_i^d, f_i^d) solving above first-order conditions results in a graph similar to figure 1 for the case of $n = 3$ showing that deviations are not profitable. Thus, the exclusive trade agreement ensures participation. (The analysis is nearly identical for the labelling setup discussed above and therefore not replicated.) As the severity of punishment $(1 - x(\mathbf{f}))$ in (52) is given by

$$\frac{f^* - f_i^0}{e^*} = \frac{6 - 3^{1/3}(-2-4n)^{2/3}(-1-n)^{1/3}}{6(1+n)} \tag{55}$$

which is increasing in n, the punishment gets more severe for larger n and participation is easier to obtain in the general case. (The limit as $n \to \infty$ equals $2^{1/3}/3^{2/3} \approx .606$.)◁

A.3 The choice of ranking technology

Consider a n-player extension of the problem of section 2.3 with prize structure $\left(\beta, \frac{1-\beta}{n-1}, \ldots, \frac{1-\beta}{n-1}\right)$. The present example shows that efficiency can also be obtained in proposition 2 for a 'difference-form' success function. The specific properties of the generalised Tullock success function which we

[52] Modelling both these aspects formally is possible and certainly provides grounds for future research.

[53] If monitoring of the deserter's abatement efforts is good, then the actual level f_i^d could be used. (In the present example setup, this leads to a rather unintuitive corner solution.) The idea, however, seems important for dealing with competing abatement agreements: As long as they are effective, there is no reason to punish them.

use in the remainder of the paper are therefore not crucial to our results. Difference-form success functions have been widely used in the literature, for instance by Che and Gale (2000), but suffer from the lack of a generally accepted, simple extension to more than two players. We define player i's probability of winning as

$$p_i(\Delta) = \frac{\exp^{\Delta_i^r}}{\sum_{j=1}^n \exp^{\Delta_j^r}}, \text{ where } \Delta = (\Delta_1, \ldots, \Delta_n),\ \Delta_i = f_i - \frac{\sum_{j\neq i} f_j}{n-1}, \text{ and } r > 0. \quad (56)$$

Setting $P = (1-\alpha)(e_i^y + (n-1)e_j^y)$, $y \in (0,1)$, $m, b > 1$ and all $j \neq i$ equal, player i's individual problem is to

$$\underset{(e_i,f_i)}{max}\ \alpha e_i^y + p_i(\Delta)\beta P + (1-p_i(\Delta))\frac{1-\beta}{n-1}P - s_i(e_i + (n-1)e_j - f_i - (n-1)f_j)^m - (e_i^b + f_i^b)$$

which, in symmetric equilibrium $e = e_i = e_j$, $f = f_i = f_j$ gives for any $p_i(\Delta)$

$$\alpha = \frac{e^{-y}\left((e-f)\left(be^b n - e^y y\right) + em((e-f)n)^m s_i\right)}{(e-f)(n-1)y},$$
$$\beta = \frac{e^{-y}\left(-b(e-f)f^b(n-1)n + f\left(m(n-1)((e-f)n)^m s_i + e^y(e-f)n^2(\alpha-1)p_i'(0)\right)\right)}{(e-f)fn^3(\alpha-1)p'(0)}$$

in which $\Delta = 0$ is the equilibrium vector of deviations.

Example 10: Plugging in the efficient efforts from (4), employing (56), and returning to the example setup from section 2.3 (i.e., $n = 2$, $y = 1/2$, $b = m = 2$, and $s_i = 1/2$), this results in a very similar efficient mechanism as under the Tullock success function

$$\alpha^* = \frac{3}{5},\ \beta^* = \frac{r + (5/6)^{2/3}}{2r} \quad (57)$$

in which $\beta^* \in (.5, 1]$ is ensured for $r \geq (5/6)^{2/3} \approx 0.89$. A picture nearly identical to figure 1 confirms, for instance, $\langle \alpha^*, \beta^*, r = 2\rangle$ as equilibrium. The precise form of ranking technology employed is thus immaterial to our results.◁

A.4 Higher moment compression

Consider now the higher moments of the redistributive pool P. In general, the t^{th} central moment of a random variable Y with real and continuous probability density $f(y)$ and mean μ is defined as

$$\mu_t = \mathbb{E}[(Y - \mathbb{E}[Y])^t] = \int_{-\infty}^{\infty} (y-\mu)^t\ f(y)\ dy, \text{ for } t = 0, 1, 2\ldots \quad (58)$$

For the first four moments, this defines the zeroth central moment μ_0 as one, the first central moment μ_1 as zero, the second central moment μ_2 as the variance σ^2, the third central moment μ_3 as skewness γ, and the fourth central moment μ_4 as kurtosis κ.

As we have seen in (34), the second moment of the redistributive pool is smaller than the second moment of the component random individual income $y(e_i, \varepsilon_i)$. We would now like to show the same for the skewness, the kurtosis and all higher moments. Given individual i.i.d. skewness $\mu_3 = \gamma$, kurtosis $\mu_4 = \kappa$, and higher moments μ_t, we get the t^{th} moment of the pooled resource M_t as

$$\begin{aligned} \mathsf{M}_t\left[\alpha, n, y\right] &= \mathbb{E}\left[\tfrac{1-\alpha}{2}\left(\tilde{y}_1(e_1)+\varepsilon_1+\tilde{y}_2(e_2)+\varepsilon_2\right)-\tfrac{1-\alpha}{2}\,\mathbb{E}\left[\tilde{y}_1(e_1)+\varepsilon_1+\tilde{y}_2(e_2)+\varepsilon_2\right]\right]^t \\ &= \left(\tfrac{1-\alpha}{2}\right)^t \mathbb{E}[\varepsilon_1+\varepsilon_2]^t \text{ because } \mathbb{E}[\varepsilon_i]=0 \\ &= \left(\tfrac{1-\alpha}{2}\right)^t \textstyle\sum_{k=0}^n C_n^k\,\mathbb{E}[\varepsilon_1^{n-k}\varepsilon_2^k] \\ &= \left(\tfrac{1-\alpha}{2}\right)^t 2\mu_t \\ &= \tfrac{(1-\alpha)^k}{2^{k-1}}\mu_t < \mu_t. \end{aligned} \tag{59}$$

