

Measures to enhance the success of global climate treaties

Johan Eyckmans · Michael Finus

Received: 30 May 2006 / Accepted: 23 November 2006 /

Published online: 3 February 2007

© Springer Science+Business Media B.V. 2007

Abstract We analyze important forces that hamper the formation of successful self-enforcing agreements to mitigate global warming from an economic point of view. The analysis combines two modules: (a) an integrated assessment model that captures the feedback between the economy, environmental damages and the climate system and (b) a game theoretic model that determines stable coalitions in the presence of free-riding incentives. We consider two types of measures to enhance the success of international environmental treaty-making: (a) transfers, aiming at balancing asymmetric gains from cooperation; (b) institutional changes, aiming at making it more difficult to upset stability of a treaty. We find that institutional changes may be as important as transfers and should therefore receive more attention in future international negotiations.

Keywords Coalition formation · Design of climate treaty protocol · Integrated assessment model · Non-cooperative game theory

JEL Classification C68 · C72 · H41 · Q25

J. Eyckmans
EHSAL, Europese Hogeschool Brussel,
Stormstraat 2, 1000 Brussels, Belgium
e-mail: Johan.Eyckmans@ehsal.be

J. Eyckmans
Centrum voor Economische Studiën, K.U. Leuven,
Leuven, Belgium

M. Finus (✉)
Department of Economics, University of Hagen,
Profilstr. 8, 58084 Hagen, Germany
e-mail: Michael.Finus@fernuni-hagen.de

1 Introduction

The voluntary provision of public goods is a well-known problem in economics (e.g., Cornes & Sandler, 1996; Endres, 1997). Non-excludability from positive externalities leads to the underprovision by private entities. In the national context, governments can mitigate this problem. They can provide the appropriate level of public goods with financial resources from taxation. However, in the international context, this is more difficult because no “world government” exists that can take on this role. Consequently, international treaties have to rely on voluntary participation and must be designed in a self-enforcing way. In the presence of free-rider incentives, this frequently means that not all countries participate in an international environmental treaty and/or the agreed level of public provision only marginally exceeds non-cooperative levels (e.g., Böhringer & Vogt, 2004; Finus & Tjøtta, 2003; Murdoch & Sandler, 1997a, b). The mitigation of global warming exemplifies this problem. In 1997, the Kyoto Protocol was signed during the third Conference of the Parties (CoP) of the United Nations Framework Convention on Climate Change (UNFCCC). 39 signatories of the Protocol listed in the so-called Annex-B, responsible for about two thirds of global emissions in 1990, promised to reduce their greenhouse gas emissions by 5.2 percent compared to 1990 levels by 2008–2012. However, in 2001, the USA withdrew from the Protocol and in the aftermaths of this decision, also other signatories started demanding to reduce their previously accepted moderate abatement targets even further. Only now, eight years after the conference in Kyoto, the Kyoto Protocol has finally entered into force after the ratification of Russia. Now, the share of the ratifying Annex-B countries in global 1990 emissions amounts to about one half and is projected to decrease to about one third by 2012 because of fast growth of emissions in non-Annex-B countries.

From the bumpy road towards an international climate agreement, it is evident that there are some fundamental characteristics associated with climate change that makes this environmental problem even more difficult to solve than other trans-boundary pollution problems. Therefore, this paper has two purposes: (a) shedding light on some fundamental forces that hamper successful treaty-making in the context of greenhouse gas mitigation and (b) considering measures to improve the success of self-enforcing climate treaties. For this purpose, we combine two economic modules in our analysis: (a) an integrated assessment model that captures the feedback between the economy, environmental damages and the climate system and (b) a non-cooperative game theoretic model that determines stable coalitions in the presence of free-riding incentives. That is, though the integrated assessment model is the same as used in Eyckmans and Tulkens (2003), the game theoretic analysis is different. In order to explain the difficulties of forming effective IEAs, we have to depart from their cooperative game theoretic framework where we will stress differences in subsequent sections.

The first measure that we consider to enhance the success of global climate treaties is transfers aiming at balancing strong asymmetries between the actors involved in climate change. We consider eight transfer schemes. They are related to the literature that pays much attention to the philosophical and moral motivation of various schemes (Rose & Stevens, 1998; Rose, Stevens, Edmonds, & Wise, 1998; Stevens & Rose, 2002). However, most of these papers consider only welfare effects

in terms of abatement costs, giving only an incomplete picture of the impacts from cooperation, letting alone the incentives to form self-enforcing agreements. But also the game theoretic oriented literature is limited in its application to the issues that we deal in this paper. Most papers assume symmetric players, rendering the analysis of transfers uninteresting (e.g., Carraro & Siniscalco, 1993; Barrett, 1994). Those papers that consider asymmetric players either consider a very particular form of asymmetry (e.g., two types of countries) or base their analysis on a climate model that only captures the dynamics of climate change in a very simplistic way (e.g., Barrett, 1997, 2001; Botteon & Carraro, 1997; Chander & Tulkens, 1997; Hoel & Schneider, 1997; Weikard, Finus, & Altamirano-Cabrera, 2006). Moreover, most papers consider only a small portfolio of transfer rules and these rules are mainly related to stylized solutions of bargaining theory (e.g., Barrett, 1997; Botteon & Carraro, 1997; Germain, Toint, & Tulkens, 1998).¹ Therefore, it is one of the main contributions of this paper to analyze a large variety of transfer schemes, studying in a systematic way their effects on coalition formation and stability based on a fully fledged assessment model with strongly asymmetric players.

The second measure is a change of the institutional rules aiming at making it more difficult to upset stability of a treaty. Therefore, we contrast open membership which is typical for public goods with exclusive membership which is typical for club goods. For this purpose, we do not test stability with the concept of the core as in Eyckmans and Tulkens (2003), but consider a modification of the concept of internal and external stability by d'Aspremont, Jacquemin, Gabszewicz, and Weymark (1983) which has been frequently applied in the non-cooperative game theoretic analysis of international environmental agreements (e.g., Barrett, 1994; Carraro & Siniscalco, 1993; Hoel, 1992). We are inspired by the recent literature in this field, pointing out that exclusive membership may be conducive to the stability of treaties (e.g., Carraro, 2000; Finus, 2003; Finus & Rundshagen, 2003). However, so far, no sound evidence could be presented on the impact of “more stability” on welfare and the environment as this requires simulations and a departure from the assumption of symmetric players as well as a static payoff structure.

In what follows, we describe the integrated assessment model in Sect. 2 and discuss fundamental features of coalition formation in Sect. 3. Subsequently, we analyze stability of coalitions and measures to improve upon the success of climate change treaties in Sect. 4. Section 5 raps up the analysis and points to future research questions.

It should be pointed out at the outset and as will be discussed in much detail Sect. 5 that we do not claim that our model reflects all aspects of the climate problem. We only claim that our model constitutes a step forward in the economics literature but makes many simplifying assumptions. These assumptions are critically reviewed and evaluated in Sect. 5. Nevertheless, we believe that our simulation exercise illustrates some institutional aspects of treaty design neglected so far.

2 Integrated Assessment Model

The CLIMNEG World Simulation Model (in the sequel referred to as CWSM) is an integrated assessment, economy-climate model that resembles closely the seminal

¹ A notable exception is for instance Bosello, Buchner, and Carraro (2003).

RICE model by Nordhaus and Yang (1996).² It captures the endogenous feedback of climate change damages on production and consumption possibilities. The economic part of CWSM consists of a longterm dynamic perfect foresight Ramsey type of optimal growth model with endogenous investment and carbon emission reduction decisions. The environmental part consists of a carbon cycle and temperature change module.

In the CWSM, the world is divided into six regions: *USA*, *JPN* (Japan), *EU* (European Union), *CHN* (China), *FSU* (Former Soviet Union) and *ROW* (Rest of the World). In every region i , and at every time t , the following budget equation describes how “potential GDP”, $Y_{i,t}$, can be “allocated” to consumption, $Z_{i,t}$, investment, $I_{i,t}$, emission abatement costs, $Y_{i,t} C_i(\mu_{i,t})$, and climate change damages, $Y_{i,t} D_i(\Delta T_t)$:

$$Y_{i,t} = Z_{i,t} + I_{i,t} + Y_{i,t} C_i(\mu_{i,t}) + Y_{i,t} D_i(\Delta T_t) \quad (1)$$

Output $Y_{i,t}$ is produced with capital and labor. Capital is build up through investment and depreciates at some fixed rate. Labour supply is assumed to be inelastic. Therefore, investment $I_{i,t}$ is the only endogenous production input and constitutes the first choice variable in the model.

Abatement costs $Y_{i,t} C_i(\mu_{i,t})$ are expressed as “loss of potential GDP”: C_i is the share of “potential GDP” devoted to abatement which is a function of $\mu_{i,t} \in [0, 1]$, measuring the relative emission reduction compared to the business-as-usual scenario without any abatement policy. The abatement cost functions are strictly convex in abatement. The specific functional form and parameter values that underly CWSM (see the Appendix) are taken from Nordhaus and Yang (1996). This is in line with the assumptions and estimates in the volume edited by Weyant and Hill (1999) but the functions are steeper than the more recent estimates used for instance in the model FUND as reported by Kemfert and Tol (2002). Damages $Y_{i,t} D_i(\Delta T_t)$ are also expressed as “loss of potential GDP”: D_i is the share of “potential GDP” destroyed by climate change damages which is a function of temperature change ΔT_t . Also damages are assumed to be strictly convex in temperature change. The functional form and the parameter values (see the Appendix) are taken from Nordhaus and Yang (1996). These estimates are in line with Pearce et al. (1996) but are relatively pessimistic (i.e. high) compared to the estimates in Tol (2005). Temperature change depends on the stock of greenhouse gases which in turn depends on emissions that accumulate in the atmosphere. Finally, emissions depend on emissions released through production minus abatement. Hence, the second choice variable in this model is the abatement level $\mu_{i,t}$.

