Schopf, Mark

Conference Paper

Self-enforcing International Environmental Agreements and Altruistic Preferences

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage"

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Schopf, Mark (2023) : Self-enforcing International Environmental Agreements and Altruistic Preferences, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2023: Growth and the "sociale Frage", ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/277598

Self-enforcing international environmental agreements and altruistic preferences

Mark Schopf*

May 12, 2023

Abstract

This paper analyses the effects of altruism on the formation of climate coalitions in the standard two-stage game of self-enforcing international environmental agreements. Altruism implies that each country values, to some extent, every other country's welfare when deciding on its coalition membership and emissions policy. In the Nash [Stackelberg] game, the fringe [coalition] countries exploit the altruism of the coalition [fringe] countries so that altruism decreases [increases] the coalition size. In any case, global emissions and global welfare are close to the non-cooperative values. However, altruism narrows the gap between the individually optimal emissions and the socially optimal emissions, so altruism increases global welfare. Our model suggests that altruism is a substitute rather than a complement for large climate coalitions.

JEL classification: C72, D64, Q54, Q58

Keywords: climate coalition, climate policy, moral behaviour, social norms

* University of Hagen, Department of Economics, Universitätsstr. 41, 58097 Hagen, Germany, phone: 004923319872449, fax: 004923319874143, email: mark.schopf@fernuni-hagen.de.

1 Introduction

The Paris Agreement, negotiated by 196 parties at the 2015 United Nations Climate Change Conference, aims to limit global warming to well below 2 degrees Celsius compared to pre-industrial levels (UN, 2015). Although there is thus broad consensus on the international goal of climate policy, a continuation of current policies would result in global warming of about 3 degrees Celsius above pre-industrial levels (UN, 2022). Consequently, the Paris Agreement with its nationally determined contributions does not reflect an international environmental agreement with globally optimal contributions. On the other hand, some world regions have introduced rather high carbon prices despite facing negative social costs of carbon (see Table 1). Although these carbon prices are still well below the global social cost of carbon (418\$/tCO_2 from Ricke et al., 2018), this behaviour can hardly be explained with perfect selfishness. Instead, it may reflect the important effects of altruistic values on environmental behaviour found in the psychological literature (see, e.g., Dietz et al., 2005; Steg, 2016; Lades et al., 2021).

Table 1: Largest carbon pricing schemes representing 22% of global CO_2 emissions.

\$/tCO_2	EU	GBR	CAN	USA	KOR	ZAF	CHN	ARG	MEX	JPN	KAZ	UKR
Price	73	58	38	28	19	10	10	5	4	2	1	1
SCC	−4	−4	−8	48	−1	3	24	3	12	6	−1	−1

Note: Price: The World Bank (2023), SCC: Ricke et al. (2018).

This paper analyses the formation of climate coalitions with altruistic preferences. In particular, each country values, to some extent, every other country's welfare when deciding on its coalition membership and emissions policy. In order to be able to compare our results with the standard literature (Carraro and Siniscalco, 1991; Barrett, 1994), we apply the canonical model of self-enforcing international environmental agreements with concave utility from own emissions and convex costs from global emissions. Without altruistic preferences, this model predicts that climate coalitions are either small or ineffective.[1]

[1]For the linear-quadratic Nash game, Finus (2001, p. 232) finds that climate coalitions consist of no more than three countries. For the linear-quadratic Stackelberg game, Finus (2001, p. 232) finds that climate coalitions are either small or ineffective, and Diamantoudi and Sartzetakis (2006, p. 254) find

We distinguish between the coalition countries taking the fringe countries' emissions as given (Nash game) and taking the reaction of the fringe countries' emissions into account (Stackelberg game) when choosing their own emissions. In both cases, altruism reduces each fringe country's emissions and raises global material welfare, i.e. global welfare in the absence of altruistic preferences. Furthermore, we get the typical results that global emissions decrease and each fringe country's emissions and material welfare increase with the coalition size. By contrast, the effect of altruism on the equilibrium coalition size depends crucially on the game structure.

In the linear-quadratic Nash game, altruism weakly reduces the coalition size, and climate coalitions consist of no more than two countries. The direct effect of altruism, namely smaller global emissions and larger global material welfare for a larger coalition size, makes it worthwhile for all other countries if some country joins the coalition. However, the indirect effect of altruism, namely smaller global emissions and larger global material welfare for a given coalition size, makes it less costly for all other countries if some country does not join the coalition. This indirect effect outweighs the direct effect for small coalition sizes and explains the small climate coalition in equilibrium.

In the linear-quadratic Stackelberg game, altruism weakly raises the coalition size, and climate coalitions can consist of up to six countries. In this case, the coalition countries take advantage of the fringe countries' altruism by becoming less ambitious in the fight against climate change, expecting the fringe countries to react by reducing their emissions more than they would without altruism. However, the coalition countries are not much more ambitious in the Stackelberg equilibrium than in the business-as-usual scenario without coalition formation.

These results suggest that altruism cannot stabilize large and effective climate coalitions. However, altruism narrows the gap between the individually optimal emissions and the socially optimal emissions, so altruism increases global welfare. Altruism thus appears to be more of a substitute than a complement for large climate coalitions.

The economic literature has developed and tested several theories for imperfect self-

that climate coalitions consist of no more than four countries when constraining the parameter space to ensure non-negative emissions.

ishness. In the case of altruistic preferences (Becker, 1974), one can distinguish between pure altruism, i.e. utility from others' utility values (Becker, 1981), paternalistic altruism, i.e. utility from others' consumption bundles (Pollak, 1988), and impure altruism, i.e. utility or warm glow from giving others (Andreoni, 1990). Alger and Weibull (2010) show that pure altruism used in this paper is evolutionary stable, and Andreoni et al. (2010) summarize the significant evidence for altruism in economic experiments. Other theories comprise reciprocal fairness (Rabin, 1993), inequality aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and Kantian behaviour (Alger and Weibull, 2013; Roemer, 2015).

These theories have also been applied in the literature on self-enforcing international environmental agreements. Buchholz et al. (2018) and Nyborg (2018) analyse the effects of reciprocal fairness when countries decide on their membership in the coalition and on their emissions. They find that reciprocal fairness can stabilize the grand coalition, but it can also stabilize an interior coalition that is either weakly larger (Nyborg, 2018) or even weakly smaller (Buchholz et al., 2018) than the interior coalition without reciprocal fairness. Lange and Vogt (2003) incorporate inequality aversion à la Bolton and Ockenfels (2000) into the canonical model of self-enforcing international environmental agreements and find that sufficiently large inequality aversion can stabilize the grand coalition. By contrast, Vogt (2016) applies inequality aversion à la Fehr and Schmidt (1999) and finds no stable coalition without transfers in his numerical model with heterogeneous countries. Recently, Eichner and Pethig (2022) analysed the effects of Kantian or moral behaviour when countries decide on their membership in the coalition and on their emissions. They find that membership moralism expands the climate coalition, and emissions moralism expands the climate coalition only in the presence of membership moralism.

Closest to our paper is van der Pol et al. (2012), who analyse the effects of altruism affecting the membership decision but not the policy decision. They find that this kind of partial altruism expands the climate coalition. We extend their model into different directions. First, we consider altruism on both stages of the game. Second, we analyse not only the Nash game but also the Stackelberg game. Third, while they solve their

model numerically with heterogeneous countries, we solve our model analytically with homogeneous countries. Finally, we replicate their results analytically to discuss the differences from our results.[2]

The remainder of the paper is organized as follows: Section 2 introduces the model, and characterizes the social optimum and the business-as-usual scenario. Section 3 analyses the effects of altruism on the Nash game of coalition formation. This section also includes a comparison with the model of van der Pol et al. (2012). Section 4 analyses the effects of altruism on the Stackelberg game of coalition formation with the coalition countries as Stackelberg leaders and the fringe countries as Stackelberg followers. Section 5 concludes.

2 Model

Consider a model with $n \geq 3$ identical countries. Each county $i \in N$ derives consumption benefits $B(e_i)$ from its emissions e_i, where $B' > 0$ and $B'' < 0$, and faces climate damages $D(e)$ from global emissions $e \coloneqq \sum_{i \in N} e_i$, where $D' > 0$ and $D'' > 0$. Then, each country's material welfare function is $W_i = B(e_i) - D(e)$. Furthermore, each country is altruistic such that it values its own material welfare by 1 and every other country's material welfare by $\alpha \in [0, 1]$.[3] Thus, the altruism parameter $\alpha = 0$ implies perfectly selfish countries, while $\alpha = 1$ implies perfectly altruistic countries. Then, each country's moral welfare function is

$$V_i = W_i + \alpha \sum_{j \in N \setminus i} W_j = (1 - \alpha) W_i + \alpha W, \tag{1}$$

[2]Daube (2019) and Goussebaïle et al. (2023) analyse the effects of altruism on climate policy with multiple countries. Daube (2019) shows that altruistic preferences lead to a partial internalization of the climate externality in the non-cooperative solution, and to a full internalization of the climate externality in the cooperative solution if and only if the altruistic preferences for all countries coincide. Goussebaïle et al. (2023) analyse the effects of altruistic foreign aid on climate change mitigation and find that paying transfers before abating emissions incentivises developing countries to choose efficient climate change mitigation and leads to the social optimum if altruistic preferences are sufficiently large. However, both papers abstract from coalition formation.

[3]Instead, if each country values its own material welfare by 1 and every other country's moral welfare by $\beta \in [0, 1/n]$, then each country's moral welfare function is $V_i = W_i + \beta \sum_{j \in N \setminus i} V_j = \tilde{W}_i + \tilde{\alpha} \sum_{j \in N \setminus i} \tilde{W}_j$ with $\tilde{W}_i = W_i/(1 + \beta)$ and $\tilde{\alpha} = \beta/[1 - \beta(n - 1)] \in [0, 1]$, and our results do not change.

where $W \coloneqq \sum_{i \in N} W_i$ is global material welfare, and the global moral welfare function is

$$V = \sum_{i \in N} \left[W_i + \alpha \sum_{j \in N \setminus i} W_j \right] = [1 + \alpha(n-1)]W, \tag{2}$$

where $V \coloneqq \sum_{i \in N} V_i$ is global moral welfare. Consequently, the socially optimal emissions (SO) are independent of the altruism parameter α, while the individually optimal emissions, i.e. the business-as-usual emissions (BAU), are not (Daube, 2019, Results 4 and 5). In particular, the socially optimal values and the individually optimal values coincide for $\alpha = 1$. In Appendix A.1, we prove that global emissions decrease and global material welfare increases with the altruism parameter in the individually optimal solution. Consequently, the relative global emissions $e^{\mathrm{BAU}}/e^{\mathrm{SO}}$ decrease and the relative global material and moral welfare $W^{\mathrm{BAU}}/W^{\mathrm{SO}} = V^{\mathrm{BAU}}/V^{\mathrm{SO}}$ increase with the altruism parameter.

In the further course of the paper we analyse the two-stage game of self-enforcing environmental agreements. At the first stage of the game, countries decide on their membership in the coalition. Thereby, internal [external] stability implies that no country will leave [join] the coalition if this reduces its moral welfare (D'Aspremont et al., 1983). At the second stage of the game, there is a coalition of m countries, and countries decide on their emissions. Thereby, each fringe country maximizes its moral welfare (1), and each coalition country $i \in M$ maximizes the sum of the coalition countries' moral welfare

$$\sum_{i \in M} V_i = \sum_{i \in M} \left[W_i + \alpha \sum_{j \in N \setminus i} W_j \right] = (1 - \alpha) \sum_{i \in M} W_i + \alpha m W. \tag{3}$$

Comparing (1) and (3), each fringe country's policy weights its own material welfare by $1 - \alpha$ and global material welfare by α, while each coalition country's policy weights the coalition's material welfare by $1 - \alpha$ and global material welfare by αm. In the following we distinguish between two game concepts. In Section 3, we analyse the Nash game, and in Section 4, we analyse the Stackelberg game with the coalition countries as Stackelberg leaders and the fringe countries as Stackelberg followers. The respective game is then solved by backward induction.

3 Nash game

At the second stage of the Nash game, each fringe country $i = f$ maximizes its moral welfare (1) over its emissions e_f, taking the other countries' emissions as given, which yields

$$B'(e_f) = [1 + \alpha(n-1)]D'(e). \tag{4}$$

Each fringe country equates marginal emissions benefits to its own marginal emissions damages $D'(e)$, plus all other countries' marginal emissions damages weighted by the altruism parameter $\alpha(n-1)D'(e)$.

Furthermore, each coalition country $i = c$ maximizes the sum of the coalition countries' moral welfare (3) over its emissions e_c, taking the other countries' emissions as given, which yields[4]

$$B'(e_c) = \frac{1 + \alpha(n-1)}{1 + \alpha(m-1)} mD'(e) \leq nD'(e). \tag{5}$$

For $\alpha = 0$, each coalition country equates marginal emissions benefits to the coalition countries' marginal emissions damages $mD'(e)$. For $\alpha > 0$, altruism implies that each coalition country accounts for all other countries' marginal emissions damages via $1 + \alpha(n-1)$, but it also implies that all other coalition countries account for each coalition country's marginal emissions benefits via $1 + \alpha(m-1)$. Note that $B'(e_f) = B'(e_c) = nD'(e)$ for $\alpha = 1$, so the Nash equilibrium and the social optimum then coincide. In the following we focus on $\alpha \in [0, 1)$.

From (4) and (5), we infer

$$\frac{B'(e_c)}{B'(e_f)} = \frac{m}{1 + \alpha(m-1)} \in (1, m]. \tag{6}$$

Consequently, each fringe country's emissions are greater than each coalition country's

[4]The second-order conditions are fulfilled.

emissions. In Appendix A.2.1, we prove[5]

Proposition 1 (Comparison of Nash equilibrium and BAU)**.**

- $e_c < e_i^{BAU} < e_f$ *and* $e < e^{BAU}$,
- $V_f > V_c$,
- $W_f > W_c, W_i^{BAU}$.

(6) implies that the coalition countries are ceteris paribus more ambitious in the fight against climate change than at BAU. This results in smaller coalition country's emissions and global emissions, which raises the free-rider incentives and leads to greater fringe country's emissions. Each fringe country's emissions being greater than each coalition country's emissions implies $V_f > V_c$ and $W_f > W_c$. Finally, global emissions being smaller and each fringe country's emissions being greater than at BAU implies $W_f > W_i^{\mathrm{BAU}}$ and, thus, $V_f > V_i^{\mathrm{BAU}}$ if $W_c \geq W_i^{\mathrm{BAU}}$ or if α is sufficiently small.

To prepare the analysis of the first stage of the Nash game, we prove in Appendix A.2.2

Lemma 1 (Effects of coalition size and altruism on emissions and welfare)**.**

- $\frac{\mathrm{d}e_f}{\mathrm{d}m} > 0$, $\frac{\mathrm{d}e}{\mathrm{d}m} < 0$ *and* $\frac{\mathrm{d}W_f}{\mathrm{d}m} > 0$,
- $\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} < 0$, $\frac{\mathrm{d}e}{\mathrm{d}\alpha} < 0$ *and* $\frac{\mathrm{d}W}{\mathrm{d}\alpha} > 0$.

From the first bullet of the lemma, we get the typical results that each fringe country's emissions increase but global emissions decrease with the coalition size, so free-rider incentives tend to increase as the coalition gets larger. The resulting higher consumption benefits and lower climate damages imply that each fringe country's material welfare increases with the coalition size and, thus, that V_f increases with the coalition size if W_c increases with the coalition size or if α is sufficiently small.

The second bullet of the lemma reveals that each fringe country's emissions and global emissions decrease with the altruism parameter and that global material welfare increases with the altruism parameter. Consequently, the relative global emissions e/e^{SO} decrease

[5]Furthermore, we there prove that global emissions are larger at the Nash equilibrium than at the social optimum.

and the relative global material and moral welfare $W/W^{\text{SO}} = V/V^{\text{SO}}$ increase with the altruism parameter.