Indeed, if ε_1 and ε_2 are independent, then for every positive function $f(\cdot)$, $f(\varepsilon_1)$ and $f(\varepsilon_2)$ are also independent. (The argument easily generalises to $n > 2$ players.) We argue in the main body of the paper that the variance is a sufficient statistic for elliptical distributions. Hence, the above additional compression of the higher moments represents a generalisation over the set of applicable distributions as each increase in the set of controlled moments enhances sufficiency. As the moment-generating function of a random variable, if it exists, completely specifies its probability distribution, this represents a general argument that the price pool P has lower risk than individual income.

Appendix B: Proofs

Proof of proposition 1. Efficient efforts are extending (2) as the pair (e^*, f^*) solving

$$\begin{aligned} y'(e) = m'(ne-nf) + c_e(e,f),\ m'(ne-nf) \\ = c_f(e,f). \end{aligned} \tag{60}$$

Let $P = (1-\alpha)\sum_{h=1}^n y(e_h)$. Since we are only interested in deviations from symmetric equilibrium, we set $e_j = e_{-i}$. Rewriting (13) for the 2-prize structure $\left(\beta^1, \frac{1-\beta^1}{n-1}, \ldots, \frac{1-\beta^1}{n-1}\right)$ results in

$$\alpha y(e_i) + \beta^1 p_i^1(\mathbf{f})P + \sum_{h=2}^{n} \frac{1-\beta^1}{n-1} p_i^h(\mathbf{f})P - s_i m\left(e_i + (n-1)e_j - f_i - (n-1)f_j\right) - c_e(e_i) - c_f(f_i)$$

which simplifies to

$$\alpha y(e_i) + \beta^1 p_i^1(\mathbf{f})P + \frac{1-\beta^1}{n-1}(1-p_i^1(\mathbf{f}))P - s_i m\left(e_i + (n-1)e_j - f_i - (n-1)f_j\right) - c_e(e_i) - c_f(f_i).$$

The symmetric $e = e_i = e_j$, $f = f_i = f_j$, first-order conditions for this problem are

$$\begin{aligned} c_e'(e) + s_i m'(ne-nf) &= \frac{1-\beta^1+\alpha(n+\beta^1-2)+(1-\alpha)(n\beta^1-1)}{n-1} p(f) y'(e) \\ c_f'(f) &= s_i m'((e-f)n) + \frac{n(1-\alpha)(n\beta^1-1)}{n-1} p'(f) y(e). \end{aligned} \tag{61}$$

Plugging in (60) and imposing $s_i = 1/n$, one obtains

$$\alpha^* = 1 - \frac{y'(e^*) - c_e'(e^*)}{y'(e^*)} \text{ and } \beta^* = \frac{1}{n} + \frac{(n-1)^2 y'(e^*)}{n^3 y(e^*) p'(\mathbf{f}^*)} \tag{62}$$

which can always be achieved by picking a suitably steep ranking technology $p(\mathbf{f}^*)$. □

Proof of proposition 2. Since under our assumptions (13) is fully separable we can split the problem into two independent problems along the respective effort dimensions. Setting $P = (1-\alpha)\sum_{h=1}^{n} y(e_h)$, the two separate problems are

$$\begin{aligned} \alpha y(e_i) + \beta^1 p_i^1(\mathbf{f}^*)P + \frac{1-\beta^1}{n-1}(1-p_i^1(\mathbf{f}^*))P - s_i m\left(e_i + (n-1)e^* - n\,f^*\right) - c_e(e_i) - c_f(f^*), \\ \alpha y(e_i^*) + \beta^1 p_i^1(\mathbf{f})P + \frac{1-\beta^1}{n-1}(1-p_i^1(\mathbf{f}))P - s_i m\left(ne^* - f_i - (n-1)f^*\right) - c_e(e^*) - c_f(f_i). \end{aligned} \tag{63}$$

1) We show that exerting productive effort $e_i = e^*$ gives a global maximum. As players are symmetric and we are looking for a profitable deviation from the efficient level we set $\mathbf{f}^* = (f_1 = f^*, \ldots, f_n = f^*)$ implying that the probability of winning is $p_i^1(\mathbf{f}^*) = 1/n$. Thus the problem simplifies to

$$\alpha y(e_i) + \frac{1}{n}P - s_i m\left(e_i + (n-1)e^* - (n)f^*\right) - c_e(e_i) - c_f(f^*) \tag{64}$$

giving the first-order condition for productive effort e_i as[54]

$$\underbrace{y'(e_i)\left(\alpha + \frac{1}{n}(1-\alpha)\right)}_{\searrow} = \underbrace{s_i m'\left(\max\{0, e_i + (n-1)e_j^* - (n)f^*\}\right)}_{\nearrow} + \underbrace{c'_{e_i}(e_i)}_{\nearrow}.$$

Notice that output is strictly increasing in e_i and is strictly concave. Thus $y''(e_i) < 0$ and $y'(e_i)$ is decreasing. Both cost functions are increasing and convex, therefore $s_i m''(\cdot) + c''(e_i) > 0$ and the rhs is increasing. As $y'(0) > s_i m'(\max\{0, (n-1)e^* - nf^*\}) + c'(0)$,[55] this confirms single crossing of rhs and lhs and ensures the existence of an equilibrium.