Both choice variables (investment and abatement) affect output, abatement costs, damage costs and therefore also consumption not only domestically but also abroad. For abatement, this is immediately evident since remaining emissions (after abatement) increases the stock of greenhouse gases which affects environmental damages in every country. However, this is also true for investment since capital is an input in the production process and atmospheric emissions are proportional to production. Technological progress is captured by CWSM but is only exogenous where the time

² An overview of the equations and parameters of the model is provided in the Appendix. A more detailed exposition of the model can be found in Eyckmans and Tulkens (2003).

path is taken from RICE. It increases the production possibilities and decreases the emission output ratio (i.e., increases energy efficiency) over time.

We measure welfare of a region i as total lifetime discounted consumption:

$$W_i(s) = \sum_{t=0}^{\Omega} \frac{Z_{i,t}}{[1 + \rho_i]^t} \quad (2)$$

where ρ_i stands for the discount rate of region i , Ω denotes the time horizon and s is a strategy vector. Vector $s = \{I_{i,t}, \mu_{i,t}\}_{i \in N, t=0, \dots, \Omega}$ consists of a time path of 35 decades³ for emission abatement and investment for all six regions and hence is of length $2 \times 35 \times 6 = 420$.

From the input data displayed in the Appendix, it is evident that we assume a relatively low discount rate of 1.5 percent, except for CHN and ROW where we assume 3 percent in order not to “overestimate” the incentives of these regions in climate change policies. These discount rates seem roughly in line with Weitzman (2001) who suggests that if a constant discount rate has to be chosen in the context of global warming, then a discount rate of 2 percent or less is appropriate to capture long-term effects. Moreover, the parameters of the model imply that USA, JPN and EU face steep and CHN and ROW flat marginal abatement costs. The regional differences in abatement costs mainly reflect differences in energy efficiency. That is, energy efficient regions face higher costs when cutting back emissions than regions characterized by low energy efficiency. Finally, damage functions are particularly steep in EU and ROW, to a lesser extent in USA and JPN and relatively flat in FSU and CHN. The high damage estimate (as a percentage of “potential GDP”) for ROW is due to the fact that climate change is believed to affect developing countries more strongly than industrialized countries because their economies tend to depend more on climate related production processes like agriculture, fishery and forestry (IPCC 2001). The low damage estimate for FSU is due to some expected benefits from moderate temperature increase like the expansion of arable land.

3 Implications of coalition formation

In this section, we discuss some fundamental features of coalition formation, abstracting from stability that is dealt with in Sect. 4.

3.1 No cooperation and full cooperation

In this subsection, we consider two benchmarks: no and full cooperation. *No cooperation* means that each region maximizes its *own* lifetime consumption with respect to its own strategy vector, taking the strategies of all other regions as given. The solution is the *non-cooperative* or Nash equilibrium strategy vector s^N . From the perspective of coalition formation, this can be interpreted as if each region forms a singleton coalition by itself, implying coalition structure $c^N = \{\{i\}, \{j\}, \dots, \{n\}\}$. In contrast, *full cooperation* maximizes the sum of lifetime consumption over all regions. That is, each region chooses abatement and investment considering not only the effect on its own region but also on all other regions. The solution is the *fully*

³ We choose a sufficiently long time period to avoid “end point bias”. However, due to discounting, only a shorter period is strategically relevant for players.

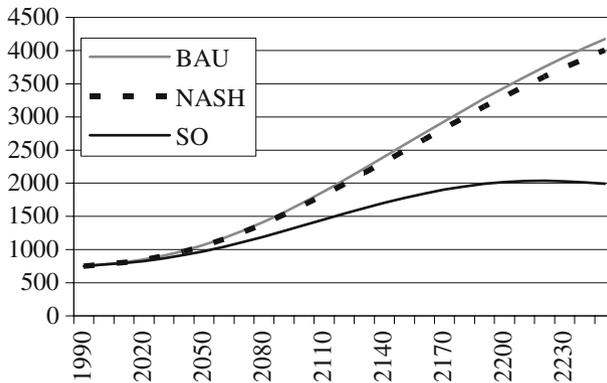


Fig. 1 Atmospheric carbon concentration (GtC)

cooperative or socially optimal strategy vector s^F . From the perspective of coalition formation, this can be interpreted as if all regions form one coalition, called the grand coalition, implying coalition structure $c^F = \{N\}$ with N the set of all players.

Because emissions constitute a negative externality⁴, strategy vector s^F differs from strategy vector s^N (Müller-Fürstenberger & Stephan, 1997). Global abatement is higher in the social optimum than in the Nash equilibrium. However, also in the Nash equilibrium, some abatement is undertaken unilaterally since—compared to the business as usual scenario (BAU)—at least national damages are taken in consideration when choosing the non-cooperative strategy vector s^N .

The difference between Nash equilibrium and social optimum shows up in a very different development of global emissions along the time axis. Whereas Nash equilibrium emissions follow very closely the business as usual path and grow steadily due to economic growth, socially optimal emissions rise only until 2150, level off and decrease afterwards. Initially, difference in annual global emissions is not so pronounced. However, by the year 2200, annual global emissions in the social optimum (20 GtC; giga tons carbon) are only one third of those in the Nash equilibrium (60 GtC). In terms of total accumulative emissions (i.e. the sum of all annual global emissions between 1990 and 2300) the difference shows up in 15,934 GtC in the Nash equilibrium but only to 7,725 GtC in the social optimum in the year 2300.

This difference also shows up in the development of carbon concentration as shown in Fig. 1. Starting from an atmospheric carbon concentration of 750 GtC in 1990, Nash concentration rises steadily and reaches 4,550 GtC in 2300. Again, the Nash equilibrium concentration path follows closely the business as usual path. In contrast, socially optimal concentration grows at a much slower rate and levels off by the year 2200, reaching a value of 1,913 GtC in 2300.

In terms of global welfare (i.e., discounted lifetime world consumption; see Table 1 for details), the difference between full cooperation (3,398.3 trillion US\$) and no cooperation (3,380.6 trillion US\$) is not so pronounced. Though the difference is not small in absolute magnitude (approximately 17.7 trillion US\$), it appears not so big in relative terms (0.52 %). However, this conclusion may be premature when not considering the development over time as shown in Fig. 2.

⁴ The mirror image is: abatement constitutes a positive externality. See Subsect. 3.2.

Table 1 Welfare and ecological implications of different coalition structures^a

| Coalition structure | | Welfare | Concentration | Emissions |
|---------------------|--|---------|---------------|-----------|
| Size | Membership | | | |
| 6 | Grand coalition (full cooperation) | 100.0 | 100.0 | 100.0 |
| 5 | USA, EU, CHN, FSU, ROW | 99.1 | 92.2 | 93.0 |
| 5 | USA, JPN, EU, CHN, ROW | 96.6 | 90.0 | 91.1 |
| 5 | JPN, EU, CHN, FSU, ROW | 95.6 | 80.6 | 81.9 |
| 4 | USA, EU, CHN, ROW | 94.5 | 82.0 | 83.2 |
| 5 | USA, JPN, CHN, FSU, ROW | 93.2 | 73.2 | 74.8 |
| 4 | EU, CHN, FSU, ROW | 91.3 | 72.3 | 73.6 |
| 4 | JPN, EU, CHN, ROW | 89.6 | 69.8 | 71.5 |
| 4 | USA, CHN, FSU, ROW | 87.4 | 64.1 | 65.7 |
| 4 | USA, JPN, CHN, ROW | 85.9 | 61.8 | 63.9 |
| 3 | EU, CHN, ROW | 84.0 | 60.7 | 62.6 |
| 3 | USA, CHN, ROW | 78.8 | 52.0 | 54.3 |
| 4 | JPN, CHN, FSU, ROW | 78.4 | 50.3 | 52.6 |
| 5 | USA, JPN, EU, FSU, ROW | 70.3 | 66.0 | 67.1 |
| 4 | USA, EU, FSU, ROW | 69.1 | 61.0 | 62.0 |
| ... | ... | ... | ... | ... |
| 2 | JPN, ROW | 46.4 | 24.7 | 26.8 |
| 5 | USA, JPN, EU, CHN, FSU | 31.0 | 26.9 | 27.5 |
| 4 | USA, EU, CHN, FSU | 29.0 | 24.5 | 25.0 |
| ... | ... | ... | ... | ... |
| 4 | USA, JPN, EU, FSU (“old Kyoto”) | 5.07 | 1.58 | 2.14 |
| ... | ... | ... | ... | ... |
| 3 | JPN, EU, FSU (“new Kyoto”) | 2.9 | 0.7 | 1.0 |
| ... | ... | ... | ... | ... |
| 2 | JPN, EU | 0.6 | 0.2 | 0.3 |
| 1 | Only singleton coalitions (no cooperation) | 0.0 | 0.0 | 0.0 |