Now we turn to the first stage of the Nash game. First note that $V_f(m) > V_c(m)$ from Proposition 1 implies that if a coalition is externally unstable, i.e. $V_c(m+1) \geq V_f(m)$, then the corresponding expansion of the coalition is accompanied by a Pareto improvement, i.e. $V_f(m+1) > V_c(m+1) \geq V_f(m) > V_c(m)$. For the detailed stability analysis, we use the following linear-quadratic specification

$$B(e_i) = ae_i - \frac{b}{2}e_i^2, \quad D(e) = \frac{d}{2}e^2. \tag{7}$$

We constrain the parameter space to ensure non-negative emissions for $m \in [2, n]$, which gives an upper bound for d/b. In Appendix A.2.3, we then prove

Proposition 2 (Stability of coalitions with policy altruism)**.**
Consider the linear-quadratic specification (7) *and suppose altruism affects the membership decision and the policy decision.*

- *Either the coalition $m = 2$ is stable or no coalition is stable.*
- *The coalition $m = 2$ is stable for $\alpha = 0$ and $n \geq 12$ (sufficient).*
- *The coalition size weakly decreases with α.*

We use a numerical example to demonstrate that there are economies in which $m = 2$ is not stable for $\alpha > 0$ and $n \geq 12$. Figure 1 depicts each coalition country's minimal emissions[6] (left-hand side figure) and the internal stability condition for $m = 2$ (right-hand side figure) dependent on α. In the numerical example, each coalition country's emissions are positive for all $m \in [2, n]$. Furthermore, $m = 2$ becomes unstable for $\alpha \geq 0.334$. Thus, there are economies in which $m = 2$ is not stable for $\alpha > 0$ and $n \geq 12$.

Proposition 2 and Figure 1 show that altruism does not stabilize larger coalitions, but even destabilizes small coalitions. This is in stark contrast to the numerical analysis of van der Pol et al. (2012), who find that the coalition size increases with the altruism parameter and that the grand coalition becomes stable for $\alpha \geq 0.401$ (with a uniform

[6] Using $e_c(m(\alpha), \alpha)$ with $m(\alpha) = \arg\min e_c(m, \alpha)$.

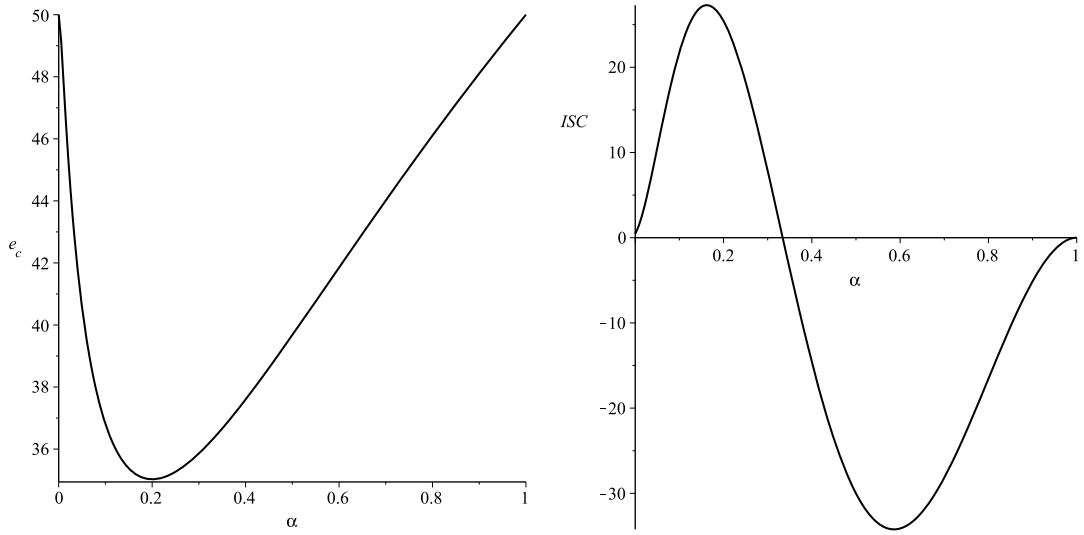

Figure 1: Each coalition country's minimal emissions (left-hand side figure) and the internal stability condition for $m = 2$ (right-hand side figure) dependent on α with $n = 100, a = 100, b = 1$ and $d = 1/10000$.

altruism parameter and without transfers). The major difference between their model and our model is that we assume altruistic preferences at both stages of the game, while they assume altruistic preferences only at the first stage of the game. At the second stage of the game, they assume that each fringe country maximizes its material welfare, while each coalition country maximizes the sum of the coalition countries' material welfare.[7] This does not alter the qualitative results at the second stage of the game, i.e. Proposition 1 and the first bullet of Lemma 1 also hold for $\alpha = 0$. However, it alters the qualitative effects of altruism on the internal stability condition. In both models, this internal stability condition reads

$$V_c(m) - V_f(m-1) = (1-\alpha)W_c(m) + \alpha W(m) \\ - \big[(1-\alpha)W_f(m-1) + \alpha W(m-1)\big] \geq 0. \tag{8}$$

[7]van der Pol et al. (2012) argue that "agents may hold different preferences when acting in different social situations, for example as consumers or as citizens." They then distinguish between the economic policy decision and the political membership decision. In our view, there is no qualitative difference between the social situations at the two stages of the game, and both decisions are made by citizens rather than by consumers, i.e. by a homo politicus with "subjective social welfare functions" rather than by a homo economicus with "personal well-being functions" (Nyborg, 2000). However, different payoff functions at different stages of the game may be justified by strategic delegation between the membership decision and the policy decision (Spycher and Winkler, 2022).

In van der Pol et al. (2012), where the policy is independent of α, altruism stabilizes coalitions if and only if

$$\frac{\partial\left[V_c(m)-V_f(m-1)\right]}{\partial\alpha} = \left[(m-1)W_c(m)+(n-m)W_f(m)\right] - \left[(m-1)W_c(m-1)+(n-m)W_f(m-1)\right] > 0. \tag{9}$$

This direct effect of altruism is positive if and only if the total material welfare of the other countries decreases when a country leaves the coalition. Then, altruism can induce a country to stay in the coalition even though its own material welfare would increase if it left the coalition. In Appendix A.2.4, we prove that the direct effect is positive for $m = 2$ (and for $m \in [2, n]$ with our linear-quadratic specification), regardless of whether or not altruistic preferences are assumed at the second stage of the game. However, the magnitude of the direct effect differs between the models. Furthermore, in our model, where the policy depends on α, altruism stabilizes coalitions if and only if

$$\frac{\mathrm{d}\left[V_c(m)-V_f(m-1)\right]}{\mathrm{d}\alpha} = \frac{\partial\left[V_c(m)-V_f(m-1)\right]}{\partial\alpha} + (1-\alpha)\frac{\mathrm{d}W_c(m)}{\mathrm{d}\alpha} + \alpha\frac{\mathrm{d}W(m)}{\mathrm{d}\alpha} - \left[(1-\alpha)\frac{\mathrm{d}W_f(m-1)}{\mathrm{d}\alpha} + \alpha\frac{\mathrm{d}W(m-1)}{\mathrm{d}\alpha}\right] > 0. \tag{10}$$

The second line of (10) represents the indirect effect of altruism. It is positive if and only if the policy effect of altruism on a country's moral welfare is greater inside than outside the coalition. In Appendix A.2.4, we prove that the policy effect inside the coalition is positive, i.e. $(1-\alpha)\frac{\mathrm{d}W_c(m)}{\mathrm{d}\alpha} + \alpha\frac{\mathrm{d}W(m)}{\mathrm{d}\alpha} > 0$, but that the policy effect outside the coalition is also positive for $m = 2$ (and for $m \in [2, n-2]$ with our linear-quadratic specification), i.e. $(1-\alpha)\frac{\mathrm{d}W_f(m-1)}{\mathrm{d}\alpha} + \alpha\frac{\mathrm{d}W(m-1)}{\mathrm{d}\alpha} > 0$. Proposition 2 reveals that the latter effect is so strong that altruism raises the free-rider incentives. In other words, the policy effect of altruism is more important for small coalitions than for large coalitions, and so important that the negative indirect effect of altruism outweighs the positive direct effect.

In order to check whether the different results of van der Pol et al. (2012) indeed stem from the different assumption concerning altruistic preferences at the second stage of the

game, and not from some other minor differences between the models, we analyse the first stage of the game without altruistic preferences at the second stage of the game. In Appendix A.2.5, we then prove

Proposition 3 (Stability of coalitions without policy altruism)**.**
Consider the linear-quadratic specification (7) *and suppose altruism affects the membership decision but not the policy decision.*

- *Either some unique coalition* $m \geq 2$ *is stable or no coalition is stable.*
- *The coalition* $m = 2$ *is stable for* $\alpha = 0$ *and* $n \geq 12$ *(sufficient).*
- *The coalition size weakly increases with* α*.*
- *The grand coalition is stable for* $\alpha \geq 4/7$ *(sufficient).*

Proposition 3 confirms the numerical result of van der Pol et al. (2012) that considering altruism only at the first stage of the game stabilizes coalitions. Furthermore, in Appendix A.2.5 we prove that global material and moral welfare then increase with the coalition size. While altruism affecting the membership decision only is beneficial for global welfare and for the climate ($\frac{\mathrm{d}e}{\mathrm{d}m} < 0$ from Lemma 1) because it expands the climate coalition, altruism affecting the membership decision and the policy decision is beneficial for global welfare and for the climate ($\frac{\mathrm{d}W}{\mathrm{d}\alpha} > 0$ and $\frac{\mathrm{d}e}{\mathrm{d}\alpha} < 0$ from Lemma 1) because it tightens the climate policy. If the same coalition is stable in both models, e.g. for $\alpha \to 0$ such that $m = 2$, then global welfare is larger and global emissions are smaller with than without altruistic preferences at the second stage of the game. By contrast, if the grand coalition is stable without altruistic preferences at the second stage of the game, e.g. for $\alpha \geq 4/7$, then global welfare is larger and global emissions are smaller without than with altruistic preferences at the second stage of the game.

Figure 2 depicts these relationships for a numerical example.[8] With [without] altruistic preferences at the second stage of the game, $m = 2$ becomes unstable for $\alpha > 0.334$ [$m = n$ becomes stable for $\alpha > 0.5$]. Global emissions are smaller and global material welfare is larger with than without altruistic preferences at the second stage of the game if and

[8] Approximating the coalition size by $m = \arg[V_c(m) - V_f(m-1) = 0]$ without altruistic preferences at the second stage of the game.

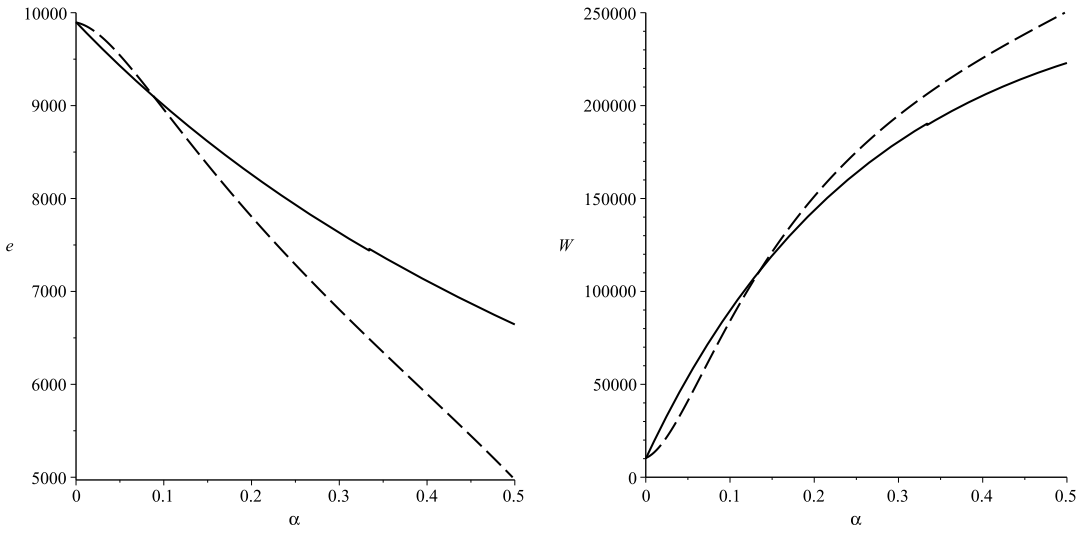

Figure 2: Global emissions (left-hand side figure) and global material welfare (right-hand side figure) with (solid curves) and without (dashed curves) altruistic preferences at the second stage of the game dependent on α with $n = 100, a = 100, b = 1$ and $d = 1/10000$.

only if $\alpha < 0.089$ and $\alpha < 0.135$, respectively. Then, the tighter climate policy outweighs the larger climate coalition, which then comprises no more than 30 and 40 out of 100 countries, respectively. Finally, the figure shows that the welfare difference is relatively small (< 27000) compared to the welfare difference between social optimum (250000) and BAU without altruistic preferences (9800).

4 Stackelberg game

At the second stage of the Stackelberg game, each fringe country $i = f$ maximizes its moral welfare (1) over its emissions e_f, taking the other countries' emissions as given, which yields (4).

Furthermore, each coalition country $i = c$ maximizes the sum of the coalition countries' moral welfare (3) over its emissions e_c, taking the other coalition countries' emissions as given, but taking (4) into account, which yields[9]

$$B'(e_c) = \frac{1+\alpha(n-1)}{1+\alpha(m-1)} mD'(e) \left[1 + (1-\alpha)\tfrac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right] \leq nD'(e), \tag{11}$$

[9]The second-order conditions are fulfilled if $B''' \geq 0$ and $D''' \leq 0$ (see Appendix A.3.1).

where

$$\frac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c} = -\frac{(n-m)[1+\alpha(n-1)]D''(e)}{(n-m)[1+\alpha(n-1)]D''(e) - B''(e_f)} \in (-1,0). \tag{12}$$

For $\alpha = 0$, each coalition country equates marginal emissions benefits to the coalition countries' marginal emissions damages $mD'(e)$, corrected for the leakage rate to the fringe countries $\left|\frac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right|$. For $\alpha > 0$, altruism implies that each coalition country accounts for all other countries' marginal emissions damages via $1+\alpha(n-1)$, but it also implies that all other coalition countries account for each coalition country's marginal emissions benefits via $1 + \alpha(m-1)$. Furthermore, altruism implies that each coalition country accounts for all fringe countries' marginal emissions benefits, which reduces the influence of the leakage rate to the fringe countries via $1 - \alpha$. Finally, altruism of the fringe countries implies that these countries react more sensitive to other countries' emissions changes, such that the altruism parameter ceteris paribus increases the leakage rate to the fringe countries. Note that $B'(e_f) = B'(e_c) = nD'(e)$ for $\alpha = 1$, so the Stackelberg equilibrium and the social optimum then coincide. In the following we focus on $\alpha \in [0,1)$.

From (4) and (11), we infer

$$\frac{B'(e_c)}{B'(e_f)} = \frac{m}{1+\alpha(m-1)}\left[1 + (1-\alpha)\frac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right] =: \tilde{\theta} \in (0,m). \tag{13}$$

Consequently, each fringe country's emissions are greater [smaller] than each coalition country's emissions for $\tilde{\theta} > [<]1$. Furthermore, $\tilde{\theta} = 1$ implies that the Stackelberg equilibrium and the BAU coincide. In Appendix A.3.2, we prove[10]

Proposition 4 (Comparison of Stackelberg equilibrium and BAU)**.**

- $e_c \gtreqless e_i^{BAU} \gtreqless e_f$ *and* $e \gtreqless e^{BAU}$ *for* $\tilde{\theta} \lesseqgtr 1$,
- $V_c > V_i^{BAU} > V_f$ *and* $V < V^{BAU}$ *for* $\tilde{\theta} < 1$,

  $V_c = V_i^{BAU} = V_f$ *and* $V = V^{BAU}$ *for* $\tilde{\theta} = 1$,

  $V_f > V_c > V_i^{BAU}$ *and* $V > V^{BAU}$ *for* $\tilde{\theta} > 1$,

[10]Furthermore, we there prove that global emissions are larger at the Stackelberg equilibrium than at the social optimum.