2) We now demonstrate global optimality of $f_i = f^*$. Assuming efficient productive effort provision, the first-order condition for reductive effort is

$$\underbrace{n\,y\left(e^*\right)(1-\alpha)\left(\beta\,n-1\right)p'(f_i, f^*)}_{=B} = \underbrace{c'_{f_i}(f_i)}_{=C\nearrow} - \underbrace{s_i m'(\max\{0, ne^* - (n-1)f^* - f_i\})}_{=A\searrow}. \tag{65}$$

Notice that the rhs is strictly increasing as we know that, with respect to f_i, $s_i m''(\cdot) \leq 0$ and thus that A is decreasing and the cost function is convex. Without further assumptions on the monitoring technology $p(\cdot)$ we cannot sign the slope of B. Notice, however, that increasing the slope of the (convex) cost function $c'_{f_i}(f_i)$ sufficiently guarantees single crossing and thus a unique maximum whatever the precise specification of $p(\cdot)$.

3) We now show that (65) identifies a global maximum for the Tullock success function.[56] Again, $s_i m(\max\{0, ne^* - (n-1)f^* - f_i\}) > 0$ for $f_i = 0$ while $p'(f_i, f^*) = 0$ and thus the lhs of (65) is zero at $f_i = 0$ while the rhs is negative. Single crossing is immediate for the case of $r \in (0,1]$ as B is (weakly) decreasing. In the general case of

$$p_i(\mathbf{f}) \;= \frac{f_i^r}{\sum_{j=1}^{n} f_j^r},\; r > 1, \tag{66}$$

[54] It is routine to verify that both first-order conditions identify a maximum.

[55] Since output is concave and the sum of cost functions is convex in e_i, the above inequality holds.

[56] A nearly identical argument can be made for any other ratio-based success function. In that more general case, however, we cannot derive an explicit existence threshold.

the function B has a single critical point and is decreasing when $f_i \geq f^* \left(\frac{(n-1)(r-1)}{r+1}\right)^{1/r}$.

To get single crossing if the two curves are increasing we need to ensure either strict concavity or convexity for the lhs and strict convexity for the rhs and prove that if $f_i = 0$, lhs is larger than the rhs. As we have not specified anything about our functions regarding the third derivative we illustrate this point using the specific $c'_{f_i}(f_i) = bf^{b-1}$ and $s_i m(\max\{0, ne^* - (n-1)f^* - f_i\}) = s_i(\max\{0, ne^* - (n-1)f^* - f_i\})^b$. We also set $s_i = \frac{1}{n}$. We find that both curves have an inflection point, thus we need to find a condition to ensure single crossing.

We first show that the rhs starts out negative and eventually becomes positive as for $f_i = 0$ we have $C - A = -s_i m'(\max\{0, ne^* - (n-1)f^*\}) < 0$. Therefore, as long as the lhs is positive and the rhs negative, the two curves cannot cross. We find that $C - A < 0$ for $f_i < f^* \frac{2}{n^{\frac{1}{b-1}}+1}$ because

$$C - A = f_i^{b-1} b - \frac{(\overbrace{ne^* - (n-1)f^*}^{=2f^*} - f_i)^{b-1} b}{n} = 0 \Leftrightarrow (2f^* - f_i)^{b-1} = n f_i^{b-1}. \tag{67}$$

Moreover, for the rhs, the inflection point occurs when the curve is negative, and it is first concave and then convex. Thus we can conclude than when the curve is above zero, it is strictly convex. We find that $(C - A)'' < 0$ for $f_i < f^* \frac{2}{n^{\frac{1}{b-3}}+1}$ and $f_i < f^* \frac{2}{n^{\frac{1}{b-3}}+1} < f^* \frac{2}{n^{\frac{1}{b-1}}+1}$ because[57]

$$(C - A)'' = f_i^{b-3}(b-2)(b-1)b - \frac{(2f^* - f_i)^{b-3}(b-2)(b-1)b}{n} = 0 \Leftrightarrow (2f^* - f_i)^{b-3} = n f_i^{b-3},$$
$$\Leftrightarrow f^* \frac{2}{n^{\frac{1}{b-1}} + 1} - f^* \frac{2}{n^{\frac{1}{b-3}} + 1} = 2 \frac{f^* \left(n^{\frac{1}{b-3}} - n^{\frac{1}{b-1}}\right)}{\left(n^{\frac{1}{b-1}} + 1\right)\left(n^{\frac{1}{b-3}} + 1\right)} \geq 0.$$

We conclude that the rhs is strictly increasing and convex when it is positive.