^a Size: size of coalition S in coalition structure $c = \{S, \{k\}, \dots, \{n\}\}$. Membership: only members in coalition S are listed. Welfare: global welfare expressed in relative terms: $(\sum_{i=1}^N (W_i(c^P) - W_i(c^N)))/(\sum_{i=1}^N (W_i(c^F) - W_i(c^N)))$ where welfare W_i is discounted lifetime consumption integrated over 1990–2300, global welfare with full cooperation is $\sum_{i=1}^N w_i(c^F) = 3,398.3$ trill US\$₁₉₉₀ (trillion US dollars expressed in 1990 levels), global welfare with no cooperation is $\sum_{i=1}^N w_i(c^N) = 3,380.6$ trill US\$₁₉₉₀ and global welfare with partial cooperation is denoted by $\sum_{i=1}^N w_i(c^P)$. Concentration: atmospheric carbon concentration M at time $t = 2,300$ expressed in relative terms: $(M(c^N) - M(c^P))/(M(c^N) - M(c^F))$ where concentration with full cooperation is $M(c^F) = 1912.907$ GtC (gigatons carbon), concentration with no cooperation is $M(c^N) = 4550.202$ GtC and concentration with partial cooperation is denoted by $M(c^P)$. Emissions: total accumulative emissions over 1990–2300 at time $t = 2,300$ expressed in relative terms: $(\sum_{i=1}^N (E_i(c^N) - E_i(c^P)))/(\sum_{i=1}^N (E_i(c^N) - E_i(c^F)))$ where emissions with full cooperation are $\sum_{i=1}^N E_i(c^F) = 7,725$ GtC, emissions with no cooperation are $\sum_{i=1}^N E_i(c^N) = 15,934$ GtC and emissions with partial cooperation are denoted by $\sum_{i=1}^N E_i(c^P)$

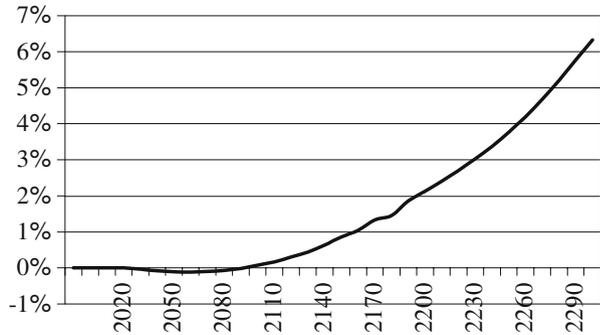
From Fig. 2, which expresses the difference between full and no cooperation in relative terms, it is evident that global welfare is substantially higher in the social optimum than in the Nash equilibrium, except in the initial years. The reason is that substantial abatement efforts are required initially in the social optimum which are only matched by lower damages in later periods given the stock pollutant nature of greenhouse gases. Thus, without discounting, the impact of full cooperation compared to no cooperation is large in economic terms in later periods. However because the negative externalities of carbon emissions occur mainly in the future and receive less weight due to discounting, the impact becomes relatively small when global welfare is measured as net present value.

Fig. 2 Relative difference of welfare between full and no cooperation*. * – Difference is computed as:

$$\Delta_t = (W_t^F - W_t^N) / W_t^N \text{ with}$$

$$W_t^F = \sum_{i \in N} \frac{Z_{it}(c^F)}{[1 + \rho_i]^t},$$

$$W_t^N = \sum_{i \in N} \frac{Z_{it}(c^N)}{[1 + \rho_i]^t} \text{ and } c^F \text{ and } c^N \text{ the coalition structure with full and no cooperation, respectively}$$



The major winners from full cooperation would be EU (9.2 trillion US\$), USA (6.4 trillion US\$) and JPN (3.2 trillion US\$). Also FSU (2.3 trillion US\$) would gain but CHN (−2.8 trillion US\$) and to a smaller extent ROW (−0.5 trillion US\$) would loose. This is because USA, JPN and EU would have to contribute below average to socially optimal abatement because of steep marginal abatement cost curves, but would benefit above average due to steep marginal damage cost curves. This “favorable” incentive structure is particular pronounced for EU. In contrast, CHN and ROW face just the opposite incentive structure that is particular “unfavorable” for CHN. Therefore, in the social optimum, CHN would have to contribute well above average to joint abatement and would benefit only very little due to a flat marginal abatement cost and a flat marginal damage cost curve.

3.2 Partial cooperation

In this subsection, we consider intermediate steps between no and full cooperation. *Partial cooperation* means that a subgroup of regions—at least two regions, but less than all regions—forms a coalition. Partial cooperation is particular important in a non-cooperative game theoretic setting because different from Eyckmans and Tulkens (2003) the grand coalition may not be stable. We assume⁵ that members of a subgroup maximize the sum of their members’ lifetime consumption. That is, every member chooses its strategy vector considering the effect on its fellow members, but ignores the effect on outsiders. Outsiders are assumed to act as singletons, as described in the non-cooperative equilibrium. Hence, the partially cooperative equilibrium strategy vector s^P can be interpreted as “partial Nash equilibrium between the coalition and outsiders” (Chander & Tulkens, 1997). The associated coalition structure is $c^P = \{S, \{k\}, \dots, \{n\}\}$ with S a non-trivial coalition (i.e., a coalition of at least two regions) and all regions that do not belong to S are singletons.

Overall, in the context of six players, the total number of coalition structures is 58: one coalition structure with no cooperation c^N , one coalition structure with full cooperation c^F and 56 coalition structures with partial cooperation c^P . There are 15 coalition structures with a coalition of two members, 20 with a coalition of three members, 15 with a coalition of four members and six with a coalition of five members.

⁵ This assumption is widely made in coalition theory. See Bloch (1997) and Yi (1997).

In Table 1, we display a selection of coalition structures and their associated welfare and ecological implications. Coalition structures are listed in descending order of global welfare. Welfare, concentration and emissions are expressed as a “closing the gap index”, that is, by how much they would “close the gap” between the social optimum and the Nash equilibrium in relative terms (see the legend of Table 1 for details).

From Table 1, it is evident that all coalition structures different from that with only singletons would generate higher global welfare (and lower global emissions and concentration). This is due to two important properties of coalition formation that apply to our model. The first property is called superadditivity. *Superadditivity* means that if a region joins a singleton or a coalition, *aggregate* welfare of all regions that are involved in this merger increases. In other words, there is a “coalitional gain” from cooperation. Superadditivity stems from the fact that coalition members can always implement the strategies they would have chosen before the merger. Hence, cooperation cannot but increase their joined payoff. The second property is called positive externality. *Positive externality* means that if a region joins a singleton or a coalition to form a bigger coalition, all outsiders that are not involved in this merger benefit from the merger. The positive externality effect can be decomposed into two effects. First, global abatement will increase after the merger because the new coalition will produce more abatement. Although outsiders to the merger will respond to the coalition’s increase of abatement by reducing their own abatement effort, this response is less than proportional.⁶ The resulting net increase of global abatement is beneficial to all players because it reduces climate change damages. Second, outsiders to the merger incur lower abatement costs because they reduce their abatement efforts in response to the higher coalitional effort.

It also appears from Table 1 that the identity of members in partial cooperation may matter more than the number of participants for the success of cooperation. Put differently, the commonly hold view that a high participation indicates success of an IEA may prove to be wrong. For instance, in our example, a coalition including the five members USA, JPN, EU, CHN, FSU ranks lower than many coalition structures comprising coalitions of only three or four members and even lower than the coalition with the two members JPN and ROW.

From a brief glance of the first 15 ranked coalition structures, it is evident that, as a tendency, the importance of membership decreases in our example along the following sequence: ROW, CHN, EU, USA, FSU and JPN. ROW’s and CHN’s important role stems from the fact that they can provide cheap abatement. Similar, JPN’s low importance is due to expensive abatement. However, there is also an additional dimension related to environmental damages. In a given coalition, the higher the marginal damages of a coalition member are, the higher the joint abatement efforts will be, everything else being equal. This explains the importance of EU for cooperation in our example. These relations also explain why the coalition comprising USA, JPN, EU and FSU that we labeled “old Kyoto coalition” ranks relatively low in our model since the two key players CHN and ROW are outsiders. A similar conclusion applies to the coalition that we labeled “new Kyoto coalition” after the withdrawal of USA. It is evident that this decision implies a dramatic drop

⁶ That is, the calibration of our model implies that there are leakage effects but they are relatively small. In other words, the slope of the reaction function of regions is negative but the absolute value is less than one. This is in line with the assumptions of most theoretical models. See Finus (2003).

in welfare and ecological variables, almost to non-cooperative levels, in our example. Thus, our model provides support for the efforts of many governments and non-governmental organization to convince the USA to rejoin the Kyoto Protocol. However, it also supports the concern of the USA and many others that an effective climate policy should include developing countries.

At the level of individual regions (not displayed), all single regions k are better off in every partially cooperative coalition structure $c^P = \{S, \{k\}, \dots, \{n\}\}$ than in the singleton coalition structure $c^N = \{\{i\}, \dots, \{j\}, \{k\}, \dots, \{n\}\}$. This is due to the positive externality property mentioned above which also implies that the most favorable condition for a singleton is if all other regions form a coalition. It is this property which makes free-riding attractive and which makes it difficult to form large stable coalitions as we will illustrate in Sect. 4.