- $W_c > W_i^{BAU} > W_f$ *and* $W < W^{BAU}$ *for* $\tilde{\theta} < 1$,
  $W_c = W_i^{BAU} = W_f$ *and* $W = W^{BAU}$ *for* $\tilde{\theta} = 1$,
  $W_f > W_c, W_i^{BAU}$ *and* $W > W^{BAU}$ *for* $\tilde{\theta} > 1$.

$\tilde{\theta} > [<]1$ implies that the coalition countries are ceteris paribus more [less] ambitious in the fight against climate change than at BAU. This results in smaller [greater] coalition country's emissions and global emissions, which raises [reduces] the free-rider incentives and leads to greater [smaller] fringe country's emissions. The coalition could always choose $\tilde{\theta} = 1$, such that $\tilde{\theta} \neq 1$ implies $V_c > V_i^{\text{BAU}}$. For $\tilde{\theta} > [<]1$ global emissions being smaller [greater] and each fringe country's emissions being greater [smaller] than at BAU implies $V_f \gtreqless V_i^{\text{BAU}} \iff W_f \gtreqless W_i^{\text{BAU}} \iff \tilde{\theta} \gtreqless 1$. Furthermore, for $\tilde{\theta} > [<]1$ global emissions being smaller [greater] than at BAU implies $V \gtreqless V^{\text{BAU}} \iff W \gtreqless W^{\text{BAU}} \iff \tilde{\theta} \gtreqless 1$. Finally, for $\tilde{\theta} > [<]1$ each fringe country's emissions being greater [smaller] than each coalition country's emissions implies $V_f \gtreqless V_c \iff W_f \gtreqless W_c \iff \tilde{\theta} \gtreqless 1$.

The partial derivative of $\tilde{\theta}$ with respect to m is positive, so the coalition countries tend to become more ambitious as the coalition gets larger. Then, the leakage rate to the fringe countries ceteris paribus becomes smaller, which tends to increase $\tilde{\theta}$, see (12). Furthermore, the coalition countries' marginal emissions damages then become greater, which outweighs the greater coalition countries' marginal emissions benefits and increases $\tilde{\theta}$, see (13). In Appendix A.3.3, we prove

Proposition 5 (Relation between coalition size and coalition's ambition)**.**
Suppose $B''' \geq 0$ *and* $D''' \leq 0$. *Then,* $m \lesseqgtr \tilde{m} \iff \tilde{\theta} \lesseqgtr 1$, *where*

$$\tilde{m} := \frac{n[1+\alpha(n-1)]D''(e^{BAU}) - B''(e_i^{BAU})}{[1+\alpha(n-1)]D''(e^{BAU}) - B''(e_i^{BAU})} \in (1,n) \tag{14}$$

and where $\frac{\mathrm{d}\tilde{m}}{\mathrm{d}\alpha} > 0$ *for* $B''' = 0$ *(sufficient).*

Thus, the coalition countries are less [more] ambitious than the fringe countries in small [large] coalitions, in which the leakage effect outweighs [is outweighed by] the marginal emissions damage effect. The partial derivative of $\tilde{m}$ with respect to α is positive, so the respective threshold coalition $\tilde{m}$ tends to get larger as countries become more

altruistic. In other words, the coalition countries tend to become less ambitious in the fight against climate change compared to the fringe countries. On the one hand, the altruism parameter ceteris paribus increases the importance of all other coalition countries' marginal emissions benefits for the optimal policy, and it increases the leakage rate to the fringe countries. On the other hand, it ceteris paribus increases the importance of all fringe countries' marginal emissions benefits for the optimal policy. Proposition 5 reveals that the former effect outweighs the latter with linear-quadratic consumption benefits.

Since $m \geq \tilde{m}$ will turn out to be the relevant coalition size and to prepare the analysis of the first stage of the Stackelberg game, we prove in Appendix A.3.4

Lemma 2 (Effects of coalition size and altruism on emissions and welfare for $m \geq \tilde{m}$)**.** *Suppose* $B''' \geq 0$ *and* $D''' \leq 0$*.*

- $\frac{\mathrm{d}e_f}{\mathrm{d}m} > 0$, $\frac{\mathrm{d}e}{\mathrm{d}m} < 0$ *and* $\frac{\mathrm{d}V_c}{\mathrm{d}m}, \frac{\mathrm{d}V_f}{\mathrm{d}m}, \frac{\mathrm{d}W_f}{\mathrm{d}m} > 0$,
- $\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} < 0$ *and* $\frac{\mathrm{d}W}{\mathrm{d}\alpha} > 0$.

From the first bullet of the lemma, we get the typical results that each fringe country's emissions increase but global emissions decrease with the coalition size, so free-rider incentives tend to increase as the coalition gets larger as in the Nash game. Contrary to the Nash game, the resulting lower climate damages ensure that not only each fringe country's moral welfare but also each coalition country's moral welfare increases with the coalition size. Finally, each fringe country's material welfare increases with the coalition size because its consumption benefits increase and the climate damages decrease with the coalition size.

The second bullet of the lemma reveals that each fringe country's emissions decrease with the altruism parameter and that global material welfare increases with the altruism parameter as in the Nash game. Consequently, the relative global material and moral welfare $W/W^{\mathrm{SO}} = V/V^{\mathrm{SO}}$ increase with the altruism parameter. Contrary to the Nash game, global emissions need not decrease with the altruism parameter.

Now we turn to the first stage of the Stackelberg game. First note that Proposition 4 implies that all coalitions with $\tilde{\theta} \leq 1$ are externally unstable because joining this coalition then increases the respective country's moral welfare from $V_f \leq V_i^{\mathrm{BAU}}$ to $V_c > V_i^{\mathrm{BAU}}$.

Consequently, global emissions are smaller and each country's moral welfare is greater at the stable Stackelberg equilibrium than at BAU, and each fringe country's welfare is greater than each coalition country's welfare. Together with Proposition 5, this gives[11]

Lemma 3 (Instability of small coalitions)**.**
Suppose $B''' \geq 0$ *and* $D''' \leq 0$*. Then, all coalitions* $m \leq \tilde{m}$ *are externally unstable, and the coalition* $m = \lfloor \tilde{m} + 1 \rfloor \geq 2$ *is internally stable.*

The coalition $m = \lfloor \tilde{m} + 1 \rfloor \geq 2$ is internally stable because leaving the coalition decreases the respective country's moral welfare from $V_c > V_i^{\text{BAU}}$ to $V_f \leq V_i^{\text{BAU}}$. The lemma indicates that the coalition size increases with the threshold coalition $\tilde{m}$, which in turn tends to increase with the altruism parameter from Proposition 5. Via this mechanism, altruism could stabilize larger coalitions.

For the detailed stability analysis, we use the linear-quadratic specification (7). We constrain the parameter space to ensure non-negative emissions for $m \in [2, n]$, which gives an upper bound for d/b similar to the Nash game. In Appendix A.3.5, we then prove

Proposition 6 (Stability of coalitions)**.**
Consider the linear-quadratic specification (7) *with* $n \geq 7$*.*

- *Some unique coalition* $m \in (\tilde{m}, \tilde{m} + 2)$ *is stable.*
- *Some unique coalition* $m \in \{2, 3\}$ *is stable for* $\alpha = 0$*.*
- *The coalition size weakly increases with* α*.*
- *Some unique coalition* $m \in \{2, 3, 4, 5, 6\}$ *is stable for* $\alpha > 0$*.*

Contrary to the Nash game, Proposition 6 reveals that altruism stabilizes larger coalitions. However, the coalition never comprises more than six countries. More importantly, the coalition is always smaller than $m = \tilde{m} + 2$. Since the Stackelberg equilibrium and BAU coincide for $m = \tilde{m}$, the emissions-reducing and welfare-enhancing effects of the coalition size from Lemma 2 are negligible. In fact, the small coalitions stem from constraining the parameter space to ensure non-negative emissions for $m \in [2, n]$, which gives

[11] The function $\lfloor \cdot \rfloor$ maps its argument to the largest weakly smaller integer.

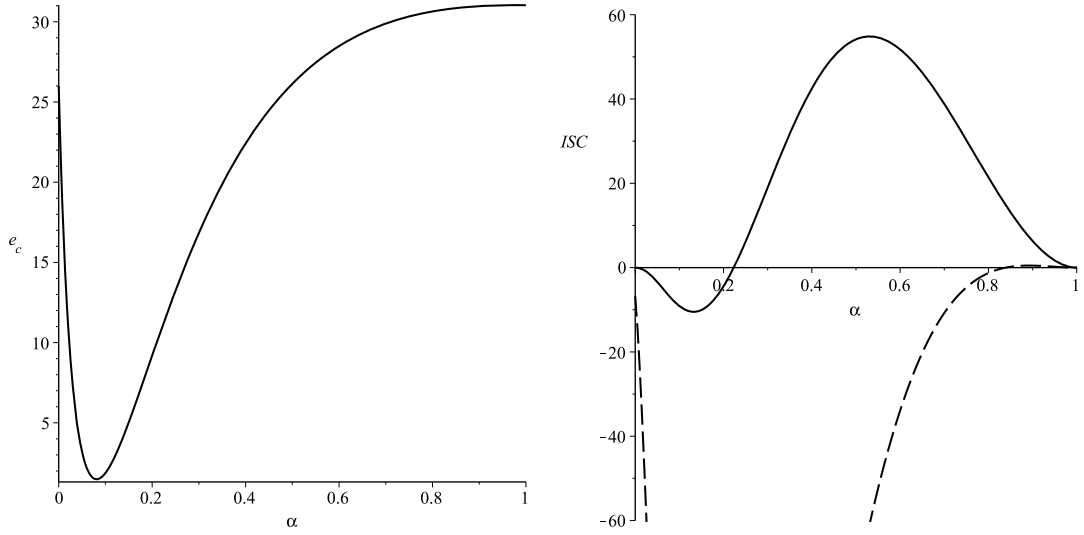

Figure 3: Each coalition country's minimal emissions (left-hand side figure) and the internal stability condition (right-hand side figure) for $m = 3$ (solid curve) and for $m = 4$ (dashed curve) dependent on α with $n = 100, a = 100, b = 1$ and $d = 1/4500$.

an upper bound for d/b and, thus, for $\tilde{m}$. From Proposition 5, this upper bound increases with the altruism parameter, which is the driving force for larger coalitions with than without altruism.

We use a numerical example to demonstrate there are economies in which the coalition is larger with than without altruism. Figure 3 depicts each coalition country's minimal emissions[12] (left-hand side figure) and the internal stability condition (right-hand side figure) for $m = 3$ (solid curve) and for $m = 4$ (dashed curve) dependent on α. In the numerical example, each coalition country's emissions are positive for all $m \in [2, n]$. Furthermore, $m = 3$ becomes stable for $\alpha \geq 0.223$, and $m = 4$ becomes stable for $\alpha \geq 0.839$. Thus, there are economies in which the coalition is larger with than without altruism. Finally, Figure 4 shows that global emissions decrease and global material welfare increases with the altruism parameter in the numerical example. Furthermore, as the coalition gets larger at $\alpha = 0.223$ and at $\alpha = 0.839$, global emissions jump downwards and global material welfare jumps upwards, but these jumps are (almost) not visible.

[12]Using $e_c(m(\alpha), \alpha)$ with $m(\alpha) = \arg\min e_c(m, \alpha)$.

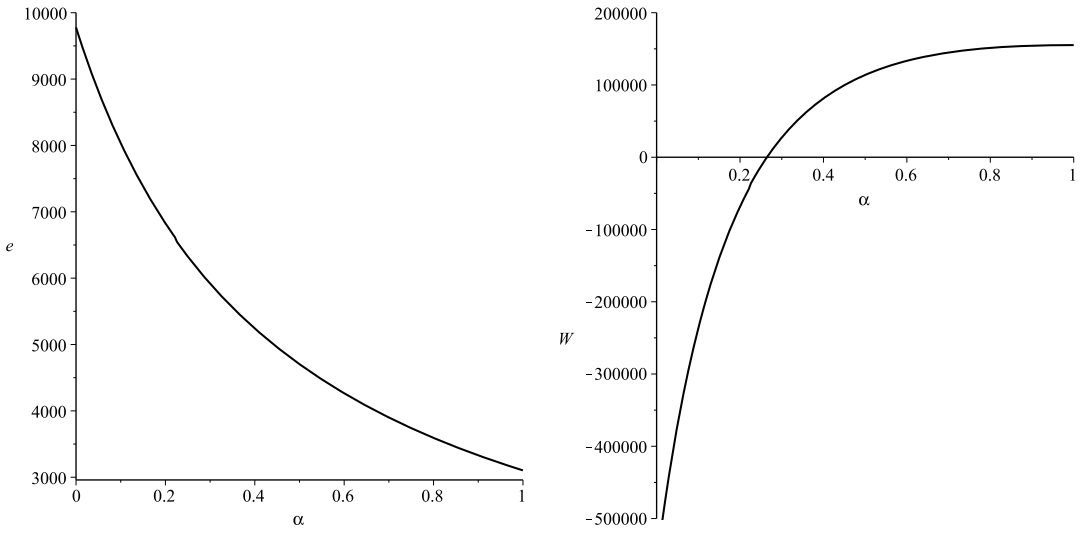

Figure 4: Global emissions (left-hand side figure) and global material welfare (right-hand side figure) dependent on α with $n = 100, a = 100, b = 1$ and $d = 1/4500$.

5 Conclusion

This paper analyses the effects of altruism on the formation of climate coalitions in the standard two-stage game of self-enforcing international environmental agreements. In the Nash [Stackelberg] game, altruism weakly decreases [increases] the coalition size. However, the coalition never comprises more than six countries, and the corresponding global emissions and global welfare are close to the non-cooperative values. Nevertheless, altruism reduces global emissions and raises global welfare by narrowing the gap between the individually optimal values and the socially optimal values. Altruism thus appears to be more of a substitute than a complement for large climate coalitions. Consequently, altruism may help explain why countries are willing to internalize their climate externalities onto other countries, but are unwilling to conclude a large and effective climate agreement.

Our analysis can be extended in several directions. For example, one could analyse the optimal strategic delegation of each country's principal to a country's agent with different altruistic preferences between the first and the second stage of the game (Spycher and Winkler, 2022). Furthermore, it may be interesting to replace the assumption of pure altruism with the assumption of paternalistic or impure altruism. In the first case, one could consider different altruistic parameters for other countries' consumption benefits

and climate damages. In the second case, one could add warm-glow transfers between countries at a third stage of the game. Finally, the results at the first stage of the Nash game and the Stackelberg game depend on the functional forms of the benefit function and the damage function, such that it may be interesting to replace our linear-quadratic specification with, e.g., an isoelastic specification (Nkuiya, 2020). These issues are beyond the scope of the present paper but may represent interesting and important tasks for future research.

References

Alger, Ingela and Jörgen W Weibull (2010), 'Kinship, incentives, and evolution', *The American Economic Review* **100**(4), 1725–1758.

Alger, Ingela and Jörgen W Weibull (2013), 'Homo moralis—preference evolution under incomplete information and assortative matching', *Econometrica* **81**(6), 2269–2302.

Andreoni, James (1990), 'Impure altruism and donations to public goods: A theory of warm-glow giving', *The Economic Journal* **100**(401), 464–477.

Andreoni, James, William T Harbaugh and Lise Vesterlund (2010), Altruism in experiments, *in* Steven N Durlauf and Lawrence E Blume, eds, 'Behavioural and Experimental Economics', The New Palgrave Economics Collection, Palgrave Macmillan London, pp. 6–13.