For the lhs, there are two inflection points: one in the increasing part and the other in the decreasing part. In the increasing part we find a condition which implies that the inflection occurs if the rhs is negative.[58] A sufficient condition for single-peakedness is therefore that

$$\frac{2^r (f^*)^r}{\left(n^{\frac{1}{b-1}} + 1\right)^r} \geq \underbrace{\frac{(n-1)\left(2(f^*)^r(r^2-1) - \sqrt{3}\sqrt{(f^*)^{2r} r^2 (r^2-1)}\right)}{2 + 3r + r^2}}_{=:\hat{f}}. \tag{68}$$

Thus if the rhs of (65) is positive, it is also strictly convex. If (68) is respected, the lhs is strictly concave or convex. Notice also that at the inflection point, the rhs is positive and the lhs is negative and therefore the lhs is larger than the rhs. The geometric intuition of (68) is shown in figure 4 for the setup of the example section 2.3. The figure shows a family of curves B for $r \in \{1, 2, 4, 10, 11\}$ with inflection points labelled $\hat{f}_2$, $\hat{f}_4$, $\hat{f}_{10}$, and $\hat{f}_{11}$, respectively. Condition (68) is fulfilled as long as the red cost curve $C - A$ is negative at the respective inflection point. This is true for $r = 2$ and $r = 4$ soon after which (68) starts failing. Uniqueness, however, is actually only lost for $r > 10$. □

[57] This is true for any $b \geq 3$.

[58] The inflection point in the decreasing part does not matter. As long as one curve is increasing and the other is decreasing they can only cross once.

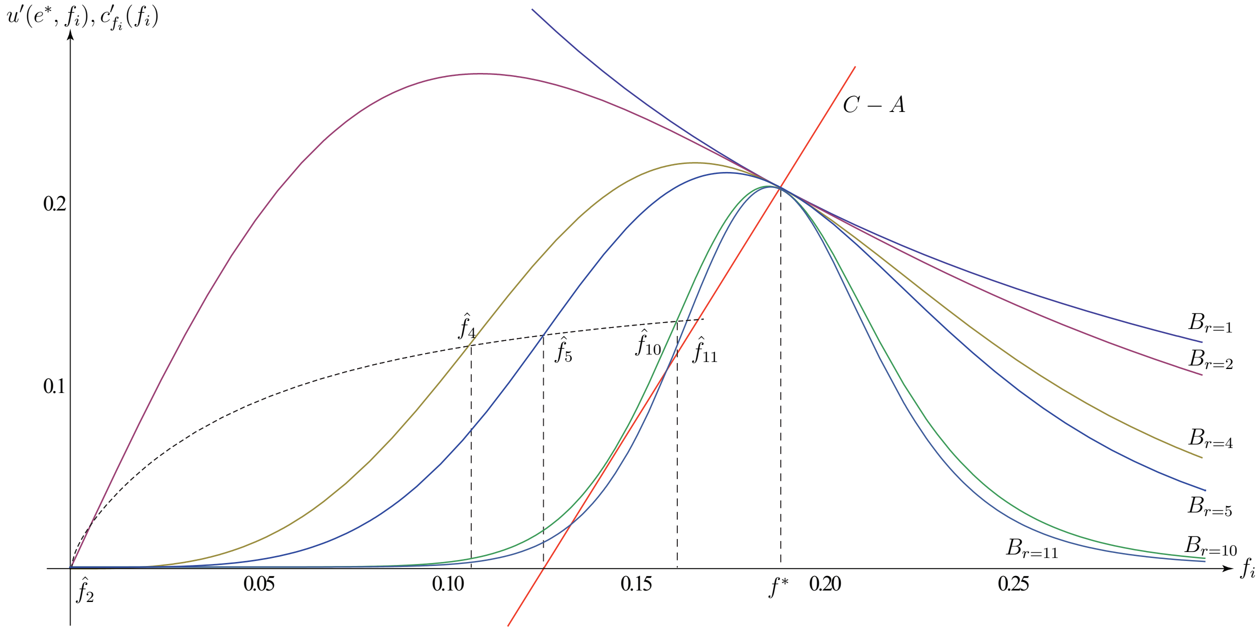


Figure 4: Single crossing in equation (65) ensures a unique global maximum at $f_i = f^*$ for the example setup of section 2.3. The dotted line gives the location of inflection points $\hat{f}$ for different r.

Proof of proposition 3. Player i's equilibrium participation utility for $P = (1-\alpha)ny(e^*)$ under the efficiency-inducing contract $C = \langle \alpha^*, \beta^*; p^*(\mathbf{f}) \rangle$ defined in (13) with prizes $\left(\beta, \frac{1-\beta}{n-1}, \ldots, \frac{1-\beta}{n-1}\right)$ is

$$\begin{aligned} u_i(e^*, f^*) &= \alpha^* y(e^*) + \tfrac{1}{n}\beta^* P + (n-1)\tfrac{1}{n}\tfrac{1-\beta^*}{n-1} P - s_i m\left(ne^* - nf^*\right) - c_e(e^*) - c_f(f^*) \\ &= \alpha y(e^*) + \tfrac{1}{n}(1-\alpha^*) ny(e^*) - s_i m\left(ne^* - nf^*\right) - c_e(e^*) - c_f(f^*) \\ &= y(e^*) - s_i m\left(ne^* - nf^*\right) - c_e(e^*) - c_f(f^*). \end{aligned} \tag{69}$$

Now consider the contract $C' = \langle \alpha', \beta'; p' \rangle$. Since second-stage efforts $e(\alpha', \beta', p'), f(\alpha', \beta', p')$ are continuous in α', it is sufficient to consider the extreme case of $\alpha' = 1$ which implements no contest at all. A deserter's utility when subjected to C' can therefore be driven down to[59]

$$u_i^s(e_i^s, f_i^s) = y(e_i^s) - s_i m(e_i^s + (n-1)e^a - f_i^s - (n-1)f^a) - c_e(e_i^s) - c_f(f_i^s) \tag{70}$$

in which e^a and f^a are the equilibrium inside agreement efforts prescribed by C'. For $\alpha' = 1$, these equal the equilibrium efforts without agreement e_i^s, f_i^s. Hence (69) and (70) are identical but implement different efforts. Since both $e^s > e^*$ and $f^s < f^*$, it is therefore individually rational to join the agreement implementing C if the alternative is C' because the cost differential

$$s_i(m(ne^s - nf^s) - m\left(ne^* - nf^*\right)) \tag{71}$$

is increasing in n and convex. As the efficient allocation is welfare maximising, this outweighs any productivity gains from free-riding $y(e^s) - y(e^*) + c_e(e^*) + c_f(f^*) - c_e(e^s) - c_f(f^s)$. Thus, every player finds it individually rational to join the reductive contest if threatened by the alternative C'. □

[59] Notice that the latter formulation requires $n > 2$ as the contest can only produce incentives if at least two players participate in the contest.