For regions that are members of a coalition, things are different. Despite the fact that coalitions can reap a gain from cooperation due to superadditivity, individual members may be worse off than in the singleton coalition structure c^N . This has already been illustrated for full cooperation in Subsect. 3.1 and may also be true for partial cooperation. In fact, only 10 out of 56 coalition structures that constitute partial cooperation are *individually profitable*, i.e., imply a gain from cooperation to *all* members in our example. None of the top 15 ranked coalition structures in Table 1 are individually profitable. No coalition with five members and only one with four members is individually profitable. Most notably, not a single coalition including CHN as a member is individually profitable, confirming the unfavorable incentive structure of this country already identified for full cooperation. This result illustrates the large asymmetries between regions. It also suggests that without transfers even moderate partial cooperation will prove very difficult. In particular, without transfers, it means that a key player of cheap abatement, namely CHN, can hardly be convinced to join a climate treaty.

4 Stability of coalition formation

4.1 A first approach

In this subsection, we have a first look at the stability of coalition structures in our example. Clearly, a necessary condition for stability is individual profitability: a region can always remain a singleton which gives it at least as much as in the singleton coalition structure due to the positive externality property. However, even if individual profitability holds, a coalition member may have an incentive to leave its coalition. This is not the case if the following condition holds:⁷

⁷ The concept of internal & external stability that we apply is due to d' Aspremont et al. (1983). It is de facto a Nash equilibrium where no player has an incentive to revise his decision about membership. For an overview of other concepts applied in the environmental context, see Finus (2003). Note that the fundamental difference to the γ -core applied in Eyckmans and Tulkens (2003) is related to internal stability. Internal stability implies that if a member of coalition S left, the remaining coalition members $S \setminus \{i\}$ would stick together whereas the γ -core assumes that $S \setminus \{i\}$ would break apart. Given the fact that coalition formation exhibits a positive externality on outsiders (due to increased abatement efforts by the coalition), the implicit punishment after a member of S leaves this coalition is stronger for the γ -core than for internal stability. This explains why in the presence of free-rider incentives the grand coalition is usually not internally stable though it lies in the core. See Tulkens (1998) and Finus and Rundshagen (2006) for details.

$$\begin{aligned} \text{internal stability: } & W_i(\{S, \{k\}, \dots, \{n\}\}) \\ & \geq W_i(\{S \setminus \{i\}, \{i\}, \{k\}, \dots, \{n\}\}) \quad \forall i \in S. \end{aligned} \tag{3}$$

If a coalition member i leaves coalition S to become a singleton, it saves abatement costs. However, not only the deviator will reduce its abatement effort but also the remaining regions in coalition $S \setminus \{i\}$ will abate less, leading to an increase of damages in every region. The importance of both welfare effects determines the incentive of a region whether remaining in or leaving coalition S pays. As a tendency, given a coalition S , the more regions join S , the higher the incentive of current members to leave their coalition will be. The reason is that more members mean higher abatement and hence higher abatement costs and lower damages. Hence, free-riding become increasingly attractive. (That is, the incentive to leave a coalition increases gradually due to the convexity of abatement cost functions and the concavity of benefits from reduced damages.) However, there is also a second dimension of stability:

$$\begin{aligned} \text{external stability: } & W_j(\{S, \{j\}, \{k\}, \dots, \{n\}\}) \\ & \geq W_j(\{S \cup \{j\}, \{k\}, \dots, \{n\}\}) \quad \forall j \notin S. \end{aligned} \tag{4}$$

External stability is the mirror image of internal stability: no singleton should have an incentive to join coalition S . The advantage of joining is that damages drop: global abatement increases and, in particular, own efforts are matched by those of other members. However, higher abatement means also higher abatement costs. Again, the importance of both welfare effects determines the incentive of joining coalition S or remaining an outsider. For the same reason as mentioned above, as a tendency, the more regions already joined coalition S , the less attractive it becomes to follow suit.

Testing for stability, it turns out that none of the 10 individually profitable coalition structures that qualify as potential candidates in our example are stable. Though 8 are internally stable, none of them is externally stable. Hence, partial cooperation is not stable—at least not without transfers and/or a change of institutional rules—because of strong asymmetric interests.

4.2 A second approach: transfers

4.2.1 Preliminaries

From the previous discussion, it became evident that a necessary condition for stability is individual profitability that is frequently violated because of large asymmetries between regions. Thus, one way out of this dilemma seems to be transfers. For instance, consider a coalition structure $c = \{S, \{k\}, \dots, \{n\}\}$ and assume that every coalition member $i \in S$ receives additionally to its welfare without transfers $W_i(c)$ a transfer $t_i(c)$ of the following type:

$$t_i(c) = [W_i(c^N) - W_i(c)] + \lambda_i \sum_{j \in S} [W_j(c) - W_j(c^N)] \tag{5}$$

where $W_i(c^N)$ is welfare in the singleton coalition structure and $W_i(c)$ is welfare without transfers in a coalition structure c which may be partial cooperation ($c = c^P$;

$2 < \#S < 6$) or full cooperation ($c = c^F$; $\#S = 6$). The first term on the right hand side in (5) puts everybody back to its no-cooperation payoff, the second allocates the aggregate gain to the coalition from cooperation to its members where λ_i is a weight, $1 \geq \lambda_i \geq 0$, $\sum_{i \in S} \lambda_i = 1$. Substituting (5) into $\hat{W}_i(c) = W_i(c) + t_i(c)$ gives welfare with transfers:⁸

$$\hat{W}_i(c) = W_i(c^N) + \lambda_i \sum_{i \in S} [W_i(c) - W_i(c^N)] \quad (6)$$

Because $\sum_{i \in S} [W_i(c) - W_i(c^N)] > 0$ is true due to superadditivity, individual profitability holds for every member in S . Note that (6) can be interpreted as the outcome of a Nash bargaining solution. The “standard” Nash bargaining solution is the special case with equal weights $\lambda_i = 1/\#S$ that implies “equal sharing”. The “general” Nash bargaining solution allows for different weights. Typically, in the game theoretic literature, weights are interpreted as bargaining power. In contrast, in the environmental literature, weights are usually derived from various moral concepts and therefore have a normative flavor. Below, we consider several of these concepts that have been put forward in this literature.

4.2.2 Motivation and fundamental features

As mentioned in the Introduction, the transfer schemes considered in this paper are closely related to the literature that has given much attention to their moral and philosophical motivation (e.g., Rose & Stevens, 1998; Rose et al., 1998; Stevens & Rose, 2002). In the following, we first discuss the input data to compute weights of the various transfer schemes. Then we introduce our schemes, briefly mention their motivation, but refer the reader for a more detailed description to the papers above. In order to save space, we already discuss some indicators that help to explain the results of our stability analysis upon which we report in the next subsection. The discussion is illustrated in Table 2.

The first three rows in Table 2 show some commonly used indicators of economic and ecological performance of different regions. They are computed using the input data gross domestic product (GDP_i), population (POP_i) and emissions (E_i) for the base year 1990 as it enters our model (see Appendix). Emissions per capita (E_i/POP_i) illustrates that USA citizens are the largest and ROW citizens the smallest emitters per head. GDP per capita (GDP_i/POP_i) indicates that CHN is the poorest and JPN the richest region in our model. Emissions per unit of GDP (E_i/GDP_i) is a commonly used indicator to measure energy efficiency. It is evident that the Japanese is the most and the Chinese the least energy efficient economy.

The subsequent rows in Table 2 display different scenarios. Scenario 0 is the benchmark case without transfers and scenario 1 to 8 represent different transfer schemes, resulting from different weights. The first column lists the numbers and names that we attach to each scenario and the second column provides the formula for computing weights λ_i . Subsequent columns display weights (λ_i) and transfers (t_i)

⁸ This procedure is valid because we assume a lump sum transfer of discounted payoffs for simplicity. That is, transfers have no effect on equilibrium economic strategy vectors in our framework, i.e., the game is a transferable utility game (TU-game) as assumed almost throughout the literature on coalition formation. A formal proof can be found in Eyckmans and Tulkens (2003).