Barrett, Scott (1994), 'Self-enforcing international environmental agreements', *Oxford Economic Papers* **46**, 878–894.

Becker, Gary S (1974), 'A theory of social interactions', *Journal of Political Economy* **82**(6), 1063–1093.

Becker, Gary S (1981), 'Altruism in the family and selfishness in the market place', *Economica* **48**(189), 1–15.

Bolton, Gary E and Axel Ockenfels (2000), 'ERC: A theory of equity, reciprocity, and competition', *The American Economic Review* **91**(1), 166–193.

Buchholz, Wolfgang, Wolfgang Peters and Aneta Ufert (2018), 'International environmental agreements on climate protection: A binary choice model with heterogeneous agents', *Journal of Economic Behavior & Organization* **154**, 191–205.

Carraro, Carlo and Domenico Siniscalco (1991), 'Strategies for the international protection of the environment', *CEPR Discussion Paper No. 568*.

D'Aspremont, Claude, Alexis Jacquemin, Jean J Gabszewicz and John A Weymark (1983), 'On the stability of collusive price leadership', *The Canadian Journal of Economics / Revue canadienne d'économique* **16**(1), 17–25.

Daube, Marc (2019), 'Altruism and global environmental taxes', *Environmental and Resource Economics* **73**(4), 1049–1072.

Diamantoudi, Effrosyni and Eftichios S Sartzetakis (2006), 'Stable international environmental agreements: An analytical approach', *Journal of Public Economic Theory* **8**(2), 247–263.

Dietz, Thomas, Amy Fitzgerald and Rachael Shwom (2005), 'Environmental values', *Annual Review of Environment and Resources* **30**, 335–372.

Eichner, Thomas and Rüdiger Pethig (2022), 'International environmental agreements when countries behave morally', CESifo *Working Paper No. 10090*.

Fehr, Ernst and Klaus M Schmidt (1999), 'A theory of fairness, competition, and cooperation', *The Quarterly Journal of Economics* **114**(3), 817–868.

Finus, Michael (2001), *Game theory and international environmental cooperation*, Edward Elgar, Cheltenham, UK and Northampton, MA, USA.

Goussebaïle, Arnaud, Antoine Bommier, Amélie Goerger and Jean-Philippe Nicolaï (2023), 'Altruistic foreign aid and climate change mitigation', *Environmental and Resource Economics* **84**(1), 219–239.

Lades, Leonhard K, Kate Laffan and Till O Weber (2021), 'Do economic preferences predict pro-environmental behaviour?', *Ecological Economics* **183**, 106977.

Lange, Andreas and Carsten Vogt (2003), 'Cooperation in international environmental negotiations due to a preference for equity', *Journal of Public Economics* **87**(9-10), 2049–2067.

Nkuiya, Bruno (2020), 'Stability of international environmental agreements under isoelastic utility', *Resource and Energy Economics* **59**, 101128.

Nyborg, Karine (2000), 'Homo economicus and homo politicus: Interpretation and aggregation of environmental values', *Journal of Economic Behavior & Organization* **42**(3), 305–322.

Nyborg, Karine (2018), 'Reciprocal climate negotiators', *Journal of Environmental Economics and Management* **92**, 707–725.

Pollak, Robert A (1988), 'Tied transfers and paternalistic preferences', *The American Economic Review* **78**(2), 240–244.

Rabin, Matthew (1993), 'Incorporating fairness into game theory and economics', *The American Economic Review* **83**(5), 1281–1302.

Ricke, Katharine, Laurent Drouet, Ken Caldeira and Massimo Tavoni (2018), 'Country-level social cost of carbon', *Nature Climate Change* **8**(10), 895–900.

Roemer, John E (2015), 'Kantian optimization: A microfoundation for cooperation', *Journal of Public Economics* **127**, 45–57.

Spycher, Sarah and Ralph Winkler (2022), 'Strategic delegation in the formation of modest international environmental agreements', *European Economic Review* **141**, 103963.

Steg, Linda (2016), 'Values, norms, and intrinsic motivation to act proenvironmentally', *Annual Review of Environment and Resources* **41**, 277–292.

The World Bank (2023), 'Carbon pricing dashboard', Online at: `https://carbonpricingdashboard.worldbank.org`.

UN (2015), 'Paris agreement', Online at: `https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement`.

UN (2022), 'Emissions gap report', Online at: `https://www.unep.org/resources/emissions-gap-report-2022`.

van der Pol, Thomas, Hans-Peter Weikard and Ekko van Ierland (2012), 'Can altruism stabilise international climate agreements?', *Ecological Economics* **81**, 112–120.

Vogt, Carsten (2016), 'Climate coalition formation when players are heterogeneous and inequality averse', *Environmental and Resource Economics* **65**(1), 33–59.

A Appendix

A.1 Business-as-usual scenario

The first-order condition of (1) reads

$$B_i' - [1 + \alpha(n-1)]D' = 0, \tag{A.1}$$

and the second-order condition reads

$$B_i'' - [1 + \alpha(n-1)]D'' < 0, \tag{A.2}$$

which is fulfilled. Differentiating (A.1) with respect to α yields

$$\begin{aligned} & B_i'' \tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} - [1 + \alpha(n-1)]D'' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} - (n-1)D' = 0 \\ \Leftrightarrow \quad & \tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} = [1 + \alpha(n-1)]D''/B_i'' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} + (n-1)D'/B_i''. \end{aligned} \tag{A.3}$$

Taking the sum over all $i \in N$ and rearranging yields

$$\begin{aligned} & \sum_{i \in N} \tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} = \sum_{i \in N} [1 + \alpha(n-1)]D''/B_i'' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} + \sum_{i \in N} (n-1)D'/B_i'' \\ \Leftrightarrow \quad & \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} = \frac{\sum_{i \in N}(n-1)D'/B_i''}{1 - \sum_{i \in N}[1 + \alpha(n-1)]D''/B_i''} < 0. \end{aligned} \tag{A.4}$$

Substituting into (A.3) yields

$$\tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} = \frac{(n-1)D'/B_i''}{1 - \sum_{i \in N}[1 + \alpha(n-1)]D''/B_i''} < 0. \tag{A.5}$$

Finally, differentiating W with respect to α and using (A.1) yields

$$\begin{aligned} \tfrac{\partial W}{\partial \alpha} &= \sum_{i \in N} B_i' \tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} - \sum_{i \in N} D' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} = \sum_{i \in N} [1 + \alpha(n-1)]D' \tfrac{\mathrm{d}e_i}{\mathrm{d}\alpha} - \sum_{i \in N} D' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} \\ &= -(1-\alpha)(n-1)D' \tfrac{\mathrm{d}e}{\mathrm{d}\alpha} > 0. \end{aligned} \tag{A.6}$$

A.2 Nash game

A.2.1 Proof of Proposition 1

From (4), (5) and $\Theta := \frac{m}{1+\alpha(m-1)}$, the equilibrium is characterized by

$$B'(e_f) = [1 + \alpha(n-1)]D', \tag{A.7}$$

$$B'(e_c) = [1 + \alpha(n-1)]\Theta D', \tag{A.8}$$

$$e = me_c + (n-m)e_f. \tag{A.9}$$

First differentiating (A.7), (A.8) and (A.9) with respect to Θ yields

$$B''(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} = [1+\alpha(n-1)]D''\tfrac{\mathrm{d}e}{\mathrm{d}\Theta}, \tag{A.10}$$

$$B''(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta} = [1+\alpha(n-1)]\left[\Theta D''\tfrac{\mathrm{d}e}{\mathrm{d}\Theta} + D'\right], \tag{A.11}$$

$$\tfrac{\mathrm{d}e}{\mathrm{d}\Theta} = m\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta} + (n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta}. \tag{A.12}$$

Solving for $\frac{\mathrm{d}e}{\mathrm{d}\Theta}$, $\frac{\mathrm{d}e_f}{\mathrm{d}\Theta}$ and $\frac{\mathrm{d}e_c}{\mathrm{d}\Theta}$ yields

$$\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} = \frac{m[1+\alpha(n-1)]^2 D''D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\Theta B''(e_f)]D''} > 0, \tag{A.13}$$

$$\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta} = -\frac{[1+\alpha(n-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\Theta B''(e_f)]D''} < 0, \tag{A.14}$$

$$\tfrac{\mathrm{d}e}{\mathrm{d}\Theta} = \frac{m[1+\alpha(n-1)]B''(e_f)D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\Theta B''(e_f)]D''} < 0. \tag{A.15}$$

Note that $\Theta = 1 \iff e_f = e_c = e_i^{\mathrm{BAU}}$. Thus, $\Theta > 1 \implies e_f > e_i^{\mathrm{BAU}} > e_c \wedge e^{\mathrm{BAU}} > e$.

Second differentiating V_f, W_f, $V_f - V_c$ and $W_f - W_c$ with respect to Θ and using (4), (5), (A.13), (A.14) and (A.15) yields

$$\begin{aligned}
&\tfrac{\mathrm{d}V_f}{\mathrm{d}\Theta} = [1+\alpha(n-m-1)]B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} + \alpha m B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta} - [1+\alpha(n-1)]D'\tfrac{\mathrm{d}e}{\mathrm{d}\Theta} \\
&= [1+\alpha(n-1)]D'\left\{(1-\alpha)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} + \alpha\left[(n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} + m\Theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta}\right] - \tfrac{\mathrm{d}e}{\mathrm{d}\Theta}\right\} \\
&= \frac{\alpha m[1+\alpha(n-1)]^2\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\Theta B''(e_f)]D''}\left\{\frac{m}{1+\alpha(m-1)}\right. \\
&\quad \left. + \frac{(1-\alpha)\{[1-\alpha(m-1)(n-m-1)][1+\alpha(n-1)]D'' - B''(e_f)\}}{\alpha[1+\alpha(m-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} - \Theta\right\},
\end{aligned} \tag{A.16}$$

$$\begin{aligned}
&\tfrac{\mathrm{d}W_f}{\mathrm{d}\Theta} = B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} - D'\tfrac{\mathrm{d}e}{\mathrm{d}\Theta} = D'\left\{[1+\alpha(n-1)]\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} - \tfrac{\mathrm{d}e}{\mathrm{d}\Theta}\right\} \\
&= \frac{m[1+\alpha(n-1)]\{[1+\alpha(n-1)]^2 D'' - B'(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\Theta B''(e_f)]D''} > 0,
\end{aligned} \tag{A.17}$$

$$\tfrac{\mathrm{d}(V_f - V_c)}{\mathrm{d}\Theta} = (1-\alpha)\tfrac{\mathrm{d}(W_f - W_c)}{\mathrm{d}\Theta} = (1-\alpha)\left[B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\Theta} - B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\Theta}\right] > 0. \tag{A.18}$$

(A.16) yields $\frac{\mathrm{d}V_f}{\mathrm{d}\Theta} > 0$ for $\alpha \leq \frac{1}{(m-1)(n-m-1)}$ $\left(\geq \frac{4}{(n-2)^2}\right)$ and $\Theta \leq \frac{m}{1+\alpha(m-1)}$, which implies $V_f > V_i^{\mathrm{BAU}}$ for $\alpha \leq \frac{4}{(n-2)^2}$. (A.17) implies $W_f > W_i^{\mathrm{BAU}}$. Finally, (A.18) implies $V_f > V_c$ and $W_f > W_c$.

Third suppose $e \leq e^{\mathrm{SO}}$. Then, the right-hand sides of (4) and (5) would be smaller than $nD'(e^{\mathrm{SO}})$, such that the left-hand sides would have to be smaller than $B'(e_i^{\mathrm{SO}})$, implying $e_c, e_f < e_i^{\mathrm{SO}}$ and contradicting $e \leq e^{\mathrm{SO}}$. Thus, $e > e^{\mathrm{SO}}$. □

A.2.2 Proof of Lemma 1

Totally differentiating (4), (5) and $e = me_c + (n-m)e_f$ yields

$$B''(e_f)\,\mathrm{d}e_f = [1+\alpha(n-1)]D''\,\mathrm{d}e + (n-1)D'\,\mathrm{d}\alpha, \tag{A.19}$$

$$\begin{aligned} B''(e_c)\,\mathrm{d}e_c = {} & \frac{1+\alpha(n-1)}{1+\alpha(m-1)} mD''\,\mathrm{d}e + \frac{(1-\alpha)[1+\alpha(n-1)]}{[1+\alpha(m-1)]^2} D'\,\mathrm{d}m \\ & + \frac{n-m}{[1+\alpha(m-1)]^2} mD'\,\mathrm{d}\alpha, \end{aligned} \tag{A.20}$$

$$\mathrm{d}e = m\,\mathrm{d}e_c + (n-m)\,\mathrm{d}e_f + (e_c - e_f)\,\mathrm{d}m. \tag{A.21}$$

Solving for $\mathrm{d}e_f$, $\mathrm{d}e_c$ and $\mathrm{d}e$ yields

$$\begin{aligned} & \Lambda\,\mathrm{d}e_f \\ & = \{(1-\alpha)[1+\alpha(n-1)]mD' - (e_f - e_c)[1+\alpha(m-1)]^2 B''(e_c)\}[1+\alpha(n-1)]D''\,\mathrm{d}m \\ & \quad - \{m^2(m-1)[1+\alpha(n-1)]^2 D'' - (n-1)[1+\alpha(m-1)]^2 B''(e_c)\}D'\,\mathrm{d}\alpha, \end{aligned} \tag{A.22}$$

$$\begin{aligned} & \Lambda\,\mathrm{d}e_c \\ & = \{(1-\alpha)\{[1+\alpha(n-1)](n-m)D'' - B''(e_f)\}D' + (e_f - e_c)[1+\alpha(m-1)]mD'' \\ & \quad \cdot B''(e_f)\}[1+\alpha(n-1)]\,\mathrm{d}m - m(n-m)\{[1+\alpha(n-1)]^2(m-1)D'' + B''(e_f)\}D'\,\mathrm{d}\alpha, \end{aligned} \tag{A.23}$$

$$\begin{aligned} & \Lambda\,\mathrm{d}e \\ & = \{(1-\alpha)[1+\alpha(n-1)]mD' - (e_f - e_c)[1+\alpha(m-1)]^2 B''(e_c)\}B''(e_f)\,\mathrm{d}m \\ & \quad + (n-m)\{[1+\alpha(m-1)]^2(n-1)B''(e_c) + m^2 B''(e_f)\}D'\,\mathrm{d}\alpha, \end{aligned} \tag{A.24}$$

where

$$\begin{aligned} \Lambda \coloneqq {} & -[1+\alpha(m-1)]\{[1+\alpha(m-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}B''(e_c) \\ & + m^2[1+\alpha(n-1)]D''B''(e_f)\} > 0. \end{aligned}$$

First (A.22) [(A.24)] yields $\frac{\mathrm{d}e_f}{\mathrm{d}m} > 0$ and $\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} < 0$ [$\frac{\mathrm{d}e}{\mathrm{d}m} < 0$ and $\frac{\mathrm{d}e}{\mathrm{d}\alpha} < 0$].

Second differentiating V_f and W_f with respect to m and using (4), (5), (A.22), (A.23) and (A.24) yields

$$\begin{aligned} \frac{\mathrm{d}V_f}{\mathrm{d}m} & = [1+\alpha(n-1)]D'\left\{(1-\alpha)\frac{\mathrm{d}e_f}{\mathrm{d}m} + \alpha\left[(n-m)\frac{\mathrm{d}e_f}{\mathrm{d}m} + \frac{m^2}{1+\alpha(m-1)}\frac{\mathrm{d}e_c}{\mathrm{d}m}\right] - \frac{\mathrm{d}e}{\mathrm{d}m}\right\} \\ & = \frac{[1+\alpha(n-1)]D'}{[1+\alpha(m-1)]\Lambda}\{(1-\alpha)[1+\alpha(n-1)]2\{[1-(m-1)(n-m-1)\alpha][1+\alpha(n-1)]D'' \\ & \quad - B''(e_f)\}mD' - (e_f - e_c)[1+\alpha(m-1)]\{[1+\alpha(n-m-1)][1+\alpha(m-1)]^2[1+\alpha(n \end{aligned}$$

$$-1)]D''B''(e_c) - [1+\alpha(m-1)]^2 B''(e_f)B''(e_c) + \alpha[1+\alpha(n-1)]m^3 D''B''(e_f)\}\}, \tag{A.25}$$

$$\frac{\mathrm{d}W_f}{\mathrm{d}m} = D'\left\{[1+\alpha(n-1)]\frac{\mathrm{d}e_f}{\mathrm{d}m} - \frac{\mathrm{d}e}{\mathrm{d}m}\right\} > 0, \tag{A.26}$$

such that $\frac{\mathrm{d}V_f}{\mathrm{d}m} > 0$ for $\alpha \leq \frac{1}{(m-1)(n-m-1)} \left(\geq \frac{4}{(n-2)^2}\right)$.