Proof of proposition 4. Since each individual output $y(e_i, \varepsilon_i)$ has variance σ^2 and covariance of two distinct variables σ_{ij}, the variance of the prize pool is (for equilibrium α)

$$\begin{aligned}
\mathbb{V}[u_i(\mathbf{e},\mathbf{f};\boldsymbol{\varepsilon})] &= \mathbb{V}[\alpha y(e_i,\varepsilon_i)] + \mathbb{V}\left[(1-\alpha)\frac{1}{n}\sum_{i=1}^{n} y(e_i,\varepsilon_i)\right] + 2\alpha(1-\alpha)\frac{1}{n}\operatorname{Cov}\left(y(e_i,\varepsilon_i),\sum_{i=1}^{n} y(e_i,\varepsilon_i)\right)\\
&= \alpha^2\,\mathbb{V}[y(e_i,\varepsilon_i)] + (1-\alpha)^2\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\left[y(e_i,\varepsilon_i)\right] + 2\alpha(1-\alpha)\frac{1}{n}\operatorname{Cov}\left(y(e_i,\varepsilon_i),\sum_{i=1}^{n} y(e_i,\varepsilon_i)\right)\\
&+ 2\frac{(1-\alpha)^2}{n^2}\operatorname{Cov}\left(\sum_{i=1}^{n} y(e_i,\varepsilon_i),\sum_{j=1}^{n} y(e_j+\varepsilon_j)\right)\\
&= \sigma^2\left[\alpha^2 + (1-\alpha)^2\frac{1}{n}\right] + \alpha(1-\alpha)\frac{1}{n}\sum_{i\neq j}\sigma_{ij} + 2\frac{(1-\alpha)^2}{n}\sum_{1\le i<j\le n}\sigma_{ij}\\
&\le \sigma^2\left[\alpha^2 + (1-\alpha)^2\frac{1}{n}\right] + 2\alpha(1-\alpha)\frac{n-1}{n}\sigma^2 + \frac{(1-\alpha)^2}{n}n(n-1)\sigma^2\\
&= \frac{\sigma^2}{n}\left[2\alpha^2 - 2\alpha + n\right]\\
&< \sigma^2 \;\; \forall\alpha\in(0,1).
\end{aligned}$$

□

Proof of proposition 5. Let us first consider independent shocks. For equilibrium α_i, β_i,

$$\begin{aligned}
\mathbb{V}[u_i(\mathbf{e}^*,\mathbf{f}^*);\boldsymbol{\varepsilon}] &= \mathbb{V}[\alpha_i y(e_i^*,\varepsilon_i) + p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i[(1-\alpha_i)y(e_i^*,\varepsilon_i)\\
&\quad + (1-\alpha_j)y(e_j^*,\varepsilon_j)] + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j)[(1-\alpha_i)y(e_i^*,\varepsilon_i) + (1-\alpha_j)y(e_j^*,\varepsilon_j)]\\
&= [\alpha_i + (p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j))(1-\alpha_i)]^2\sigma^2\\
&\quad + [p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j)]^2(1-\alpha_j)^2\sigma^2\\
&= \alpha_i^2\sigma^2 + 2\alpha_i(1-\alpha_i)[p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j)]\sigma^2\\
&\quad + [p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j)]^2((1-\alpha_i)^2 + (1-\alpha_j)^2)\sigma^2\\
&\le \alpha_i^2\sigma^2 + 2\alpha_i(1-\alpha_i)\beta_i\sigma^2 + ((1-\alpha_i)^2 + (1-\alpha_j)^2)[p_i(\mathbf{e}^*,\mathbf{f}^*)^2(-1+\beta_i+\beta_j)^2\\
&\quad + 2p_i(\mathbf{e}^*,\mathbf{f}^*)(1-\beta_j)(-1+\beta_i+\beta_j) + (1-\beta_j)^2]\sigma^2\\
&\le \{\alpha_i^2 + 2\alpha_i(1-\alpha_i)\beta_i + ((1-\alpha_i)^2 + (1-\alpha_j)^2)\beta_i^2\}\sigma^2.
\end{aligned}$$

Then, $\mathbb{V}[u_i(\mathbf{e}^*,\mathbf{f}^*);\boldsymbol{\varepsilon}] < \sigma^2$ for all i if and only if

$$1 > \beta_i^2(1-\alpha_j)^2 + (\beta_i - \alpha_i\beta_i - \alpha_i)^2 + 4\alpha_i\beta_i(1-\alpha_i). \tag{72}$$