Table 2 Background information of scenarios ^a

| Indicators | USA | JPN | EU | CHN | FSU | ROW | Total |
|---|--|------|------|------|------|------|-------|
| Scenario 0: No Transfers | - | - | - | - | - | - | - |
| Scenario 1: Egalitarian 1 | $\lambda_i = 1/\#S$ | 3 | 0 | 31 | 18 | 12 | 67 |
| Scenario 2: Egalitarian 2 | $\lambda_i = \frac{POP_i}{\sum_{i \in S} POP_i}$ | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 1 |
| Scenario 3: Ability to Pollute | $\lambda_i = \frac{[E_i/POP_i]^{-1}}{\sum_{i \in S} [E_i/POP_i]^{-1}}$ | -3.4 | -6.3 | 5.8 | 0.6 | 3.5 | 9.9 |
| Scenario 4a: Ability to Pay ($\eta = 0.25$) | $\lambda_i = \frac{[GDP_i/POP_i]^{-\eta}}{\sum_{i \in S} [GDP_i/POP_i]^{-\eta}}$ | 25 | 26 | 0 | 4 | 0 | 70 |
| Scenario 4b: Ability to Pay ($\eta = 1$) | $\lambda_i = \frac{[GDP_i/POP_i]^{-1}}{\sum_{i \in S} [GDP_i/POP_i]^{-1}}$ | 0.05 | 0.02 | 0.22 | 0.05 | 0.05 | 1 |
| Scenario 4c: Ability to Pay ($\eta = 10$) | $\lambda_i = \frac{[GDP_i/POP_i]^{-10}}{\sum_{i \in S} [GDP_i/POP_i]^{-10}}$ | -5.5 | -2.7 | 6.8 | -1.3 | 10.9 | 17.7 |
| Scenario 5: Ecological Reward | $\lambda_i = \frac{[E_i/GDP_i]^{-1}}{\sum_{i \in S} [E_i/GDP_i]^{-1}}$ | 25 | 24 | 0 | 22 | 0 | 97 |
| Scenario 6: Ecological Subsidy | $\lambda_i = \frac{E_i/GDP_i}{\sum_{i \in S} E_i/GDP_i}$ | 0.04 | 0.09 | 0.35 | 0.06 | 0.38 | 1 |
| | | -5.7 | -7.7 | 9.1 | -1.3 | 7.2 | 16.3 |
| | | 26 | 24 | 0 | 22 | 0 | 94 |
| | | 0.11 | 0.11 | 0.30 | 0.17 | 0.21 | 1 |
| | | -4.5 | -7.3 | 8.2 | 0.7 | 4.2 | 13.1 |
| | | 25 | 26 | 0 | 7 | 0 | 81 |
| | | 0.01 | 0.01 | 0.73 | 0.08 | 0.16 | 1 |
| | | -6.1 | -9.0 | 15.7 | -0.9 | 3.3 | 19.0 |
| | | 25 | 25 | 0 | 15 | 0 | 91 |
| | | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1 |
| | | -6.4 | -9.3 | 20.6 | -2.3 | 0.5 | 21.1 |
| | | 26 | 25 | 0 | 22 | 15 | 114 |
| | | 0.16 | 0.30 | 0.02 | 0.03 | 0.11 | 1 |
| | | -3.6 | -3.9 | 3.2 | -1.8 | 2.4 | 9.3 |
| | | 22 | 15 | 10 | 26 | 1 | 74 |
| | | 0.06 | 0.03 | 0.46 | 0.32 | 0.09 | 1 |
| | | -5.2 | -8.7 | 11.1 | 3.4 | 2.2 | 16.7 |
| | | 25 | 26 | 0 | 1 | 8 | 86 |

Table 2 continued

| Indicators | E_i/POP_i | USA | JPN | EU | CHN | FSU | ROW | Total |
|------------|------------------------------------|-------|-------|-------|------|------|------|-------|
| | GDP _i /POP _i | 5432 | 2364 | 2379 | 590 | 3685 | 548 | 1132 |
| | E_i/GDP_i | 21827 | 23734 | 18630 | 326 | 2956 | 1492 | 4003 |
| | | 249 | 100 | 128 | 1808 | 1246 | 367 | 283 |

| | | | | | | | | |
|--------------|--|-------------|------|------|------|------|------|-----|
| Scenario 7: | $\lambda_i = \frac{E_i/POP_i}{\sum_{i \in S} E_i/POP_i}$ | λ_i | 0.34 | 0.27 | 0.01 | 0.04 | 0.02 | 1 |
| Status Quo 1 | t_i | t_i | 3.0 | -4.5 | 2.9 | -1.6 | 0.9 | 6.8 |
| Scenario 8: | VIS_i | VIS_i | 0 | 22 | 23 | 25 | 16 | 93 |
| | $\lambda_i = \frac{GDP_i/POP_i}{\sum_{i \in S} GDP_i/POP_i}$ | λ_i | 0.16 | 0.16 | 0.04 | 0.25 | 0.04 | 1 |
| Status Quo 2 | t_i | t_i | -0.4 | -6.4 | 3.5 | 2.0 | 1.2 | 6.8 |
| | VIS_i | VIS_i | 15 | 25 | 0 | 0 | 12 | 54 |

^a Indicators are calculated using base data in Table 7 (Appendix) with $GDP_i = Y_i^0$ (gross domestic product), $POP_i = L_i$ (population) and $E_i = E_i^0$ (emissions). Emissions per capita (E_i/POP_i) are measured in kilogram carbon per capita, GDP per capita (GDP_i/POP_i) in US\$ per capita and emission intensity of GDP_i/E_i in gram carbon per US\$. λ_i : weight of transfer scheme (see section 4.2.1), t_i : transfer (trillion US\$), VIS_i : number of violations of internal stability. "Total" for transfers in the last column refers to the sum of all positive transfers

under the assumption of full cooperation. Though the values will differ for other partially cooperative coalition structures, they give at least a rough indication of the welfare implications of different scenarios. We also display the number of violations of internal stability in terms of region i (VIS_i), considering all 58 coalition structures. Since there are 31 coalition structures where region i is a member of a coalition of at least two members, the maximum number of possible violations is 31. Again, the number of violations can only be a crude indicator but provides already valuable insights into the stability implications of our different scenarios. For instance, the absence of transfers (scenario 0) favors in particular EU ($VIS_i = 0$) and puts CHN ($VIS_i = 31$) at a big disadvantage, confirming our observations from above.

Scenario 1 assumes equal weights. That is, each member receives the same share from the gains from cooperation. This solution may be seen a focal point of fairness in the sense of Schelling (1960). However, it treats all participants equally, regardless, how much they contribute to cooperation and regardless of the size or other characteristics of their economy. This is different in *scenario 2* where weights are related to population. The normative idea behind this rule may be summarized as: “one man one vote”. We call scenario 1 and 2 “egalitarian 1” and “egalitarian 2”. For the grand coalition, “egalitarian 1” means a larger share of the gains from cooperation for CHN and ROW compared to “no transfers” in our example. Transfers flow from USA, JPN and EU to CHN, FSU and ROW. That this means a substantial redistribution of welfare, and, in fact, again, an asymmetry, but opposite to that in the no transfer case, is evident from the change of the number of “VIS”. A similar but even more pronounced pattern holds for “egalitarian 2” because population is strongly concentrated in CHN and ROW.

Scenario 3 relates weights to the inverse of the emission per capita ratio and scenario 4 does this with respect to the inverse of the GDP per capita ratio. Scenario 3’s motivation is that every man should have the same “right to pollute”. It is also associated with historical responsibility for the current stock of greenhouse gases. Since USA has the highest emission per capita ratio, they receive the lowest weight and because this is completely reversed for CHN and ROW, both regions receive high weights. *Scenario 4* allocates the gains from cooperation to the “poor” and thus uses environmental policy as a vehicle to transfer money from the “rich” to the “poor”. The parameter η is usually referred to as the “degree of inequality aversion”. We consider three values where $\eta \rightarrow +\infty$ would correspond to the “Rawlsian maximin rule”. In our model, already $\eta = 10$ approximates this rule since all weights are zero, except China’s weight that is 1. However, even a value of “only” $\eta = 0.25$ implies a substantial reshuffling of the gains from cooperation from the “rich” to the “poor”. We follow the literature and call scenario 4 “ability to pay” and because of its similar design scenario 3 “ability to pollute”. Again, “extreme” weights in scenario 3 and 4 show up in a high number of “VIS” in industrialized countries and low numbers for developing countries and economies in transition in our example.

Scenario 5 and 6 are related to the ecological dimension and are therefore called “ecological reward” and “ecological subsidy”, respectively. Both scenarios relate weights to energy efficiency that is measured as “emissions per output”. “Ecological reward” means that those regions with a high energy efficiency receive a high share of the gains from cooperation whereas this is the opposite with “ecological subsidy”. Whereas the motivation for rewards is more or less self-evident, the motivation for subsidies has to argue that dirty regions should receive sufficient resources to clean up their environment. In line with the mainstream of evidence, in our model, energy

efficiency is very high in JPN and EU but very low in CHN and FSU. Due to low emissions associated with its low stage of economic development, ROW's energy efficiency is relatively high.

Scenario 7 and 8 are the mirror images of scenarios 3 and 4. Whereas scenarios 3 and 4 implied a major redistribution of the gains from cooperation based on “ecological or welfarism justice”, scenario 7 and 8 more or less preserve the “status quo”. They also have been called sovereignty rules in the literature. That is, these transfer schemes acknowledge “political reality” that huge transfers are politically not feasible. Both scenarios give relatively high weights to USA, JPN and EU.

Overall, Table 2 suggests that except for the “pragmatic solutions” “status quo 1” and “status quo 2”, the amount of transfers may be very large which is evident by comparing the total amount of transfers under the various scenarios (last column) and recalling that the total gain from cooperation amounts to 17.7 trillion \$.⁹ It is already evident that many transfer schemes, though they may be grounded on well-defendable moral motives, would change the asymmetry of the no transfer scenario, but may introduce another asymmetry. In particular, the high and very concentrated violations of internal stability under some scenarios already indicate that though some transfer schemes may be regarded as “just”, they may not be very useful in encouraging a high voluntary and self-enforcing participation.

4.2.3 Results

Table 3 lists internally and externally stable coalition structures that are stable under at least one scenario in our example. “X” means that this coalition structure is internally and externally stable (as defined in Sect. 4.1) under a particular scenario; what “X” means will be explained in Sect. 4.3.

Whereas in the no transfer case no coalition was stable, now at least one coalition is stable under every transfer scheme. With exception of scenario 4c, only one or two coalitions are stable. No coalition with more than three members is stable and most stable coalitions comprise only two members. This stresses that individual profitability is a necessary but by no means a sufficient condition for stability. Also with transfers, the free-rider incentive increases with the size of coalitions. Nevertheless, also small coalitions can make a difference by closing the gap between no and full cooperation under most scenarios by 50 or more percent.