Third differentiating W with respect to α and using (4), (5) and (A.21) yields

$$\begin{aligned}\frac{\mathrm{d}W}{\mathrm{d}\alpha} &= D'\left\{[1+\alpha(n-1)]\left[(n-m)\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} + \frac{m^2}{1+\alpha(m-1)}\frac{\mathrm{d}e_c}{\mathrm{d}\alpha}\right] - n\frac{\mathrm{d}e}{\mathrm{d}\alpha}\right\} \\ &= -\frac{(1-\alpha)(n-m)}{1+\alpha(m-1)}D'\left\{(m-1)[1+\alpha(n-1)]\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} + \frac{\mathrm{d}e}{\mathrm{d}\alpha}\right\} > 0. \end{aligned} \tag{A.27}$$

□

A.2.3 Proof of Proposition 2

The linear-quadratic equilibrium is defined by

$$a - be_f = [1+\alpha(n-1)]de, \tag{A.28}$$

$$a - be_c = \frac{1+\alpha(n-1)}{1+\alpha(m-1)}mde, \tag{A.29}$$

$$e = (n-m)e_f + me_c. \tag{A.30}$$

Solving for e_f, e_c and e yields

$$e_f = \frac{1+\alpha(m-1)+m(m-1)(1-\alpha)[1+\alpha(n-1)]\frac{d}{b}}{\Omega}\frac{a}{b} > 0, \tag{A.31}$$

$$e_c = \frac{1+\alpha(m-1)+(n-m)(m-1)(1-\alpha)[1+\alpha(n-1)]\frac{d}{b}}{\Omega}\frac{a}{b}, \tag{A.32}$$

$$e = \frac{1}{\Omega}\frac{na}{b} > 0, \tag{A.33}$$

where

$$\Omega \coloneqq [1+\alpha(m-1)] + [1+\alpha(n-1)]\{[1+\alpha(m-1)]n + (1-\alpha)m(m-1)\}\frac{d}{b} > 0.$$

Note that $\frac{\partial^2 e_c\Omega}{\partial m^2} > 0$. For $\alpha = 0$, $e_c\Omega$ is minimal at $m = \frac{n-1}{2}$, and then $e_c\Omega$ is non-negative if and only if $\frac{d}{b} \leq \frac{4}{(n-1)^2}$, which is thus an upper bound for $\frac{d}{b}$.

Using (A.31), (A.32) and (A.33) yields

$$V_f = [1+\alpha(n-m-1)]\left[ae_f - \frac{b}{2}e_f^2 - \frac{d}{2}e^2\right] + \alpha m\left[ae_c - \frac{b}{2}e_c^2 - \frac{d}{2}e^2\right]$$

$$= \frac{a^2}{2b\Omega^2}[1+\alpha(n-1)]\{[1+\alpha(m-1)]^2 + (1-\alpha)^2[1+\alpha(n-1)]m(m-1)\{m^2+2n - m - \alpha(m-1)(n^2-nm-2n+m)\}\frac{d}{b} - [1+\alpha(m-1)][n^2-2m^2-2n+2m + \alpha(n^2m-2nm^2-3n^2+4m^2+4n-4m) - 2\alpha^2(n-1)(n-m)(m-1)]\left(\frac{d}{b}\right)^2\}, \tag{A.34}$$

$$V_c = [1+\alpha(n-m)]\left[ae_f - \frac{b}{2}e_f^2 - \frac{d}{2}e^2\right] + [1+\alpha(m-1)]\left[ae_c - \frac{b}{2}e_c^2 - \frac{d}{2}e^2\right]$$
$$= V_f - \frac{a^2}{2b\Omega^2}n^2(m-1)(1-\alpha)^2[1+\alpha(n-1)]^2[m+1+\alpha(m-1)]\left(\frac{d}{b}\right)^2. \tag{A.35}$$

The internal stability condition reads

$$V_c(m) - V_f(m-1) = \frac{a^2n^2(m-1)(1-\alpha)^2[1+\alpha(n-1)]^2\left(\frac{d}{b}\right)^2}{2b\Omega(m)^2\Omega(m-1)^2}\Phi(m), \tag{A.36}$$

where

$$\begin{aligned}\Phi(m) \coloneqq & -[1+\alpha(m-1)][m-3+\alpha(m-2)^2] - [1+\alpha(n-1)]\{2[(n-m)(m-1)+(m \\ & -3)^3+6(m-3)^2+11(m-3)+4] + 2\alpha\{(n-m)[2(m-2)^2+5(m-2)+1] \\ & +(m-2)[(m-2)^3+4(m-2)^2+6(m-2)+2]\} + \alpha^2(m-2)[2(n-m)(m^2 \\ & +m-3)+(m-1)(m-2)] + 2\alpha^3(n-m)(m-1)(m-2)^2\}\frac{d}{b} - [1+\alpha(n-1)]^2 \\ & \cdot\{[n(m+1)+m(m-1)^2][n+m(m-3)] + \alpha\{(n-m)^2[2(m-2)^2+10(m-2) \\ & +3]+2(n-m)[(m-2)^4+7(m-2)^3+16(m-2)^2+14(m-2)+2]+m^2(m \\ & -2)^2\} + \alpha^2(n-m)(m-2)\{(n-m)[(m-2)^2+9(m-2)+6]+2(m-2)^3+8 \\ & \cdot(m-2)^2+12(m-2)+4\} + \alpha^3(n-m)(2n-2m+1)(m-1)(m-2)^2\}\left(\frac{d}{b}\right)^2 \\ & <0 \impliedby m \geq 3,\end{aligned} \tag{A.37}$$

such that all coalitions $m \geq 3$ are internally unstable, which proves the first bullet of the proposition. Furthermore,

$$\begin{aligned}\Phi(2)|_{\alpha=0} &= 1 - 2(n-4)\frac{d}{b} - (n-2)(3n+2)\left(\frac{d}{b}\right)^2 \\ &= \left[1 - \frac{(n-1)^2}{4}\frac{d}{b}\right]^2 + \frac{(n-12)^2+18(n-12)+89}{2}\left[1-\frac{(n-1)^2}{4}\frac{d}{b}\right]\frac{d}{b} \\ &\quad + \frac{(n-12)^4+36(n-12)^3+438(n-12)^2+1860(n-12)+817}{16}\left(\frac{d}{b}\right)^2 \\ &> 0 \impliedby n \geq 12,\end{aligned} \tag{A.38}$$

such that $m = 2$ is internally stable for $\alpha = 0$ and $n \geq 12$, which proves the second bullet of the proposition. Finally,

$$\frac{\Phi(2)}{1+\alpha} = 1 - \frac{2[n-4+\alpha(n-2)]}{(1+\alpha)/[1+\alpha(n-1)]}\frac{d}{b} - \frac{(n-2)[3n+2+\alpha(3n-2)]}{(1+\alpha)/[1+\alpha(n-1)]^2}\left(\frac{d}{b}\right)^2, \quad \text{(A.39)}$$

where

$$\frac{\partial\left(\frac{\Phi(2)}{1+\alpha}\right)}{\partial\left(\frac{d}{b}\right)} < 0 \iff n \geq 4, \quad \text{(A.40)}$$

$$\begin{aligned}\frac{\partial\left(\frac{\Phi(2)}{1+\alpha}\right)}{\partial\alpha} = &-\frac{2(n-2)[(n+1)(3n-4)+(2+\alpha)\alpha(n-1)(3n-2)]}{(1+\alpha)^2/[1+\alpha(n-1)]}\left(\frac{d}{b}\right)^2 \\ &- \frac{2(n-2)[n-3+(2+\alpha)\alpha(n-1)]}{(1+\alpha)^2}\frac{d}{b} < 0 \iff n \geq 3, \quad \text{(A.41)}\end{aligned}$$

$$\begin{aligned}\frac{\partial\left(\frac{\Phi(2)}{1+\alpha}\right)}{\partial n} = &-\frac{2[3n-2+6\alpha(n^2-n-1)+\alpha^2(6n^2-15n+8)]}{(1+\alpha)/[1+\alpha(n-1)]}\left(\frac{d}{b}\right)^2 \\ &- \frac{2[1+2\alpha(n-2)+\alpha^2(2n-3)]}{(1+\alpha)}\frac{d}{b} < 0 \iff n \geq 2, \quad \text{(A.42)}\end{aligned}$$

such that $m = 2$ is internally stable if $\frac{d}{b}$, α and n are sufficiently small, which proves the third bullet of the proposition. □

A.2.4 Effects of altruism on the internal stability condition

From (9), the direct effect of altruism on the internal stability condition reads

$$\begin{aligned}&\frac{\partial\left[V_c(m)-V_f(m-1)\right]}{\partial\alpha} \\ &= \left[(m-1)W_c(m)+(n-m)W_f(m)\right] - \left[(m-1)W_c(m-1)+(n-m)W_f(m-1)\right], \quad \text{(A.43)}\end{aligned}$$

which is positive if $(m-1)\frac{\mathrm{d}W_c(m)}{\mathrm{d}m} + (n-m)\frac{\mathrm{d}W_f(m)}{\mathrm{d}m} > 0$. Using (4), (5) and (A.21) yields

$$\begin{aligned}&(m-1)\frac{\mathrm{d}W_c(m)}{\mathrm{d}m} + (n-m)\frac{\mathrm{d}W_f(m)}{\mathrm{d}m} \\ &= (m-1)\left[B'(e_c)\frac{\mathrm{d}e_c}{\mathrm{d}m} - D'\frac{\mathrm{d}e}{\mathrm{d}m}\right] + (n-m)\left[B'(e_f)\frac{\mathrm{d}e_f}{\mathrm{d}m} - D'\frac{\mathrm{d}e}{\mathrm{d}m}\right] \\ &= D'\left\{(m-1)\left[\frac{1+\alpha(n-1)}{1+\alpha(m-1)}m\frac{\mathrm{d}e_c}{\mathrm{d}m} - \frac{\mathrm{d}e}{\mathrm{d}m}\right] + (n-m)\left[[1+\alpha(n-1)]\frac{\mathrm{d}e_f}{\mathrm{d}m} - \frac{\mathrm{d}e}{\mathrm{d}m}\right]\right\} \\ &= \frac{(m-1)(e_f-e_c) + (n-m)\left\{[2-m+\alpha(m-1)]\frac{\mathrm{d}e_f}{\mathrm{d}m} - [1+\alpha(n-1)]^{-1}\frac{\mathrm{d}e}{\mathrm{d}m}\right\}}{[1+\alpha(n-1)]^{-1}[1+\alpha(m-1)]/D'}, \quad \text{(A.44)}\end{aligned}$$

such that $\frac{\partial\left[V_c(m)-V_f(m-1)\right]}{\partial\alpha} > 0$ for $m \in \{2, n\}$. Using the linear-quadratic specification (7), it can be shown that $\frac{\partial\left[V_c(m)-V_f(m-1)\right]}{\partial\alpha} > 0$ for $m \in [2, n]$. The corresponding Maple

file is available on request.

From (10), the indirect effects of altruism on the internal stability condition read

$$(1-\alpha)\tfrac{\mathrm{d}W_c(m)}{\mathrm{d}\alpha}+\alpha\tfrac{\mathrm{d}W(m)}{\mathrm{d}\alpha}-\left[(1-\alpha)\tfrac{\mathrm{d}W_f(m-1)}{\mathrm{d}\alpha}+\alpha\tfrac{\mathrm{d}W(m-1)}{\mathrm{d}\alpha}\right]. \tag{A.45}$$

Using (4), (5) and (A.21) yields

$$\begin{aligned}
&(1-\alpha)\tfrac{\mathrm{d}W_c(m)}{\mathrm{d}\alpha}+\alpha\tfrac{\mathrm{d}W(m)}{\mathrm{d}\alpha}\\
&=(1-\alpha)\left[B'(e_c(m))\tfrac{\mathrm{d}e_c(m)}{\mathrm{d}\alpha}-D'\tfrac{\mathrm{d}e(m)}{\mathrm{d}\alpha}\right]\\
&\quad+\alpha\left[mB'(e_c(m))\tfrac{\mathrm{d}e_c(m)}{\mathrm{d}\alpha}+(n-m)B'(e_f(m))\tfrac{\mathrm{d}e_f(m)}{\mathrm{d}\alpha}-nD'\tfrac{\mathrm{d}e(m)}{\mathrm{d}\alpha}\right]\\
&=[1+\alpha(n-1)]D'\left[m\tfrac{\mathrm{d}e_c(m)}{\mathrm{d}\alpha}+\alpha(n-m)\tfrac{\mathrm{d}e_f(m)}{\mathrm{d}\alpha}-\tfrac{\mathrm{d}e(m)}{\mathrm{d}\alpha}\right]\\
&=-(1-\alpha)[1+\alpha(n-1)](n-m)D'\tfrac{\mathrm{d}e_f(m)}{\mathrm{d}\alpha}>0
\end{aligned} \tag{A.46}$$

and

$$\begin{aligned}
&(1-\alpha)\tfrac{\mathrm{d}W_f(m-1)}{\mathrm{d}\alpha}+\alpha\tfrac{\mathrm{d}W(m-1)}{\mathrm{d}\alpha}\\
&=(1-\alpha)\left[B'(e_f(m-1))\tfrac{\mathrm{d}e_f(m-1)}{\mathrm{d}\alpha}-D'\tfrac{\mathrm{d}e(m-1)}{\mathrm{d}\alpha}\right]\\
&\quad+\alpha\left[(m-1)B'(e_c(m-1))\tfrac{\mathrm{d}e_c(m-1)}{\mathrm{d}\alpha}+(n-m+1)B'(e_f(m-1))\tfrac{\mathrm{d}e_f(m-1)}{\mathrm{d}\alpha}-nD'\tfrac{\mathrm{d}e(m-1)}{\mathrm{d}\alpha}\right]\\
&=[1+\alpha(n-1)]D'\left\{[1+\alpha(n-m)]\tfrac{\mathrm{d}e_f(m-1)}{\mathrm{d}\alpha}+\frac{\alpha(m-1)^2}{1+\alpha(m-2)}\tfrac{\mathrm{d}e_c(m-1)}{\mathrm{d}\alpha}-\tfrac{\mathrm{d}e(m-1)}{\mathrm{d}\alpha}\right\}\\
&=-\frac{(1-\alpha)[1+\alpha(n-1)]D'}{1+\alpha(m-2)}\left\{[\alpha(m-2)(n-m)-1]\tfrac{\mathrm{d}e_f(m-1)}{\mathrm{d}\alpha}+\tfrac{\mathrm{d}e(m-1)}{\mathrm{d}\alpha}\right\},
\end{aligned} \tag{A.47}$$

such that $(1-\alpha)\frac{\mathrm{d}W_c(m)}{\mathrm{d}\alpha}+\alpha\frac{\mathrm{d}W(m)}{\mathrm{d}\alpha}>0$ and $(1-\alpha)\frac{\mathrm{d}W_f(2-1)}{\mathrm{d}\alpha}+\alpha\frac{\mathrm{d}W(2-1)}{\mathrm{d}\alpha}=[1+\alpha(n-1)]\frac{\mathrm{d}W_i^{\mathrm{BAU}}}{\mathrm{d}\alpha}>0$ from Appendix A.1. Using the linear-quadratic specification (7), it can be shown that $(1-\alpha)\frac{\mathrm{d}W_f(m-1)}{\mathrm{d}\alpha}+\alpha\frac{\mathrm{d}W(m-1)}{\mathrm{d}\alpha}>0$ for $m\in[2,n-2]$. The corresponding Maple file is available on request.