Under non-independent shocks, the covariance term has to be added

$$\begin{aligned}
&\operatorname{Cov}([\alpha_i + A(1-\alpha_i)]y(e_i^*,\varepsilon_i), A(1-\alpha_j)y(e_j^*,\varepsilon_j))\\
&= [\alpha_i + A(1-\alpha_i)]A(1-\alpha_j)\operatorname{Cov}[y(e_i^*,\varepsilon_i), y(e_j^*,\varepsilon_j)] && (73)\\
&\le [\alpha_i + \beta_i(1-\alpha_i)]\beta_i(1-\alpha_j)\sigma_{ij} && (74)\\
&\le [\alpha_i + \beta_i(1-\alpha_i)]\beta_i(1-\alpha_j)\sigma^2
\end{aligned}$$

with $A = [p_i(\mathbf{e}^*,\mathbf{f}^*)\beta_i + (1-p_i(\mathbf{e}^*,\mathbf{f}^*))(1-\beta_j)]$. As for the independent shocks, we use the fact

$A \leq \beta_i$ (equations (73) & (74)). Then, using the condition (72) for independent shocks we get

$$1 > \beta_i(3 - 2\alpha_i - \alpha_j)(2\alpha_i + \beta_i - \beta_i\alpha_j) + (\beta_i - \alpha_i\beta_i - \alpha_i)^2 \tag{75}$$

which establishes our claim. □

Appendix C: Stochastic Dominance Relations

This appendix provides analytical tools and new results on stochastic dominance which are useful to establish proposition 6.

Definition 1 (Shaked and Shanthikumar (1994))**.** *Let X and Y be two random variables. X is said to be smaller than Y in the increasing concave order and denoted $X \leq_{icv} Y$ if and only if*

$$\mathbb{E}[v(X)] \leq \mathbb{E}[v(Y)] \tag{76}$$

for all increasing, concave functions $v(\cdot)$.

In the economics literature, increasing concave orders are typically referred to as *second order stochastic dominance*. Similarly, increasing concave orders with equal means are usually called *mean preserving spreads*. An alternative, equivalent definition is:

Definition 2 (Shaked and Shanthikumar (1994))**.** *Let X and Y be two random variables with F_X and F_Y their continuous cumulative functions. $X \leq_{icv} Y$ if and only if*

$$\int_{-\infty}^{z} F_X(t)dt \geq \int_{-\infty}^{z} F_Y(t)dt. \tag{77}$$

The following result generalises theorem 5 of Müller (2001) for univariate elliptical distributions and increasing concave orders. The proof is provided for completeness and is an adaptation of that in Müller (2001).

Theorem 7. *Let $X \sim \mathcal{L}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{L}(\mu_y, \sigma_y^2)$, which are two elliptically distributed random variables. If $\sigma_y^2 \leq \sigma_x^2$ and $\mu_y \leq \mu_x$ then $X \leq_{icv} Y$.*

Proof. Given the properties of elliptical distributions, the distribution of X is equal to that of $Y + Z$ with $Z \sim \mathcal{L}(\mu_x - \mu_y, \sigma_x^2 - \sigma_y^2)$.[60] Then,

$$\begin{aligned} \mathbb{E}[v(X)] &= \mathbb{E}[v(Y+Z)] \\ &= \mathbb{E}[\mathbb{E}[v(Y+Z)/Y]] && (78) \\ &\leq \mathbb{E}[v(Y + \mathbb{E}[Z])] && (79) \\ &\leq \mathbb{E}[v(Y)]. && (80) \end{aligned}$$

The Jensen inequality for concave functions leads to (79). Then, using $\mathbb{E}[Z] \leq 0$ (80) follows. □

[60] See Fang, Kotz, and Ng (1987) or Landsman and Valdez (2003) for details.

Theorem 8. *Let $X \sim \mathcal{L}(\mu_x, \sigma_x^2)$ and $Y \sim \mathcal{L}(\mu_y, \sigma_y^2)$, which are two elliptically distributed random variables such that $\mu_y = \mu_x = 0$. Then the following statements are equivalent: (i) $X \leq_{\text{icv}} Y$, and (ii) $\sigma_y^2 \leq \sigma_x^2$.*

Proof. Every probability density $(F_i)'$ is symmetric and continuous. Zero mean implies that the symmetry of the probability density is with respect to the ordinate. Then, every cumulative distribution has an inflection point at zero. Furthermore, X has a probability density with bigger tails than Y because of the ranking of the variance. Consequently of the tails and the inflection point, the cumulative distribution of X is above (below) the one of Y for all value inferior (superior) to zero such that the surfaces between the two cumulative distributions before and after zero are equal. Then,

$$\int_{-\infty}^{z} F_X(t)dt \geq \int_{-\infty}^{z} F_Y(t)dt \qquad (81)$$

which implies that $X \leq_{\text{icv}} Y$. Notice that this implication is true by application of Theorem 7.

Moreover, it is well-known that if $X \leq_{\text{icv}} Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$ then $\sigma_y^2 \leq \sigma_x^2$ follows (see, for instance, Rothschild and Stiglitz (1970)). □

References

Aldy, J. E., and R. N. Stavins (2007): *Architectures for Agreement*. Cambridge University Press, New York.

Allison, I., N. L. Bindoff, R. Bindschadler, P. Cox, N. de Noblet, M. England, J. Francis, N. Gruber, A. Haywood, D. Karoly, G. Kaser, C. L. Quéré, T. Lenton, M. Mann, B. McNeil, A. Pitman, S. Rahmstorf, E. Rignot, H. Schellnhuber, S. Schneider, S. Sherwood, R. Somerville, K.Steffen, E. Steig, M. Visbeck, and A. Weaver (2009): *The Copenhagen Diagnosis, 2009: Updating the world on the Latest Climate Science*. The University of New South Wales, Climate Change Research Centre (CCRC), Sydney, Australia.

Arnott, R., and J. E. Stiglitz (1991): "Moral Hazard and Nonmarket Institutions: Dysfunctional Crowding Out of Peer Monitoring?," *American Economic Review*, 81, 179–90.