Under most scenarios, the number of violations VIS_i as listed in Table 2 already provides a good indication which regions participate in a stable coalition as listed in Table 3: only regions with a low VIS_i number form stable coalitions. It is interesting to observe that no coalition including only the key industrialized regions USA, JPN and EU is stable in our model. Moreover, scenario 1 to 4b and 6 do not involve any of the key industrialized regions, but only FSU, CHN and ROW. This indicates that it is not that straightforward to establish stable cooperation between heterogeneous players through transfers, i.e., between industrialized countries, countries in transition and developing countries. As conjectured already in the previous subsection,

⁹ “Total transfers” in the last column of Table 3 are the sum of all positive transfers (=sum of all negative transfers) but not the sum of all transfers that are zero by definition. Total transfers are an indicator of the amount of financial resources redistributed by the transfer scheme. It is important to note that because transfers are computed according to (5) it can happen that total transfers exceed the total gain from cooperation. For instance, suppose $W_i(c^N) - W_i(c) > 0$ and $\lambda_i = 1$, then $t_i(c) > \sum_{i \in S} [W_i(c) - W_i(c^N)]$.

Table 3 Stable coalition structures^a

| Coalition structure | Welfare | No transfers | Egalitarian 1 | Egalitarian 2 | Ability to pollute | Ability to pay | | | Ecological reward | Ecological subsidy | Status Quo 1 | Status Quo 2 |
|---------------------------------------|---|--------------|---------------|---------------|--------------------|----------------|------------|-------------|-------------------|--------------------|--------------|--------------|
| | | | | | | $\eta = 0.25$ | $\eta = 1$ | $\eta = 10$ | | | | |
| Size | Membership | 0 | 1 | 2 | 3 | 4a | 4b | 4c | 5 | 6 | 7 | 8 |
| 3 | EU, CHN, ROW | | | | | | | | | | X | |
| 3 | CHN, FSU, ROW | | X,X | | | | | | | | | |
| 3 | JPN, CHN, ROW | | | | | | | | | | X | |
| 3 | USA, FSU, ROW | X | | | | | | | | | | |
| 3 | USA, JPN, ROW | X | | | | | | | | | | X,X |
| 2 | EU, ROW | X | | | X | | | X | X | | | X |
| 2 | CHN, ROW | X | | X,X | X,X | X,X | | X,X | | | | X |
| 3 | JPN, FSU, ROW | X | | | | | | | | | | |
| 2 | USA, ROW | X | | | X | X | X | X | X | | | X |
| 2 | FSU, ROW | X | | | X | X | X | X | X | | | X |
| 2 | JPN, ROW | X | | | X | X | X | X | X,X | | | |
| 3 | USA, CHN, FSU | | | | | | | | | | X,X | |
| 2 | EU, CHN | | | | | | | X,X | | | | |
| 2 | USA, CHN | | | | | | | X,X | | | | |
| 2 | JPN, CHN | | | | | | | X,X | | | | X |
| 2 | CHN, FSU | | | | | | | X,X | | | | X,X |
| 2 | EU, FSU | | | | | | | X | | | | |
| 2 | USA, FSU | | | | | | | X | | | | |
| 2 | USA, EU | | | | | | | X | | | | |
| 2 | JPN, FSU | | | | | | | X | | | | |
| 2 | JPN, EU | | | | | | | X | | | | |
| Number of stable cs, open membership | | | | | | | | | | | | |
| | Max. welfare of stable cs, open membership | 0 | 68.3 | 54.6 | 54.6 | 54.6 | 54.6 | 54.6 | 46.4 | 8.7 | 20.5 | 59.0 |
| Number of stable cs, excl. membership | | | | | | | | | | | | |
| | Max. welfare of stable cs, excl. membership | 60.7 | 68.3 | 54.6 | 57.5 | 54.6 | 57.5 | 57.5 | 57.5 | 57.5 | 84.0 | 59.0 |

^a List of coalition structures that are internally and externally stable in some scenarios. Coalition structures are sorted in descending order of global welfare; welfare measured in relative terms as described in the legend of Table 1. X means internally and externally stable under open membership, X means internally and externally stable under exclusive membership, assuming unanimity voting

this is partly due to the fact that most transfer schemes would just replace one asymmetry by another. Only the two status quo scenarios are more balanced. It is exactly for this reason why there are stable coalitions with “mixed membership” under scenario 7 and 8.

However, there is also a second dimension of stability, namely, external stability that explains participation in stable coalitions. Here, the arguments are more or less reversed. That is, whereas asymmetries hamper internal stability, they foster external stability. For instance, consider scenario 1 with equal sharing. As pointed out in the previous subsection, this transfer scheme favors CHN, FSU and ROW at the expenses of USA, JPN and EU. Therefore, any coalition including apart from CHN, FSU and ROW one of the industrialized countries is not internally stable in our example. However, this means that the coalition between CHN, FSU and ROW is externally stable because no key industrialized country has an incentive to join this coalition.

Taken together, we it has been illustrated that transfers can make a difference but fail to stabilize full cooperation. Membership in stable coalitions crucially depends on the design of the transfer rule. Participation of industrialized and non-industrialized regions requires a “balanced transfer scheme” that may not be in line with equity considerations.

4.3 A third approach: changing membership rules

4.3.1 Introduction

From the previous discussion, it became evident that the stability problem has two dimensions: internal and external stability. In terms of internal stability one could argue that stability could be improved by punishing those regions that leave the coalition very harshly. Though this is certainly true, it is also evident that punishment must be credible. In fact, in our model, there is an implicit punishment as we assume that coalitions and single players play a partial Nash equilibrium in economic strategies (see Sect. 3). Hence, if region i leaves coalition S , the remaining regions $S \setminus \{i\}$ reoptimize their economic strategies. This means for instance that the remaining regions revise their abatement targets downward which have a negative impact on the culprit through higher damages. Though this means not the harshest possible punishment, it means a credible punishment. After a deviation, regions $S \setminus \{i\}$ revise their equilibrium economic strategies which is a best reply under the “new” condition.¹⁰

It could also be argued that the problem of internal instability could be mitigated by simply not allowing members to leave. However, it is evident that such a change of the institutional rule would not be in line with the notion of voluntary participation. Alternatively, however, one may consider improving stability by not allowing outsiders to join a coalition. That such a modification is at least in theory possible is evident when noting that, literally, external stability means an *open membership rule*. That is, every outsider can join the coalition without asking for acceptance. Hence, an alternative institutional scheme could be an *exclusive membership rule*. For instance, suppose that an outsider wants to accede to an

¹⁰ The γ -core also assumes a best reply after player i has left coalition S but $S \setminus \{i\}$ breaks apart implying lower abatement by these players and hence harsher punishment. See endnote 7.

agreement. Then, the current members of a coalition could vote either by majority or unanimity voting about accession of the new member. Of course, such voting procedures are usually not part of international environmental agreements. However, there are many examples of other international treaties or organizations like for instance the World Trade Organization (WTO), the European Union (EU) and the Security Council of the United Nations Organization (UNO) that control membership through exclusive membership. At least a priori, there seems to be no reason why exclusive membership should not be considered as an alternative rule. This is particular true since not all examples deal only with club goods (Cornes & Sandler, 1996). For instance, UNO, fostering security and political stability worldwide through peace keeping missions, may be seen as an organization that provides a public good.

From a purely technical point of view, it is easy to see that an institutional change may imply that some internally stable coalitions that are externally unstable under open membership may become externally stable under exclusive membership. This will be the case if first an outsider has an incentive to join a coalition S , and second if its application is turned down because a single member (unanimity voting) or a majority of members (majority voting) would loose from accession. Thus, the interesting question (because it cannot be answered by theoretical reasoning as mentioned in the Introduction) is whether “more stability” translates into higher global welfare.

Before turning to the results, it is important to pause for a second and to realize that a change of membership is not just a “technical” trick. For instance, it could be argued that current members never turn down an application because more members mean more contributors to the public good. However, in our model, a coalition $S \cup \{i\}$ increases abatement efforts compared to coalition S , which may be regarded as “too ambitious” by at least some of the current members. Consequently, one could argue that all current members of S could just allow region i to join, asking it to increase its abatement effort, but that all members S do not change their economic strategies. Alternatively, one could argue that region i must not join coalition S , but may just increase its abatement effort, which current members neither can avoid nor they have an interest to do so. However, both alternatives can never be in the interest of the “potential accessor” i in our model. The reason is simple: under coalition structure $c = \{S, \{i\}, \{k\}, \dots, \{n\}\}$, $s = (s, s_{-i})$ is the equilibrium economic strategy vector that implies that s_i is a best reply to s_{-i} (and vice versa). Consequently, if under coalition structure $c' = \{S \cup \{i\}, \{k\}, \dots, \{n\}\}$ s_{-i} is the same as under $c = \{S, \{i\}, \{k\}, \dots, \{n\}\}$, it cannot be an improvement for region i to change its strategy to s'_i when joining coalition S . A similar argument applies if region i only changes its strategy without joining coalition S .

4.3.2 Results

In Table 3, “**X**” indicates stable coalition structures under exclusive membership in our example where we restrict attention to unanimity voting. In the case of no transfers (scenario 0), the impact is dramatic. Now seven coalition structures are stable where the most successful achieves a welfare level of 60.7 percent. Also for the various transfer schemes the maximum welfare is usually raised, though the changes are not so pronounced as in the case of no transfers. The largest difference can be

observed for scenario 7, Status Quo 1. Obviously, exclusive membership makes not much difference for scenarios 1 to 4c because industrialized regions have no interest anyway in joining non-industrialized regions.