A.2.5 Proof of Proposition 3

Without altruistic preferences at the second stage of the game, the emissions for a given coalition size are given by substituting $\alpha=0$ into (A.31), (A.32) and (A.33), and the material welfare levels for a given coalition size are given by substituting $\alpha=0$ into (A.34) and (A.35). Using these results, the internal stability condition reads

$$V_c(m)-V_f(m-1)=(1-\alpha)W_c(m)+\alpha W(m)-\left[(1-\alpha)W_f(m-1)+\alpha W(m-1)\right]$$

$$= \frac{(1+2\alpha)n^2(m-1)\frac{a^2}{2b}\left(\frac{d}{b}\right)^2}{\left[1+(m^2+n-3m+2)\frac{d}{b}\right]^2\left[1+(m^2+n-m)\frac{d}{b}\right]^2}\varphi(m), \quad \text{(A.48)}$$

where

$$\begin{aligned}\varphi(m) \coloneqq \frac{\alpha}{1+2\alpha}\Bigg\{&4N_3+3+2[2N_3^2+N_3(2M_2^2+4M_2+13)+2M_2^3+9M_2^2+7M_2+9]\frac{d}{b}\\&+[7N_3^2+N_3[4M_2^3+22M_2^2+22M_2+34]+4M_2^5+23M_2^4+54M_2^3+85M_2^2\\&+50M_2+27]\left(\frac{d}{b}\right)^2\Bigg\}-\Bigg\{m-3+2[(n-m)(m-1)+(m-3)(m^2+2)+4]\frac{d}{b}\\&+[n-m+m(m-2)][(n-m)(m+1)+m(m^2-m+2)]\left(\frac{d}{b}\right)^2\Bigg\}, \quad \text{(A.49)}\end{aligned}$$

where $N_i \coloneqq n-i$ and $M_i \coloneqq m-i$. From (A.37) and (A.49), $\Phi(m)|_{\alpha=0} = \varphi(m)|_{\alpha=0}$, and from the proof of Proposition 2, $\Phi(2)|_{\alpha=0} > 0$ for $n \geq 12$, which proves the second bullet of the proposition. From (A.49), $\varphi(m)$ increases with $\frac{\alpha}{1+2\alpha}$ and, thus, with α, which proves the third bullet of the proposition. Furthermore,

$$\begin{aligned}\varphi(n) = &\frac{2N_2+1}{1+2\alpha}\left[\alpha-\frac{1}{2}+\frac{3}{2(2N_2+1)}\right]\left[1-\frac{(n-1)^2}{4}\frac{d}{b}\right]^2+\frac{10N_2^3+33N_2^2+36N_2+9}{2(1+2\alpha)}\\&\cdot\left[\alpha-\frac{1}{2}+\frac{7N_2^2+22N_2+27}{2(10N_2^3+33N_2^2+36N_2+9)}\right]\left[1-\frac{(n-1)^2}{4}\frac{d}{b}\right]\frac{d}{b}\\&+\frac{50N_2^5+305N_2^4+720N_2^3+790N_2^2+358N_2+17}{16(1+2\alpha)}\\&\cdot\left[\alpha-\frac{4}{7}+\frac{N_3(5N_3^2+24N_3+8)}{7(10N_3^3+55N_3^2+100N_3+56)}\right]\left(\frac{d}{b}\right)^2, \quad \text{(A.50)}\end{aligned}$$

such that $m = n$ is internally stable for $\alpha \geq 4/7$, which proves the third bullet of the proposition. Furthermore,

$$\begin{aligned}\tfrac{\partial\varphi(m)}{\partial m} = &\Bigg\{1+2(n-m+3m^2-7m+3)\frac{d}{b}+[(n-m)^2+(6m^2-6m-2)(n-m)+5m^4\\&-14m^3+14m^2-8m]\left(\frac{d}{b}\right)^2\Bigg\}\Bigg/\Bigg\{m-3+2[(m-1)(n-m)+(m-3)(m^2+2)\\&+4]\frac{d}{b}+[n-m+m(m-2)][(m+1)(n-m)+m(m^2-m+2)]\left(\frac{d}{b}\right)^2\Bigg\}\varphi(m)-\Psi, \quad \text{(A.51)}\end{aligned}$$

where $\Psi > 0$ for $n \geq 4$ and $m \geq 3$. The corresponding Maple file is available on request. $\varphi(\underline{m}) \leq 0$ for some $\underline{m} \geq 3$ implies $\frac{\partial \varphi(m)}{\partial m} < 0$ and, thus, $\varphi(\bar{m}) < 0$ for all $\bar{m} \geq \underline{m}$. Furthermore, $\varphi(m) \geq 0 \iff V_c(m) - V_f(m-1) \geq 0 \iff V_f(m-1) - V_c(m) \leq 0$. Thus, an internally stable coalition m implies an externally unstable coalition $m-1$. Consequently, there is at most one internally and externally stable coalition, which proves the first bullet of the proposition. Finally, note that

$$W = \frac{\{b^2 - b[n(n-2) - 2m(m-1)]d - m(m-1)^2(n-m)d^2\}na^2}{2b[b + (m^2+n-m)d]^2}, \tag{A.52}$$

$$\begin{aligned}\frac{\partial W}{\partial m} = {} & \frac{[(4m-2)(n-m) + (m-1)^2]n^2ad^2}{2[b + (m^2+n-m)d]^2}e_c + \frac{(m-1)^4n^2(n-2)a^2d^3}{2(n-1)^2b[b + (m^2+n-m)d]^3} \\ & \cdot \left[2\left(\frac{n-m}{m-1}\right)^3 + (4n-1)\left(\frac{n-m}{m-1}\right)^2 + (n-1)\left(\frac{n-m}{m-1}\right) + \frac{n^2}{n-2}\right] > 0, \end{aligned} \tag{A.53}$$

such that $[1 + \alpha(n-1)]\frac{\partial W}{\partial m} = \frac{\partial V}{\partial m} > 0$. □

A.3 Stackelberg game

A.3.1 Derivation of (12)

The first-order condition of (3) reads

$$[1 + \alpha(m-1)]\left\{B'(e_c) - \frac{m[1 + \alpha(n-1)]}{1 + \alpha(m-1)}D'\left[1 + \tfrac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right] + \frac{\alpha m(n-m)}{1 + \alpha(m-1)}B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}e_c}\right\} = 0. \tag{A.54}$$

Substituting (4) and rearranging yields (11). Differentiating (4) with respect to e_c yields

$$B''(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}e_c} - [1 + \alpha(n-1)]D''\left[1 + \tfrac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right] = 0. \tag{A.55}$$

Solving for $\frac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}$ yields (12). The second-order condition of (3) reads

$$\begin{aligned}[1 + \alpha(m-1)]\Bigg\{ & B''(e_c) - \frac{m[1 + \alpha(n-1)]}{1 + \alpha(m-1)}D''\left[1 + \tfrac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right]^2 + \frac{\alpha m(n-m)}{1 + \alpha(m-1)}B''(e_f)\left(\tfrac{\mathrm{d}e_f}{\mathrm{d}e_c}\right)^2 \\ & - \frac{m}{1 + \alpha(m-1)}\{[1 + \alpha(n-1)]D' - \alpha B'(e_f)\}\frac{(n-m)[1 + \alpha(n-1)]}{\{(n-m)[1 + \alpha(n-1)]D'' - B''(e_f)\}^2} \\ & \cdot \left[B''(e_f)D'''\left[1 + \tfrac{\mathrm{d}(n-m)e_f}{\mathrm{d}e_c}\right] - D''B'''(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}e_c}\right]\Bigg\} < 0, \end{aligned} \tag{A.56}$$

which is fulfilled if $D''' \leq 0, B''' \geq 0$.

A.3.2 Proof of Proposition 4

From (4), (11) and (12), the equilibrium is characterized by

$$B'(e_f) = [1+\alpha(n-1)]D', \tag{A.57}$$

$$B'(e_c) = [1+\alpha(n-1)]\theta D', \tag{A.58}$$

$$e = me_c + (n-m)e_f. \tag{A.59}$$

First differentiating (A.57), (A.58) and (A.59) with respect to θ yields

$$B''(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} = [1+\alpha(n-1)]D''\tfrac{\mathrm{d}e}{\mathrm{d}\theta}, \tag{A.60}$$

$$B''(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} = [1+\alpha(n-1)]\left[\theta D''\tfrac{\mathrm{d}e}{\mathrm{d}\theta} + D'\right], \tag{A.61}$$

$$\tfrac{\mathrm{d}e}{\mathrm{d}\theta} = m\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} + (n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta}. \tag{A.62}$$

Solving for $\frac{\mathrm{d}e}{\mathrm{d}\theta}$, $\frac{\mathrm{d}e_f}{\mathrm{d}\theta}$ and $\frac{\mathrm{d}e_c}{\mathrm{d}\theta}$ yields

$$\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} = \frac{m[1+\alpha(n-1)]^2 D''D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c)+m\theta B''(e_f)]D''} > 0, \tag{A.63}$$

$$\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} = -\frac{[1+\alpha(n-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c)+m\theta B''(e_f)]D''} < 0, \tag{A.64}$$

$$\tfrac{\mathrm{d}e}{\mathrm{d}\theta} = \frac{m[1+\alpha(n-1)]B''(e_f)D'}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c)+m\theta B''(e_f)]D''} < 0. \tag{A.65}$$

Note that $\theta = 1 \iff e_f = e_c = e_i^{\mathrm{BAU}}$. Thus, $\theta \gtreqless 1 \iff e_f \gtreqless e_i^{\mathrm{BAU}} \gtreqless e_c \wedge e^{\mathrm{BAU}} \gtreqless e$.

Second differentiating V_i and W_i with respect to θ and using (4), (11), (12), (A.63), (A.64) and (A.65) yields

$$\begin{aligned}
\tfrac{\mathrm{d}V_f}{\mathrm{d}\theta} &= [1+\alpha(n-m-1)]B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + \alpha m B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} - [1+\alpha(n-1)]D'\tfrac{\mathrm{d}e}{\mathrm{d}\theta} \\
&= [1+\alpha(n-1)]D'\left\{(1-\alpha)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + \alpha\left[(n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + m\theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta}\right] - \tfrac{\mathrm{d}e}{\mathrm{d}\theta}\right\} \\
&= \frac{\alpha m[1+\alpha(n-1)]^2\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c)+m\theta B''(e_f)]D''} \\
&\quad \cdot \left\{\frac{(1-\alpha)\{[1+\alpha(n-1)]D'' - B''(e_f)\}}{\alpha\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} + 1 - \theta\right\} \qquad \text{(A.66)} \\
&= \frac{\alpha m[1+\alpha(n-1)]^2\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c)+m\theta B''(e_f)]D''} \\
&\quad \cdot \left\{\frac{(1-\alpha)\{[1+\alpha(n-1)]^2 D'' - B''(e_f)\}}{\alpha[1+\alpha(m-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} + \tilde{\theta} - \theta\right\}, \qquad \text{(A.67)} \\
\tfrac{\mathrm{d}V_c}{\mathrm{d}\theta} &= \alpha(n-m)B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + [1+\alpha(m-1)]B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} - [1+\alpha(n-1)]D'\tfrac{\mathrm{d}e}{\mathrm{d}\theta}
\end{aligned}$$

$$= [1+\alpha(n-1)]D'\left\{(1-\alpha)\theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} + \alpha\left[(n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + m\theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta}\right] - \tfrac{\mathrm{d}e}{\mathrm{d}\theta}\right\}$$

$$= \frac{[1+\alpha(m-1)][1+\alpha(n-1)]^2\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2(\tilde{\theta}-\theta)}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\theta B''(e_f)]D''}, \tag{A.68}$$

$$\tfrac{\mathrm{d}V}{\mathrm{d}\theta} = [1+\alpha(n-1)]\tfrac{\mathrm{d}W}{\mathrm{d}\theta} = [1+\alpha(n-1)]\left\{(n-m)B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + mB'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} - nD'\tfrac{\mathrm{d}e}{\mathrm{d}\theta}\right\}$$

$$= [1+\alpha(n-1)]D'\left\{[1+\alpha(n-1)]\left[(n-m)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} + m\theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta}\right] - n\tfrac{\mathrm{d}e}{\mathrm{d}\theta}\right\}$$

$$= \frac{m[1+\alpha(n-1)]^3\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\theta B''(e_f)]D''}$$
$$\cdot\left\{-\frac{(1-\alpha)(n-1)B''(e_f)}{[1+\alpha(n-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} + 1 - \theta\right\} \tag{A.69}$$

$$= \frac{m[1+\alpha(n-1)]^3\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\theta B''(e_f)]D''}$$
$$\cdot\left\{\frac{(1-\alpha)(n-m)\{[1+\alpha(n-1)]^2D'' - B''(e_f)\}}{[1+\alpha(m-1)][1+\alpha(n-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} + \tilde{\theta} - \theta\right\}, \tag{A.70}$$

$$\tfrac{\mathrm{d}W_f}{\mathrm{d}\theta} = B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} - D'\tfrac{\mathrm{d}e}{\mathrm{d}\theta} > 0, \tag{A.71}$$

$$\tfrac{\mathrm{d}W_c}{\mathrm{d}\theta} = B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} - D'\tfrac{\mathrm{d}e}{\mathrm{d}\theta} = D'\left\{[1+\alpha(n-1)]\theta\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta} - \tfrac{\mathrm{d}e}{\mathrm{d}\theta}\right\}$$

$$= \frac{[1+\alpha(n-1)]^2\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}(D')^2}{B''(e_c)B''(e_f) - [1+\alpha(n-1)][(n-m)B''(e_c) + m\theta B''(e_f)]D''}$$
$$\cdot\left\{-\frac{\alpha m(n-m)\{(n-m)[1+\alpha(n-1)]^2D'' - B''(e_f)\}}{[1+\alpha(n-1)]\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} + \tilde{\theta} - \theta\right\}, \tag{A.72}$$

$$\tfrac{\mathrm{d}(V_f - V_c)}{\mathrm{d}\theta} = (1-\alpha)\tfrac{\mathrm{d}(W_f - W_c)}{\mathrm{d}\theta} = (1-\alpha)\left[B'(e_f)\tfrac{\mathrm{d}e_f}{\mathrm{d}\theta} - B'(e_c)\tfrac{\mathrm{d}e_c}{\mathrm{d}\theta}\right] > 0. \tag{A.73}$$

(A.66) [(A.69)] yields $\frac{\mathrm{d}V_f}{\mathrm{d}\theta} > 0$ $\left[\frac{\mathrm{d}V}{\mathrm{d}\theta} > 0 \text{ and } \frac{\mathrm{d}W}{\mathrm{d}\theta} > 0\right]$ for $\theta \leq 1$, which implies $V_f < V_i^{\mathrm{BAU}}$ [$V < V^{\mathrm{BAU}}$ and $W < W^{\mathrm{BAU}}$] for $\tilde{\theta} < 1$. (A.67) [(A.70)] yields $\frac{\mathrm{d}V_f}{\mathrm{d}\theta} > 0$ $\left[\frac{\mathrm{d}V}{\mathrm{d}\theta} > 0 \text{ and } \frac{\mathrm{d}W}{\mathrm{d}\theta} > 0\right]$ for $\theta \leq \tilde{\theta}$, which implies $V_f > V_i^{\mathrm{BAU}}$ [$V > V^{\mathrm{BAU}}$ and $W > W^{\mathrm{BAU}}$] for $\tilde{\theta} > 1$. (A.68) yields $\frac{\mathrm{d}V_c}{\mathrm{d}\theta} \gtreqless 0$ for $\theta \gtreqless \tilde{\theta}$, which implies $V_c > V_i^{\mathrm{BAU}}$ for $\tilde{\theta} \neq 1$. (A.71) implies $W_f \gtreqless W_i^{\mathrm{BAU}}$ for $\tilde{\theta} \gtreqless 1$. (A.72) yields $\frac{\mathrm{d}W_c}{\mathrm{d}\theta} < 0$ for $\theta \geq \tilde{\theta}$, which implies $W_c > W_i^{\mathrm{BAU}}$ for $\tilde{\theta} < 1$. Finally, (A.73) implies $V_f \gtreqless V_c$ and $W_f \gtreqless W_c$ for $\tilde{\theta} \gtreqless 1$.