Banks, J., R. Blundell, and A. Brugiavini (2001): "Risk Pooling, Precautionary Saving and Consumption Growth," *Review of Economic Studies*, 68, 757–79.

Barrett, S. (1994): "Self-Enforcing International Environmental Agreements," *Oxford Economic Papers*, 46, 878–94.

——— (1998): "Political Economy of the Kyoto Protocol," *Oxford Review of Economic Policy*, 14(4), 20–39.

——— (2003): *Environment and statecraft: The strategy of environmental treaty-making*. Oxford University Press, New York.

——— (2006): "Climate Treaties and 'Breakthrough' Technologies," *American Economic Review*, 96, 22–5.

Beccherle, J., and J. Tirole (2011): "Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements," *Journal of Public Economics*, 95, 1339–48.

Boadway, R., Z. Song, and J.-F. Tremblay (2011): "The efficiency of voluntary pollution abatement when countries can commit," *European Journal of Political Economy*, 27, 352–68.

Bosetti, V., C. Carraro, E. De Cian, E. Massetti, and M. Tavoni (2013): "Incentives and Stability of International Climate Coalitions: An Integrated Assessment," *Energy Policy*, 55, 44–56.

Breton, M., L. Sbragia, and G. Zaccour (2010): "Dynamic Models for International Environmental Agreements," *Environmental and Resource Economics*, 45, 25–48.

Buchholz, W., R. Cornes, and D. Rübbelke (2011): "Interior matching equilibria in a public good economy: An aggregative game approach," *Journal of Public Economics*, 95, 639–45.

Buchholz, W., and K. Konrad (1994): "Global Environmental Problems and the Strategic Choice of Technology," *Journal of Economcs*, 60(3), 299–321.

Buob, S., and G. Stephan (2011): "To mitigate or to adapt: How to confront global climate change," *European Journal of Political Economy*, 27, 1–16.

Carraro, C., C. Marchiori, and S. Oreffice (2009): "Endogenous Minimum Participation in International Environmental Treaties," *Environmental & Resource Economics*, 42(3), 411–425.

Chamberlain, G. (1983): "A characterisation of the distributions that imply mean-variance utility functions," *Journal of Economic Theory*, 29, 185–201.

Chander, P. (2007): "The gamma-core and coalition formation," *International Journal of Game Theory*, 35(4), 539–56.

Chander, P., and H. Tulkens (1995): "A core-theoretic solution for the design of cooperative agreements on transfrontier pollution," *International Tax and Public Fincance*, 2, 279–93.

——— (2009): "Cooperation, stability and self-enforcement in international environmental agreements: A conceptual discussion," in *The Design of Climate Policy*, ed. by R. Guesnerie, and H. Tulkens, CESifo Seminar Series. MIT Press, Cambridge, Mass.

Che, Y.-K. C., and I. Gale (2000): "Difference-form contests and the robustness of all-pay auctions," *Games and Economic Behavior*, 30, 22–43.

Dasgupta, S., B. Laplante, C. Meisner, D. Wheeler, and J. Yan (2009): "The impact of sea level rise on developing countries: a comparative analysis," *Climatic Change*, 93, 379–88.

Deaton, A. (1991): "Saving and liquidity constraints," *Econometrica*, 59, 1221–48.

Demange, G. (2008): "Sharing aggregate risks under moral hazard," *Paris-Jourdan Sciences Economiques, Laboratoire d'Économie Appliquée*, Working Paper #2008-27.

Diamantoudi, E., and E. S. Sartzetakis (2006): "Stable International Environmental Agreements: An Analytical Approach," *Journal of Public Economic Theory*, 8, 247–63.

Dijkstra, B. R. (2007): "An investment contest to influence environmental policy," *Resource and Energy Economics*, 29(4), 300–24.

DIXIT, A. (1987): "Strategic Behavior in Contests," *American Economic Review*, 77, 891–98.

ENGEL, S. (2004): "Achieving environmental goals in a world of trade and hidden action: the role of trade policies and eco-labeling," *Journal of Environmental Economics and Management*, 48(3), 1122–45.

FANG, K., S. KOTZ, AND K. NG (1987): *Symmetric Multivariate and Related Distributions*. London: Chapman & Hall.

FRANKE, J. (2012): "Affirmative action in contest games," *European Journal of Political Economy*, 28, 105–18.

FRANKE, J., C. KANZOW, W. LEININGER, AND A. SCHWARTZ (2011): "Effort maximization in asymmetric contest games with heterogeneous contestants," *Economic Theory*, 52(2), 589–630.

FU, Q., AND J. LU (2012): "Micro foundations of multi-prize lottery contests: a perspective of noisy performance ranking," *Social Choice and Welfare*, 38, 497–517.

GERBER, A., AND P. C. WICHARDT (2009): "Providing public goods in the absence of strong institutions," *Journal of Public Economics*, 93(34), 429–39.

GERSBACH, H., AND R. WINKLER (2011): "International emission permit markets with refunding," *European Economic Review*, 55(6), 759–73.

——— (2012): "Global refunding and climate change," *Journal of Economic Dynamics and Control*, 36, 1775–95.

GERSHKOV, A., J. LI, AND P. SCHWEINZER (2009): "Efficient Tournaments within Teams," *Rand Journal of Economics*, 40(1), 103–19.

GIEBE, T., AND P. SCHWEINZER (2013): "Consuming your way to efficiency: public goods provision through non-distortionary tax lotteries," *CESifo Working Paper*, #4228.