Taken together, the results stress that institutional rules may be as important as transfers for the stability of international environmental agreements. In particular, if transfers are not available, difficult to implement or politically not feasible, then a change of institutional rules may be an alternative measure to increase the success of cooperation. This is even more true when considering that such institutional changes involve basically no costs. Moreover, now with exclusive membership, morally motivated transfer schemes (scenario 1 to 6) are with one exception inferior in welfare terms to no transfers and the two status quo rules. This indicates that in a second-best-world with free-rider incentives moral motives may not always be a good guide to establish effective agreements. This clearly contrasts to Eyckmans and Tulkens (2003) who show that for their transfer scheme (which is based on some normative notion) the grand coalition is in the γ -core.

5 Summary and conclusion

We analyzed coalition formation in the context of global warming. Our approach combines a game theoretic analysis with numerical simulations based on a dynamic integrated assessment model that captures the feedback between the economy, the climate system and environmental damages. The model comprises six world regions: USA, Japan, European Union, China, Former Soviet Union and “Rest of the World”. Stability of coalitions was tested with the concept of internal and external stability. From our simulations, we mention six key results.

First, in the context of global warming, the difference between full and no cooperation was large in ecological terms. It was not so large in welfare terms due to discounting and because the benefits from cooperation will occur mainly in the future.

Second, also partial cooperation can be an important step to mitigate global warming. However, the identity may be more important than the number of members for the success of partial cooperation. This indicated that concluding success only from a high participation without measuring the effectiveness of an agreement may be misleading. Moreover, we found that coalitions that do not comprise key players with low marginal abatement costs (e.g., CHN and ROW) and/or high marginal damages (e.g., EU and ROW) will not achieve much at the global level. This finding stressed that including developing countries (apart from industrialized countries) in future climate treaties will be of great importance.

Third, without transfers and under open membership rule, there was no stable coalition in our model. This result illustrated the difficulties of cooperation in the case of global warming because of large asymmetries between regions.

Fourth, the success of cooperation could be improved through the eight transfer schemes considered in this paper, though no more than three out of six world regions participated in stable cooperation. Hence, we argued that making agreements individually profitable to all participants through transfers is a necessary but by no means a sufficient condition to establish successful self-enforcing treaties. Strong free-rider incentives are an obstacle to higher participation and success.

Fifth, the two “status quo” transfer schemes performed better than the six “morally motivated” transfer schemes. The latter group of schemes implies large transfers from industrialized to developing countries. It therefore mitigated the asymmetric distribution of the gains from cooperation observed for no transfers, but at the expenses of introducing a “new” asymmetry. This did not allow for stable coalitions with participation of key industrialized countries that are also important for the success of joint climate policy. In contrast, our “status quo” transfer schemes implied a more moderate redistribution and therefore were more successful in attracting both industrialized and developing countries for stable cooperation. Consequently, we concluded that moral motives may not always be a good guide to establish effective agreements if they are not in line with fundamental free-rider incentives.

Sixth, changing the institutional design from open to exclusive membership could make a big difference and were as important as transfers. Therefore, we concluded that in future environmental treaties open membership should not be taken for granted despite that it may seem an obvious rule in the context of public goods.

Evidently, models are limited in their applicability. From a normative viewpoint, these limitations seem less severe when focusing more on the qualitative instead on the quantitative results which seems sensible given the uncertainty associated with cost-benefit data. From a positive viewpoint, these limitations may appear more severe. We mention three aspects. First, our model divides the world in six regions whereas in reality we roughly count 200 countries. In particular, the level of ROW’s aggregation is very high which may lead to an overestimation of the perspectives of stable cooperation. Second, our integrated assessment model CWSM, as well as most CGE-models analyzing climate change, assumes damages to be certain and continuous. On the one hand, this may lead to an overestimation of the incentives for cooperation. On the other hand, there are voices that claim that damage estimates in RICE (that are the basis of our model CWSM) are too low because of an incomplete representation of the Earth’s carbon cycle and climate system (e.g., Joos, Müller-Fürstenberger & Stephan, 1999; Kaufmann, 1997). Moreover, provided some agents would attach some probability to uncertain catastrophic events caused by global warming, then our assumptions imply an underestimation. It is also clear that there are other aspects apart from cost and benefits that determine the position of governments whether to participate in a climate treaty. Third, also our long time horizon may overestimate the foresight of politicians. However, a much shorter period would simply ignore the climate problem and would provide no incentive for cooperation at all. Moreover, discounting adjusts “too much” foresight downward.

For future research, we would like to mention three of certainly many possible options. First, our analysis was based on a stylized simulation model and theoretical assumptions to keep the analysis tractable. Subsequent research should focus on richer strategy sets for the players (e.g. the possibility to change membership of coalitions over time) and should use simulation models describing in more detail the physics (e.g. carbon cycle and temperature systems), the economics (e.g. explicit representation of energy systems) and the political dimension of the problem. Second, it would be interesting to gain more theoretical insights in how transfer schemes should be designed so that they mitigate free-rider incentives in an “optimal way”. Third, in the absence of transfers, it would be interesting to analyze how abatement duties should be allocated to improve upon the success of self-enforcing treaties. We assumed that the coalition chooses abatement levels by maximizing coalitional

welfare. However, in the presence of free-rider incentives and because of the need for self-enforcing agreements, it may be the case that less ambitious abatement targets and/or a departure from a cost-effective abatement allocation may be more successful if this allows for a sufficiently higher participation in a stable agreement.

Acknowledgement This paper has been written while M. Finus was a visiting scholar at the Katholieke Universiteit Leuven, Centrum voor Economische Studiën (Leuven, Belgium) and at Fondazione Eni Enrico Mattei, FEEM (Venice, Italy). He acknowledges the financial support by the CLIMNEG 2 project funded by the Belgian Federal Science Policy Office and the kind hospitality of FEEM. Both authors acknowledge research assistance by Carmen Dunsche.

Appendix

The CLIMNEG World Simulation Model comprises the following equations:

$$\begin{aligned}
 Y_{i,t} &= Z_{i,t} + I_{i,t} + Y_{i,t}C_i(\mu_{i,t}) + Y_{i,t}D_i(\Delta T_t) \\
 Y_{i,t} &= a_{i,t}K_{i,t}^\gamma L_{i,t}^{1-\gamma} \\
 C_i(\mu_{i,t}) &= b_{i,1}\mu_{i,t}^{b_{i,2}} \\
 D_i(\Delta T_t) &= \theta_{i,1}\Delta T_t^{\theta_{i,2}} \\
 K_{i,t+1} &= [1 - \delta_K]K_{i,t} + I_{i,t} \quad K_{i,0} \text{ given} \\
 E_{i,t} &= \alpha_{i,t}[1 - \mu_{i,t}]Y_{i,t} \\
 M_{t+1} &= [1 - \delta_M]M_t + \beta \sum_{i \in N} E_{i,t} \quad M_0 \text{ given} \\
 F_t &= \frac{4.1 \ln(M_t)}{\ln(2)} + F_t^x \\
 T_t^0 &= T_{t-1}^0 + \tau_3 [T_{t-1}^a - T_{t-1}^0] \\
 T_t^a &= T_{t-1}^a + \tau_1 [F_t - \lambda T_{t-1}^a] - \tau_2 [T_{t-1}^a - T_{t-1}^0] \\
 \Delta T_t &= \frac{T_t^a}{2.50}
 \end{aligned}$$

Table 4 List of variables

| | |
|--------------|---|
| $Y_{i,t}$ | Production (billion 1990 US\$) |
| $Z_{i,t}$ | Consumption (billion 1990 US\$) |
| $I_{i,t}$ | Investment (billion 1990 US\$) |
| $K_{i,t}$ | Capital stock (billion 1990 US\$) |
| $C_{i,t}$ | Cost of abatement (billion 1990 US\$) |
| $D_{i,t}$ | Damage from climate change (billion 1990 US\$) |
| $E_{i,t}$ | Carbon emissions (gigatons = billion tons of C) |
| $\mu_{i,t}$ | Emission abatement |
| M_t | Atmospheric carbon concentration (gigatons = billion tons of C) |
| F_t | Radiative forcing (Watt per m ²) |
| F_t^x | Exogenous radiative forcing (Watt per m ²) |
| T_t^a | Temperature increase in the atmosphere (°C) |
| T_t^0 | Temperature increase in the deep ocean (°C) |
| ΔT_t | Change of temperature increase in the atmosphere (°C) |

Table 5 Global parameter values

| | | |
|----------------|---------------------------------------|--------|
| $a_{i,t}$ | Productivity | RICE |
| $L_{i,t}$ | Population | RICE |
| $\alpha_{i,t}$ | Emission-output rate | RICE |
| δ_K | Capital depreciation rate | 0.10 |
| γ | Capital productivity parameter | 0.25 |
| β | Airborne fraction of carbon emissions | 0.64 |
| δ_M | Atmospheric carbon removal rate | 0.0833 |
| τ_1 | Parameter temperature relationship | 0.226 |
| τ_2 | Parameter temperature relationship | 0.44 |
| τ_3 | Parameter temperature relationship | 0.02 |
| λ | Parameter temperature relationship | 1.41 |
| M_0 | Initial carbon concentration | 590 |
| T_0^a | Initial temperature atmosphere | 0.50 |
| T_0^b | Initial temperature deep ocean | 0.10 |