Third suppose $e \leq e^{\mathrm{SO}}$. Then, the right-hand sides of (4) and (11) would be smaller than $nD'(e^{\mathrm{SO}})$, such that the left-hand sides would have to be smaller than $B'(e_i^{\mathrm{SO}})$, implying $e_c, e_f < e_i^{\mathrm{SO}}$ and contradicting $e \leq e^{\mathrm{SO}}$. Thus, $e > e^{\mathrm{SO}}$. □

A.3.3 Proof of Proposition 5

Totally differentiating (4), (11) and $e = me_c + (n-m)e_f$ yields

$$B''(e_f)\,\mathrm{d}e_f = [1+\alpha(n-1)]D''\,\mathrm{d}e + (n-1)D'\,\mathrm{d}\alpha, \tag{A.74}$$

$$B''(e_c)\,\mathrm{d}e_c = \lambda_e\,\mathrm{d}e - \lambda_{e_f}\,\mathrm{d}e_f + \lambda_m\,\mathrm{d}m + \lambda_\alpha\,\mathrm{d}\alpha, \tag{A.75}$$

$$\mathrm{d}e = m\,\mathrm{d}e_c + (n-m)\,\mathrm{d}e_f + (e_c - e_f)\,\mathrm{d}m, \tag{A.76}$$

where

$$\begin{aligned}
\lambda_e &\coloneqq \frac{1+\alpha(n-1)}{1+\alpha(m-1)} mD'' \frac{\alpha(n-m)[1+\alpha(n-1)]D'' - B''(e_f)}{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)} \\
&\quad + \frac{1+\alpha(n-1)}{1+\alpha(m-1)} mD' \frac{(1-\alpha)(n-m)[1+\alpha(n-1)]B''(e_f)}{[(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2} D''' > 0 \iff D''' \leq 0, \\
\lambda_{e_f} &\coloneqq \frac{1+\alpha(n-1)}{1+\alpha(m-1)} mD' \frac{(1-\alpha)(n-m)[1+\alpha(n-1)]D''}{[(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2} B'''(e_f) \gtreqless 0 \iff B''' \gtreqless 0, \\
\lambda_m &\coloneqq \frac{1+\alpha(n-1)}{[1+\alpha(m-1)]^2} D' \frac{1-\alpha}{[(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2} \{\alpha(n-m)^2[1+\alpha(n-1)]^2 \\
&\quad \cdot [D'']^2 - [n+\alpha(m^2+n-2m)][1+\alpha(n-1)]D''B''(e_f) + [B''(e_f)]^2]\} > 0, \\
\lambda_\alpha &\coloneqq \frac{n-m}{[1+\alpha(m-1)]^2} mD' \frac{1}{[(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2} \{[1+\alpha(n-1)][1+\alpha(m \\
&\quad - 1)][1+2\alpha(n-1)+\alpha^2(n-1)(m-1)][(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]D''(1 \\
&\quad - \tilde{\theta})/(1-\alpha) + [1+\alpha(n-1)][1+\alpha(n-1)(m^2-1)+\alpha(n-1)(m+2)]D''B''(e_f) \\
&\quad + [B''(e_f)]^2\} > 0 \iff \tilde{\theta} \leq 1.
\end{aligned}$$

Solving for $\mathrm{d}e_f$, $\mathrm{d}e_c$ and $\mathrm{d}e$ yields

$$\begin{aligned}
&\lambda\,\mathrm{d}e_f \\
&= [1+\alpha(n-1)]D''[m\lambda_m - (e_f - e_c)B''(e_c)]\,\mathrm{d}m \\
&\quad - \{(n-1)[m\lambda_e - B''(e_c)]D' - m[1+\alpha(n-1)]D''\lambda_\alpha\}\,\mathrm{d}\alpha,
\end{aligned} \tag{A.77}$$

$$\begin{aligned}
&\lambda\,\mathrm{d}e_c \\
&= -\{[1+\alpha(n-1)]D''[(n-m)\lambda_m + (e_c - e_f)\lambda_{e_f}] - B''(e_f)[\lambda_m + (e_c - e_f)\lambda_e]\}\,\mathrm{d}m \\
&\quad + \{(n-1)[(n-m)\lambda_e - \lambda_{e_f}]D' - \{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}\lambda_\alpha\}\,\mathrm{d}\alpha,
\end{aligned} \tag{A.78}$$

$$\begin{aligned}
&\lambda\,\mathrm{d}e \\
&= B''(e_f)[m\lambda_m - (e_f - e_c)B''(e_c)]\,\mathrm{d}m \\
&\quad - \{(n-1)[m\lambda_{e_f} - (n-m)B''(e_c)]D' - mB''(e_f)\lambda_\alpha\}\,\mathrm{d}\alpha,
\end{aligned} \tag{A.79}$$

where

$$\lambda \coloneqq [1+\alpha(n-1)]D''[m\lambda_{e_f} - (n-m)B''(e_c)] - [m\lambda_e - B''(e_c)]B''(e_f) > 0$$
$$\Longleftarrow\ D''' \leq 0, B''' \geq 0.$$

First differentiating (13) with respect to m and using (A.77) and (A.79) yields

$$\begin{aligned}\tfrac{\mathrm{d}\tilde{\theta}}{\mathrm{d}m} = &-\frac{1}{[1+\alpha(n-1)][1+\alpha(m-1)]\lambda\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}^2 D'}\\
&\cdot\{(n-m)\lambda_m\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}\{[1+\alpha(m-1)]\{(n-m)[1+\alpha\\
&\cdot(n-1)]D'' - B''(e_f)\}^2 B''(e_c) + m^2[1+\alpha(n-1)]\{\alpha(n-m)[1+\alpha(n-1)]D''\\
&- B''(e_f)\}D''B''(e_f)\} - (e_c - e_f)(1-\alpha)m(n-m)[1+\alpha(n-1)]^2 B''(e_c)\{[B''(e_f)]^2\\
&\cdot D''' - [1+\alpha(n-1)]D''B'''(e_f)\}D'\},\end{aligned} \tag{A.80}$$

such that $e_c \geq e_f \Longleftrightarrow \tilde{\theta} \leq 1 \implies \frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}m} > 0$ if $D''' \leq 0, B''' \geq 0$. $\tilde{\theta} \leq 1 \implies \frac{\mathrm{d}\tilde{\theta}}{\mathrm{d}m} > 0$ implies that $\tilde{\theta}(\underline{m}) \geq 1 \implies \tilde{\theta}(\bar{m}) > 1$ for $\bar{m} > \underline{m}$. Thus, (13) implicitly defines $\tilde{m}$ with $m \lessgtr \tilde{m} \Longleftrightarrow \tilde{\theta} \lessgtr 1$ if $D''' \leq 0, B''' \geq 0$. Using $\tilde{\theta} = 1$ in (13) and solving for m yields (14).

Second differentiating (14) with respect to α yields

$$\begin{aligned}\tfrac{\mathrm{d}\tilde{m}}{\mathrm{d}\alpha} = &-\frac{(n-1)^2 D''(e^{\mathrm{BAU}})B''(e_i^{\mathrm{BAU}})}{\{[1+\alpha(n-1)]D''(e^{\mathrm{BAU}}) - B''(e_i^{\mathrm{BAU}})\}^2}\\
&-\frac{(n-1)[1+\alpha(n-1)]B''(e_i^{\mathrm{BAU}})D'''(e^{\mathrm{BAU}})}{\{[1+\alpha(n-1)]D''(e^{\mathrm{BAU}}) - B''(e_i^{\mathrm{BAU}})\}^2}\tfrac{\mathrm{d}e^{\mathrm{BAU}}}{\mathrm{d}\alpha}\\
&+\frac{(n-1)[1+\alpha(n-1)]D''(e^{\mathrm{BAU}})B'''(e_i^{\mathrm{BAU}})}{\{[1+\alpha(n-1)]D''(e^{\mathrm{BAU}}) - B''(e_i^{\mathrm{BAU}})\}^2}\tfrac{\mathrm{d}e_i^{\mathrm{BAU}}}{\mathrm{d}\alpha},\end{aligned} \tag{A.81}$$

where $\frac{\mathrm{d}e^{\mathrm{BAU}}}{\mathrm{d}\alpha} = \frac{\mathrm{d}ne_i^{\mathrm{BAU}}}{\mathrm{d}\alpha} < 0$ from Appendix A.1. Thus, $\frac{\mathrm{d}\tilde{m}}{\mathrm{d}\alpha} > 0$ if $D''' \leq 0, B''' \leq 0$. □

A.3.4 Proof of Lemma 2

First (A.77) [(A.79)] yields $\frac{\mathrm{d}e_f}{\mathrm{d}m} > 0$ [$\frac{\mathrm{d}e}{\mathrm{d}m} < 0$] for $e_f \geq e_c \Longleftrightarrow m \geq \tilde{m}$, and (A.78) yields $\frac{\mathrm{d}e_c}{\mathrm{d}m} < 0$ for $e_c \geq e_f \Longleftrightarrow m \leq \tilde{m}$ if $D''' \leq 0, B''' \geq 0$. Furthermore, using (13) in (A.77) yields

$$\begin{aligned}&-(n-1)m\lambda_e|_{D'''=0}D' + m[1+\alpha(n-1)]D''\lambda_\alpha\\
&= -\frac{m^2[1+\alpha(n-1)]D'D''}{(1-\alpha)^2(m-1)[1+\alpha(m-1)][(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2}\{(1-\alpha)^3(n-1)\\
&\cdot[1+\alpha(n-1)](n-m)^2[D'']^2 + (1-\alpha)[(1-\alpha)^2(n+m-1) + 2\alpha(1-\alpha)nm + \alpha^2 nm]\\
&\cdot(n-m)D''[(n-m)[1+\alpha(n-1)]D'' - B''(e_f)](\tilde{\theta}-1) + [1+\alpha(m-1)][(n-m)[1\\
&+\alpha(n-1)]D'' - B''(e_f)]^2(\tilde{\theta}-1)^2\} < 0 \ \Longleftarrow\ \tilde{\theta} \geq 1,\end{aligned}$$

such that $\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} < 0$ for $e_f \geq e_c \iff m \geq \tilde{m}$, and (A.79) yields $\frac{\mathrm{d}e}{\mathrm{d}\alpha} < 0$ for $e_c \geq e_f \iff m \leq \tilde{m}$ if $D''' \leq 0, B''' \geq 0$.

Second differentiating V_i and W_i with respect to m and using (4), (11), (12), (A.77), (A.78) and (A.79) yields

$$\begin{aligned}
\frac{\mathrm{d}V_f}{\mathrm{d}m} &= [1+\alpha(n-1)]D'\left\{(1-\alpha)\frac{\mathrm{d}e_f}{\mathrm{d}m} + \alpha\left[(n-m)\frac{\mathrm{d}e_f}{\mathrm{d}m} + m\theta\frac{\mathrm{d}e_c}{\mathrm{d}m}\right] - \frac{\mathrm{d}e}{\mathrm{d}m}\right\} \\
&= \frac{[1+\alpha(n-1)]D'}{[1+\alpha(m-1)]\lambda\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} \\
&\quad \cdot \{(1-\alpha)m\{[1+\alpha(n-1)]^2 D'' - B''(e_f)\}\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}\lambda_m \\
&\quad + (e_f - e_c)\{\alpha m^2\{\alpha(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}\{[1+\alpha(n-1)]D''\lambda_{e_f} - B''(e_f) \\
&\quad \cdot \lambda_e\}B''(e_f)\} - [1+\alpha(m-1)]\{[1+\alpha(n-m-1)][1+\alpha(n-1)]D'' - B''(e_f)\}\{(n \\
&\quad - m)[1+\alpha(n-1)]D'' - B''(e_f)\}B''(e_c)\}, \qquad \text{(A.82)} \\
\frac{\mathrm{d}V_c}{\mathrm{d}m} &= [1+\alpha(n-1)]D'\left\{(1-\alpha)\theta\frac{\mathrm{d}e_c}{\mathrm{d}m} + \alpha\left[(n-m)\frac{\mathrm{d}e_f}{\mathrm{d}m} + m\theta\frac{\mathrm{d}e_c}{\mathrm{d}m}\right] - \frac{\mathrm{d}e}{\mathrm{d}m}\right\} \\
&= \frac{(e_f - e_c)[1+\alpha(n-1)]D'\{\alpha(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}}{\lambda\{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}} \\
&\quad \cdot \{m\{[1+\alpha(n-1)]D''\lambda_{e_f} - B''(e_f)\lambda_e\} - \{(n-m)[1+\alpha(n-1)]D'' - B''(e_f)\}B''(e_c)\}, \qquad \text{(A.83)} \\
\frac{\mathrm{d}W_f}{\mathrm{d}m} &= D'\left\{[1+\alpha(n-1)]\frac{\mathrm{d}e_f}{\mathrm{d}m} - \frac{\mathrm{d}e}{\mathrm{d}m}\right\}, \qquad \text{(A.84)} \\
\frac{\mathrm{d}W_c}{\mathrm{d}m} &= D'\left\{[1+\alpha(n-1)]\theta\frac{\mathrm{d}e_c}{\mathrm{d}m} - \frac{\mathrm{d}e}{\mathrm{d}m}\right\} \\
&= -\frac{D'}{[1+\alpha(m-1)]\lambda}\{\alpha m(n-m)\{[1+\alpha(n-1)]^2 D'' - B''(e_f)\}\lambda_m + (e_c - e_f)[1+\alpha(m \\
&\quad - 1)]\{[1+\alpha(n-1)]\theta\{[1+\alpha(n-1)]D''\lambda_{e_f} - B''(e_f)\lambda_e\} + B''(e_f)B''(e_c)\}\}, \qquad \text{(A.85)}
\end{aligned}$$

such that $\frac{\mathrm{d}V_f}{\mathrm{d}m} > 0$ for $e_f \geq e_c \iff m \geq \tilde{m}$, and $\frac{\mathrm{d}V_c}{\mathrm{d}m} \gtreqless 0$ for $e_f \gtreqless e_c \iff m \gtreqless \tilde{m}$ if $D''' \leq 0, B''' \geq 0$. Furthermore, $\frac{\mathrm{d}e_f}{\mathrm{d}m} > 0$ and $\frac{\mathrm{d}e}{\mathrm{d}m} < 0$ implies $\frac{\mathrm{d}W_f}{\mathrm{d}m} > 0$ for $m \geq \tilde{m}$, and $e_c \geq e_f$ implies $\frac{\mathrm{d}W_c}{\mathrm{d}m} < 0$ for $m \leq \tilde{m}$.