GRANKVIST, G., U. DAHISTRAND, AND A. BIEL (2004): "The Impact of Environmental Labelling on Consumer Preference: Negative vs Positive Labels," *Journal of Consumer Policy*, 27(2), 213–30.

GREEN, E. J. (1987): "Lending and the smoothing of uninsurable Income," in *Contractual Arrangements for Intertemporal Trade*, ed. by E. Prescott, and N. Wallace, pp. 3–25. University of Minnesota Press, Minneapolis, MN.

GREEN, J. R., AND N. STOKEY (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349–64.

GUESNERIE, R., AND H. TULKENS (eds.) (2009): *The Design of Climate Policy*, CESifo Seminar Series. MIT Press, Cambridge, Mass.

HARBAUGH, R., J. W. MAXWELL, AND B. ROUSSILLON (2011): "Label Confusion: The Groucho Effect of Uncertain Standards," *Management Science*, 57, 1512–27.

HARSTAD, B. (2010): "The Dynamics of Climate Agreements," *Northwestern University, Center for Mathematical Studies in Economics and Management Science*, Discussion paper #1474.

HEAYES, A. G. (1997): "Environmental Regulation by Private Contest," *Journal of Public Economics*, 63, 407–28.

HURLEY, T. M., AND J. F. SHOGREN (1997): "Environmental Conflicts and the SLAPP," *Journal of Environmental Economics and Management*, 33(3), 253–73.

JIA, H. (2008): "A stochastic derivation of the ratio form of contest success functions," *Public Choice*, 135, 125–30.

KNIVETON, D. R., C. D. SMITH, AND R. BLACK (2012): "Emerging migration flows in a changing climate in dryland Africa," *Nature Climate Change*, 2, 444–7.

KOLSTAD, C. D. (2007): "Systematic uncertainty in self-enforcing international environmental agreements," *Journal of Environmental Economics and Management*, 53(1), 68 – 79.

KONRAD, K. (2008): *Strategy and Dynamics in Contests*. Oxford University Press, Oxford.

KOTCHEN, M. J., AND S. W. SALANT (2011): "A free lunch in the commons," *Journal of Environmental Economics and Management*, 61, 245–53.

LANDSMAN, Z., AND E. VALDEZ (2003): "Tail Conditional Expectations for Elliptical Distributions," *North American Actuarial Journal*, 7(4), 55–71.

LAZEAR, E., AND S. ROSEN (1981): "Rank Order Tournaments as Optimal Labor Contracts," *Journal of Political Economy*, 89, 841–64.

LERNER, J., AND J. TIROLE (2006): "A Model of Forum Shopping, with Special Reference to Standard Setting Organizations," *American Economic Review*, 96, 1091–13.

MITROVICA, J. X., N. GOMEZ, AND P. U. CLARK (2009): "The Sea-Level Fingerprint of West Antarctic Collapse," *Science*, 323(5915), 753.

MOLDOVANU, B., AND A. SELA (2001): "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91(3), 542–58.

MONTERO, J.-P. (2008): "A Simple Auction Mechanism for the Optimal Allocation of the Commons," *American Economic Review*, 98, 496–518.

MORGAN, P. (2000): "Financing Public Goods by Means of Lotteries," *Review of Economic Studies*, 67, 761–84.

MÜLLER, A. (2001): "Stochastic Ordering of Multivariate Normal Distributions," *Annals of the Institute of Statistical Mathematics*, 53(3), 567–75.

NALEBUFF, B. J., AND J. E. STIGLITZ (1983): "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *Bell Journal of Economics*, 14, 21–43.

NICHOLLS, R. J. (1995): "Synthesis of vulnerability analysis studies," in *Proceedings of WORLD COAST 1993*, ed. by Ministry of Transport, Public Works and Water Management, the Netherlands, pp. 181–216.

NICHOLLS, R. J., S. HANSON, C. HERWEIJER, N. PATMORE, S. HALLEGATTE, J. CORFEE-MORLOT, J. CHATEAU, AND R. MUIR-WOOD (2008): "Ranking Port Cities with High Exposure and Vulnerability to Climate Extremes: Exposure Estimates," *OECD Environment Working Papers*, #1.

NORDHAUS, W. D. (2006): "After Kyoto: Alternative Mechanisms to Control Global Warming," *American Economic Review*, 96, 31–4.

Owen, J., and R. Rabinovitch (1983): "On the class of elliptical distributions and their applications to the theory of portfolio choice," *Journal of Finance*, 38, 745–52.

Polanski, A., and E. Stoja (2010): "Incorporating Higher Moments into Value at Risk Forecasting," *Journal of Forecasting*, 29, 523–35.

Rothschild, M., and J. Stiglitz (1970): "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225–43.

Schweinzer, P., and E. Segev (2012): "The optimal prize structure of symmetric Tullock contests," *Public Choice*, 153, 69–82.

Shaked, M., and J. Shanthikumar (1994): *Stochastic Orders and Their Applications*. Probability and Mathematical Statistics, San Diego: Academic Press.

Siegel, R. (2009): "Asymmetric Contests with Conditional Investments," *American Economic Review*, 100, 2230–60.

Skaperdas, S. (1996): "Contest Success Functions," *Economic Theory*, 7(2), 283–90.

Stern, N. (2006): *Stern review on the economics of climate change*. HM Treasury, London, UK.

Wildasin, D. E. (1995): "Factor Mobility, Risk and Redistribution in the Welfare State," *Scandinavian Journal of Economics*, 97, 527–46.

Wilson, R. (1968): "The Theory of Syndicates," *Econometrica*, 36(1), 119–22.