Table 6 Regional parameter values

| | $\theta_{i,1}$ | $\theta_{i,2}$ | $b_{i,1}$ | $b_{i,2}$ | ρ_i |
|-----|----------------|----------------|-----------|-----------|----------|
| USA | 0.01102 | 2.0 | 0.07 | 2.887 | 0.015 |
| JPN | 0.01174 | 2.0 | 0.05 | 2.887 | 0.015 |
| EU | 0.01174 | 2.0 | 0.05 | 2.887 | 0.015 |
| CHN | 0.01523 | 2.0 | 0.15 | 2.887 | 0.030 |
| FSU | 0.00857 | 2.0 | 0.15 | 2.887 | 0.015 |
| ROW | 0.02093 | 2.0 | 0.10 | 2.887 | 0.030 |

Table 7 Variables in 1990 (reference year)^a

| | Y_i^0 | (%) | K_i^0 | (%) | L_i^0 | (%) | E_i^0 | (%) |
|-------|------------|-------|------------|-------|-----------|-------|---------|-------|
| USA | 5,464.796 | 25.9 | 14,262.510 | 26.3 | 250.372 | 4.8 | 1.360 | 20.5 |
| JPN | 2,932.055 | 13.9 | 8,442.250 | 15.6 | 123.537 | 2.4 | 0.292 | 10.9 |
| EU | 6,828.042 | 32.4 | 18,435.710 | 34.0 | 366.497 | 7.0 | 0.872 | 28.9 |
| CHN | 370.024 | 1.8 | 1,025.790 | 1.9 | 1,133.683 | 21.5 | 0.669 | 3.0 |
| FSU | 855.207 | 4.1 | 2,281.900 | 4.2 | 289.324 | 5.5 | 1.066 | 6.8 |
| ROW | 4,628.621 | 22.0 | 9,842.220 | 18.1 | 3,102.689 | 58.9 | 1.700 | 29.9 |
| World | 21,078.750 | 100.0 | 54,290.380 | 100.0 | 5,266.100 | 100.0 | 5.959 | 100.0 |

^a Y_i^0 and K_i^0 billion US\$, L_i^0 million people and E_i^0 gigatons carbon equivalent. Data are taken from Nordhaus and Yang (1996)

References

- Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46, 804–878.
- Barrett, S. (1997). Heterogeneous international environmental agreements. In C. Carraro (Ed.), *International environmental negotiations: Strategic policy issues* (pp. 9–25). Cheltenham: Edward Elgar.
- Barrett, S. (2001). International cooperation for sale. *European Economic Review*, 45, 1835–1850.
- Bloch, F. (1997). Non-cooperative models of coalition formation in games with spillovers. In C. Carraro, & D. Siniscalco (Eds.), *New directions in the economic theory of the environment* (ch. 10, pp. 311–352). Cambridge: Cambridge University Press.
- Böhringer, C., & Vogt, C. (2004). The dismantling of a breakthrough: The Kyoto protocol as symbolic policy. *European Journal of Political Economy*, 20, 597–618.
- Bosello, F., Buchner, B., & Carraro, C. (2003). Equity, development, and climate change control. *Journal of the European Economic Association*, 1, 601–611.

- Botteon, M., & Carraro, C. (1997). Burden-sharing and coalition stability in environmental negotiations with asymmetric countries. In C. Carraro (Ed.), *International environmental negotiations: Strategic policy issues* (ch. 3, pp. 26–55). Cheltenham: Edward Elgar.
- Carraro, C. (2000). Roads towards international environmental agreements. In H. Siebert (Ed.), *The economics of international environmental problems* (pp. 169–202). Tübingen: Mohr Siebeck.
- Carraro, C., & Siniscalco, D. (1993). Strategies for the international protection of the environment. *Journal of Public Economics*, 52, 309–328.
- Chander, P., & Tulkens, H. (1997). The core of an economy with multilateral environmental externalities. *International Journal of Game Theory*, 26, 379–401.
- Cornes, R., & Sandler, T. (1996). *The theory of externalities, public goods and club goods* (2nd ed.). Cambridge: Cambridge University Press
- d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., & Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics*, 16, 17–25
- Endres, A. (1997). Negotiating a climate convention – the role of prices and quantities. *International Review of Law and Economics*, 17, 147–156
- Eyckmans, J., & Tulkens, H. (2003). Simulating coalitionally stable burden sharing agreements for the climate change problem. *Resource and Energy Economics*, 25, 299–327
- Finus, M. (2003). Stability and design of international environmental agreements: The case of transboundary pollution. In H. Folmer, & T. Tietenberg (Eds.), *International yearbook of environmental and resource economics* (2003/4, ch. 3, pp. 82–158). Cheltenham, UK: Edward Elgar.
- Finus, M., & Rundshagen, B. (2003). Endogenous coalition formation in global pollution control. A partition function approach. In C. Carraro (Ed.), *Endogenous formation of economic coalitions* (ch. 6, pp. 199–241). Cheltenham, UK: Edward Elgar.
- Finus, M., & Rundshagen, B. (2006). A micro-foundation of core-stability in positive externality coalition games. *Journal of Institutional and Theoretical Economics*, 162, 329–346
- Finus, M., & Tjøtta, S. (2003). The Oslo protocol on sulfur reduction: The great leap forward? *Journal of Public Economics*, 87, 2031–2048.
- Germain, M., Toint, P. L., & Tulkens, H. (1998). Financial transfers to ensure cooperative international optimality in stock pollutant abatement. In S. Faucheux, J. Gowdy, & I. Nicolai (Eds.), *Sustainability and firms: Technological change and the changing regulatory environment* (ch. 11, pp. 205–219). Cheltenham UK: Edward Elgar.
- Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics*, 2, 141–159.
- Hoel, M., & Schneider, K. (1997). Incentives to participate in an international environmental agreement. *Environmental and Resource Economics*, 9, 153–170.
- IPCC (2001). *Climate change 2001: Mitigation* (Contribution of working group III to the third assessment report of the intergovernmental panel on climate change. Cambridge: Cambridge University Press.
- Joos, F., Müller-Fürstenberger, G., & Stephan, G. (1999). Correcting the carbon cycle representation: How important is it for the economics of climate change? *Environmental Modeling and Assessment*, 4, 133–140.
- Kaufmann, R. K. (1997). Assessing the DICE model: Uncertainty associated with the emission and retention of greenhouse gases. *Climatic Change*, 35, 435–448.
- Kemfert, C., & Tol, R. (2002). Equity, international trade and climate policy. *International Environmental Agreements: Politics, Law and Economics*, 2, 23–48.
- Müller-Fürstenberger, G., & Stephan, G. (1997). Environmental policy and cooperation beyond the nation state: An introduction and overview. *Structural Change and Economic Dynamics*, 8, 99–114.
- Murdoch, J. C., & Sandler, T. (1997a). Voluntary cutbacks and pretreaty behavior: The Helsinki protocol and sulfur emissions. *Public Finance Review*, 25, 139–162.
- Murdoch, J. C., & Sandler, T. (1997b). The voluntary provision of a pure public good: The case of reduced CFC emissions and the Montreal protocol. *Journal of Public Economics*, 63, 331–349.
- Nordhaus, W., & Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *American Economic Review*, 86, 741–765.
- Pearce, D. W., Cline, W. R., Achanta, A. N., Fankhauser, S., Pachauri, R. K., Tol, R. S. J., & Vellinga, P. (1996). The social costs of climate change: greenhouse damage and the benefits of control. In J. P. Bruce, H. Lee, & E. F. Haites (Eds.), *Climate change 1995: Economic and social dimensions—contribution of working group III to the second assessment report of the intergovernmental panel on climate change* (pp. 179–224). Cambridge: Cambridge University Press.

- Rose, A., & Stevens, B. (1998). A dynamic analysis of fairness in global warming policies: Kyoto, Buenos Aires, and Beyond. *Journal of Applied Economics*, 1, 329–362.
- Rose, A., Stevens, B., Edmonds, J., & Wise, M. (1998). International equity and differentiation in global warming policy. *Environmental and Resource Economics*, 12, 25–51.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Stevens, B., & Rose, A. (2002). A dynamic analysis of the marketable permits approach to global warming policy: A comparison of spatial and temporal flexibility. *Journal of Environmental Economics and Management*, 44, 45–69
- Tol, R. (2005). The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties. *Energy Policy*, 33, 2064–2074
- Tulkens, H. (1998). Cooperation versus free-riding in international environmental affairs: Two approaches. In N. Hanley & H. Folmer (Eds.), *Game theory and the environment* (ch. 2, pp. 30–44). Cheltenham, UK: Edward Elgar.
- Weikard, H.-P., Finus, M., & Altamirano-Cabrera, J.-C. (2006). The impact of surplus sharing on the stability of international climate agreements. *Oxford Economic Papers*, 58, 209–232.
- Weitzman, M. L. (2001). Gamma discounting. *American Economic Review*, 91, 260–271.
- Weyant J., & Hill, J. N. (1999). Introduction and overview. *Energy Journal* Special Issue: The costs of the Kyoto protocol: A multi-model evaluation, vii–xliv.
- Yi, S.-S. (1997). Stable coalition structures with externalities. *Games and Economic Behavior*, 20, 201–237.