Third differentiating W with respect to α and using (4), (11), (12), (A.77), (A.78) and (A.79) yields

$$\begin{aligned}
\frac{\mathrm{d}W}{\mathrm{d}\alpha} &= (n-m)\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} + mB'(e_c)\frac{\mathrm{d}e_c}{\mathrm{d}\alpha} - nD'\frac{\mathrm{d}e}{\mathrm{d}\alpha} \\
&= D'\left\{(n-m)[1+\alpha(n-1)]\frac{\mathrm{d}e_f}{\mathrm{d}\alpha} + m[1+\alpha(n-1)]\theta\frac{\mathrm{d}e_c}{\mathrm{d}\alpha} - n\frac{\mathrm{d}e}{\mathrm{d}\alpha}\right\} \\
&= \frac{(n-1)(D')^2}{\lambda}\left\{\frac{(\theta-1)m^2(n-m)^2(1-\alpha)[1+\alpha(n-1)]^3 B''(e_f)D'D'''}{[1+\alpha(m-1)][(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2}\right. \\
&\quad - (1-\alpha)(n-1)(n-m)B''(e_c) + m\{n - [1+\alpha(n-1)]\theta\}\lambda_{e_f} \\
&\quad \left. + \frac{m^2(n-m)(1-\alpha)[1+\alpha(n-1)]^3(D'')^3}{(n-1)(m-1)^3[1+\alpha(m-1)][(n-m)[1+\alpha(n-1)]D'' - B''(e_f)]^2}\right.
\end{aligned}$$

$$\cdot \{(n-1)^2(n-m)^2[1+\alpha(m^2-1)] + (n-1)(n-m)[1+\alpha(m-1)][3n-2m-1+\alpha$$
$$\cdot (n-1)(m^2-1)]\psi + \{3(n-m)^2 + 2(m-1)(n-m) + (m-1)^3 + \alpha(n-1)(m-1)$$
$$\cdot [(m^2-1)(n-m)\alpha + (m-1)^3\alpha + 4(n-m) + 2(m-1)^2]\}\psi^2 + (n-m)[1+\alpha(m$$
$$-1)]\psi^3\}\Bigg\} > 0 \iff \tilde{\theta} \geq 1, D''' \leq 0, B''' \geq 0, \tag{A.86}$$

where $\psi \coloneqq \frac{(\tilde{\theta}-1)\{(n-m)[1+\alpha(n-1)]D''-B''(e_f)\}}{(1-\alpha)[1+\alpha(n-1)]D''}$, such that $\frac{\mathrm{d}W}{\mathrm{d}\alpha} > 0$ for $e_f \geq e_c \iff m \geq \tilde{m}$ if $D''' \leq 0, B''' \geq 0$. □

A.3.5 Prove of Proposition 6

The linear-quadratic equilibrium is defined by

$$a - be_f = [1+\alpha(n-1)]de, \tag{A.87}$$
$$a - be_c = \frac{1+\alpha(n-1)}{1+\alpha(m-1)}mde\left[1-(1-\alpha)\frac{(n-m)[1+\alpha(n-1)]d}{(n-m)[1+\alpha(n-1)]d-b}\right], \tag{A.88}$$
$$e = (n-m)e_f + me_c. \tag{A.89}$$

Solving for e_f, e_c and e yields

$$e_f = \frac{a}{b\omega}\Bigg\{1+\alpha(m-1) + [1+\alpha(n-1)]\{(n-m)[1+\alpha(m-1)] + (1-\alpha)m(m-1)\}\frac{d}{b}$$
$$- m(n-m)(1-\alpha)[1+\alpha(n-1)]^2\left(\frac{d}{b}\right)^2\Bigg\}, \tag{A.90}$$
$$e_c = \frac{a}{b\omega}\Bigg\{1+\alpha(m-1) - [1+\alpha(n-1)](n-m)\{m-2-2\alpha(m-1)\}\frac{d}{b}$$
$$+ (n-m)^2(1-\alpha)[1+\alpha(n-1)]^2\left(\frac{d}{b}\right)^2\Bigg\} > 0 \iff \alpha \geq \frac{1}{2}, \tag{A.91}$$
$$e = \frac{a}{b\omega}n[1+\alpha(m-1)]\Bigg\{1+(n-m)[1+\alpha(n-1)]\frac{d}{b}\Bigg\} > 0, \tag{A.92}$$

where

$$\omega \coloneqq 1+\alpha(m-1) + [1+\alpha(n-1)]\{2(n-m)[1+\alpha(m-1)] + m^2\}\frac{d}{b}$$
$$+ (n-m)[1+\alpha(n-1)]^2\{(1-\alpha)(n-m) + \alpha nm\}\left(\frac{d}{b}\right)^2 > 0.$$

Note that $\frac{\partial^2 e_c\omega}{\partial m^2} > 0$ for $a \leq \frac{1}{2}$. For $\alpha = 0$, $e_c\Omega$ is minimal at $m = \frac{n+2}{2} + \frac{n-2}{2}\frac{d}{b+d}$, and then $e_c\Omega$ is non-negative if and only if $\frac{d}{b} \leq \frac{4}{n(n-4)}$, which is thus an upper bound for $\frac{d}{b}$. It can

be shown that $e_f\Omega$ is positive for $m \in [2, n]$ and $\alpha \in [0, 1]$ if $\frac{d}{b} \leq \frac{4}{n(n-4)}$ and $n \geq 6$. The corresponding Maple file is available on request.

Using (A.90), (A.91) and (A.92) yields

$$\begin{aligned} V_f &= [1+\alpha(n-m-1)]\left[ae_f - \frac{b}{2}e_f^2 - \frac{d}{2}e^2\right] + \alpha m\left[ae_c - \frac{b}{2}e_c^2 - \frac{d}{2}e^2\right] \\ &= V_c + \frac{a^2}{2b\omega^2}n^2(m-\tilde{m})(1-\alpha)^2[1+\alpha(n-1)]^2\left\{1+[1+\alpha(n-1)]\frac{d}{b}\right\}\left\{m+1\right. \\ &\quad \left. +\alpha(m-1)+(n-m)[1+\alpha(n-1)][1+\alpha(2m-1)]\frac{d}{b}\right\}\left(\frac{d}{b}\right)^2, \qquad \text{(A.93)} \\ V_c &= [1+\alpha(n-m)]\left[ae_f - \frac{b}{2}e_f^2 - \frac{d}{2}e^2\right] + [1+\alpha(m-1)]\left[ae_c - \frac{b}{2}e_c^2 - \frac{d}{2}e^2\right] \\ &= \frac{a^2}{2b\omega}[1+\alpha(n-1)]\left\{1+\alpha(m-1)-(n-m)\{n+m-2-2\alpha^2(n-1)(m-1)\right. \\ &\quad \left. +\alpha(nm-3n-3m+4)\}\frac{d}{b}+(n-m)^2(1-\alpha)^2[1+\alpha(n-1)]\left(\frac{d}{b}\right)^2\right\}, \qquad \text{(A.94)} \end{aligned}$$

where $\tilde{m} = \frac{1+n[1+\alpha(n-1)]d/b}{1+[1+\alpha(n-1)]d/b}$.

The internal stability condition reads

$$V_c(m) - V_f(m-1) = \frac{a^2n^2(1-\alpha)^2[1+\alpha(n-1)]^2\left(\frac{d}{b}\right)^2}{2b\omega(m)\omega(m-1)^2}\phi(m), \qquad \text{(A.95)}$$

where

$$\begin{aligned} \phi(m) \coloneqq & -(m-1)[m-3+\alpha(m-2)^2] - (m-1)[1+\alpha(n-1)]\{[(4m^2-12m+4)(n \\ & -m)+m^3-2m^2-4m+2]\alpha^2+[(2m^2-2m-10)(n-m)+2m^3-5m^2-6]\alpha \\ & \cdot(1-\alpha)+[(2(m-3))(n-m)+m^3-3m^2+4m-8](1-\alpha)^2\}\frac{d}{b}-[1+\alpha(n \\ & -1)]^2\{(m-1)[(5m^2-17m+6)(n-m)^2+(3m^3-6m^2-12m+6)(n-m) \\ & +m^3-5m^2+1]\alpha^2+(m-1)[(m^2+2m-17)(n-m)^2+(3m^3-9m^2+4m \\ & -20)(n-m)+2m^3-8m^2+2m-5]\alpha(1-\alpha)+[(m^2-4m+2)(n-m)^2-8 \\ & \cdot(m-1)(n-m)-m^2-4m+5](1-\alpha)^2\}\left(\frac{d}{b}\right)^2+(n-m+1)[1+\alpha(n-1)]^3 \\ & \cdot\{(m-1)[(2m^2-10m+4)(n-m)^2+(2m^3-8m^2-m+2)(n-m)-2m^2 \\ & +m]\alpha^2+(m-1)[(2m-12)(n-m)^2-(2m^2+9)(n-m)-2m^2-1]\alpha(1-\alpha) \\ & -[2(n-m)^2+(m^2-1)(n-m)+m^2-1](1-\alpha)^2\}\left(\frac{d}{b}\right)^3+(n-m)(n-m \\ & +1)^2[1+\alpha(n-1)]^4\{n(2m^2-3m+1)\alpha^2+(3n-2m+1)(m-1)\alpha(1-\alpha)+(n \end{aligned}$$

$$- m)(1-\alpha)^2\} \left(\frac{d}{b}\right)^4. \tag{A.96}$$

Substituting $m = \frac{1}{1+\beta}(\tilde{m}+2)+\frac{\beta}{1+\beta}n$, $n = N_7+7$, $\alpha = \frac{1}{1+\gamma}$, $\frac{d}{b} = \frac{1}{1+\delta}\frac{4}{n(n-4)}$ with $\beta, \gamma, \delta \geq 0$ yields $\phi\left(\frac{1}{1+\beta}(\tilde{m}+2)+\frac{\beta}{1+\beta}n\right) < 0$, which implies $\phi(m) < 0$ for $m \geq \tilde{m}+2$. The corresponding Maple file is available on request. Consequently, all coalitions $m \geq \tilde{m}+2$ are internally unstable, while all coalitions $m \leq \tilde{m}$ are externally unstable from Lemma 3. Suppose $\tilde{m}$ is an integer. Then, $m = \tilde{m}$ is externally unstable and $m = \tilde{m}+2$ is internally unstable, $m = \tilde{m}+1$ is internally and externally stable from Lemma 3. Suppose $\tilde{m}$ is not an integer. Then, $m = \lfloor \tilde{m} \rfloor$ is externally unstable and $m = \lfloor \tilde{m}+3 \rfloor$ is internally unstable, such that $m = \lfloor \tilde{m}+1 \rfloor$ is internally stable from Lemma 3 and $m = \lfloor \tilde{m}+2 \rfloor$ is externally stable. If $m = \lfloor \tilde{m}+1 \rfloor$ is externally stable [unstable], then $m = \lfloor \tilde{m}+2 \rfloor$ is internally stable [unstable], such that some unique coalition $m \in (\tilde{m}, \tilde{m}+2)$ is stable. This proves the first bullet of the proposition. Furthermore, $\frac{\partial \tilde{m}}{\partial (d/b)} = \frac{(n-1)[1+\alpha(n-1)]}{\{1+[1+\alpha(n-1)]d/b\}^2} > 0$, such that substituting $\frac{d}{b} = \frac{4}{n(n-4)}$ into $\tilde{m}$ yields an upper bound $\bar{\tilde{m}}$:

$$\tilde{m} \leq \bar{\tilde{m}} := \frac{n[n+4\alpha(n-1)]}{(n-2)^2+4\alpha(n-1)}, \tag{A.97}$$

where

$$\frac{\partial \bar{\tilde{m}}}{\partial \alpha} = \frac{4n(n-1)^2(n-4)}{[(n-2)^2+4\alpha(n-1)]^2} \gtreqless 0 \iff n \gtreqless 4, \tag{A.98}$$

$$\begin{aligned} \frac{\partial \bar{\tilde{m}}}{\partial n} &= \frac{16\alpha(n-1)^2 - 8\alpha(n-1)(n-2) - 4n(n-2)}{[(n-2)^2+4\alpha(n-1)]^2} \gtreqless 0 \\ &\iff \alpha \gtreqless \frac{n-2+\sqrt{(n-2)(5n-2)}}{4(n-1)} \in [0.576, 0.809]. \end{aligned} \tag{A.99}$$

Thus, $\bar{\tilde{m}}$ is minimal for $\alpha = 0$ and $n = 7$ with $\bar{\tilde{m}} = 1.96$, and it is maximal for $\alpha = 1$ and $n \to \infty$ with $\bar{\tilde{m}} = 5$. $m \in (\tilde{m}, \tilde{m}+2)$ and $\tilde{m} \leq \bar{\tilde{m}}$ then imply $m \in \{2,3\}$ for $\alpha = 0$ and $m \in \{2,3,4,5,6\}$ for $\alpha > 0$. This proves the second bullet of the proposition and the fourth bullet of the proposition, respectively. Furthermore,

$$\begin{aligned} \phi(3)|_{\alpha=0} &= -8\frac{d}{b} + (n^2+10n-23)\left(\frac{d}{b}\right)^2 + 2(n-1)^2(n-2)\left(\frac{d}{b}\right)^3 + (n-2)^2(n-3)^2\left(\frac{d}{b}\right)^4 \\ &= -8\frac{d}{b}\left[1-\frac{n(n-4)}{4}\frac{d}{b}\right]^3 - (5N_{26}^2+226N_{26}+2519)\left(\frac{d}{b}\right)^2\left[1-\frac{n(n-4)}{4}\frac{d}{b}\right]^2 \\ &\quad - 0.5(2N_{26}^4+174N_{26}^3+5587N_{26}^2+78148N_{26}+399316)\left(\frac{d}{b}\right)^3\left[1-\frac{n(n-4)}{4}\frac{d}{b}\right] \\ &\quad - 0.0625(N_{26}^6+122N_{26}^5+5951N_{26}^4+145224N_{26}^3+1778248N_{26}^2+8867520N_{26} \\ &\quad + 2064240)\left(\frac{d}{b}\right)^4, \end{aligned} \tag{A.100}$$

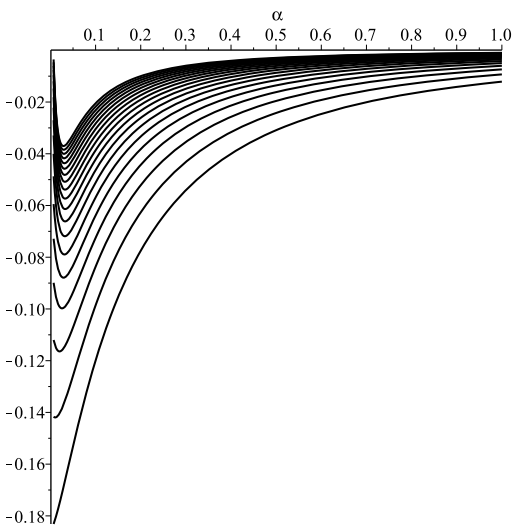

Figure 5: Derivative of $\underline{d/b}$ with respect to α for $n \in [7, 25]$.

such that $m = 3$ is internally unstable for $\alpha = 0$ and $n \geq 26$. Finally,

$$\begin{aligned}\phi(3) = &-2\alpha - 2[1+\alpha(n-1)][4+(2n-11)\alpha+(2n-6)\alpha^2]\frac{d}{b} + [1+\alpha(n-1)]^2[n^2+10n \\ &- 23 + (2n^2-28n+68)\alpha - (3n^2-24n+29)\alpha^2]\left(\frac{d}{b}\right)^2 + (n-2)[1+\alpha(n-1)]^3 \\ &\cdot [2(n-1)^2 + (8n^2-10n-20)\alpha + (n-3)(6n-26)\alpha^2]\left(\frac{d}{b}\right)^3 + (n-3)(n-2)^2 \\ &\cdot [1+\alpha(n-1)]^4[n-3+4(n-1)\alpha+(5n+7)\alpha^2]\left(\frac{d}{b}\right)^4, \end{aligned} \tag{A.101}$$

such that $\phi(3)$ decreases with d/b, and increases with $(d/b)^3$ and $(d/b)^4$, which implies that $\phi(3)$ is positive if and only if d/b is greater than some unique threshold $\underline{d/b} = [\arg \phi(3) = 0]$. Figure 5 shows that the derivative of this threshold with respect to α is negative for $n \in [7, 25]$. This proves the third bullet of the proposition. □