Barrett's paradox of cooperation in the case of quasi-linear utilities

E. Accinelli* Atefeh Afsar † Filipe Martins‡
Bruno M. P. M. Oliveira§ Jorge Oviedo¶ Alberto A. Pinto ‖
Luis Quintas ** José Martins ††

Abstract

This paper fits in the theory of international agreements by studying the success of stable coalitions of agents seeking the preservation of a public good. Extending Baliga and Maskin [3], we consider a model of homogeneous agents with quasi-linear utilities of the form $u_j(r_j; r) = r^\alpha - r_j$, where r is the aggregate contribution and the exponent α is the elasticity of the gross utility. We prove that membership of the stable coalitions grows from 1 up to attaining the grand coalition as the value of α increases in its natural range $(0, 1)$. We show that when the size of the stable coalition increases, the ratio of the welfare of the stable coalitions against the welfare of the competitive singleton coalition grows with α. However, we prove that the growth of the size of stable coalitions occurs with a much greater loss of the coalition members when compared to the utilities of the free-riders, which leads to the well-known Barrett's paradox of cooperation [4] even in the extreme case where the stable coalition misses only a single free-rider. The paradox of cooperation only breaks when α is large enough so that the grand coalition is stable.

*Facultad de Economía, Universidad Autónoma de San Luís Potosí, México. E-mail: elvio.accinelli@eco.uaslp.mx

†Department of Biological Sciences, The University of Texas at Dallas. E-mail: atefeh.afsar@utdallas.edu

‡Centro de Matemática, Universidade do Porto – CMUP, Portugal. E-mail: luis.f.martins@fc.up.pt

§Faculdade de Ciências da Nutrição e Alimentação, Universidade do Porto, Portugal and LIAAD–INESC TEC. E-mail: bmpmo@fcna.up.pt

¶Instituto de Matemáticas, Universidad Nacional de San Luis, Argentina. E-mail: joviedo@unsl.edu.ar

‖Departmento de Matemática, Universidade do Porto, Portugal and LIAAD–INESC TEC. E-mail: aapinto@fc.up.pt

**Instituto de Matemáticas, Universidad Nacional de San Luis, Argentina. E-mail: lquintas@unsl.edu.ar

††Politécnico de Leiria, Portugal and LIAAD–INESC TEC. E-mail: jmmartins@ipleiria.pt

- **Keywords**: public and common goods; free-riding; coalition formation; stability; paradox of cooperation.
- **JEL classification: C7, D7, H4.**
- **MSC2020 classification: 91-10, 91B18, 91B76, 91A12, 91A40.**

1 Introduction

Scott Barrett [4] argued that international environmental agreements are, typically, not successful, since when cooperation matters the most, stable coalitions may achieve only little. This claim has been known in the literature as *Barrett's paradox of cooperation* or simply the *paradox of cooperation*. The decision of whether or not to cooperate for the provision or maintenance of a non-excludable and non-rivalrous good can be based on countries' private interests, in some way a Hobbesian approach where the common interest is served only by a country when the neglect of the collective interest implies the loss of its individual well-being.

Baliga and Maskin [3] proposed a model for pollution-emitting communities that are negatively affected by these emissions, showing that any agreement involving two or more communities is vulnerable to free-riding. Hence, even when stable coalitions are formed, pollution reduction turns out to be no greater than in the case where negotiation is ruled out. The stable coalition coincides with the competitive equilibrium for the preservation of the public good, consisting of only a single agent paying the preservation costs and all the other agents acting as free-riders and not contributing.

Extending Baliga and Maskin [3], we consider a model of homogeneous agents with quasi-linear utilities of the form $u_j(r_j; r) = r^\alpha - r_j$, where r is the aggregate contribution and the exponent $\alpha \in (0, 1)$ coincides with the elasticity of the gross utility with respect to the aggregate contribution, where $\alpha = 1/2$ in [3]. The Nash–Cournot equilibria of the game where all agents competitively choose their contributions are the *low-cooperation* strategies: the aggregate effort of all the agents coincides with the stand-alone effort of a single agent optimizing his/her utility on his/her own. We observe that low-cooperation strategies are relevant in the dynamic framework previously analyzed in [1, 2] since they are stable equilibrium points for the myopic or adaptive dynamics. Hence, for this extended version of the Baliga and Maskin [3] model, it still holds that a single agent paying the preservation costs and all the other agents as free-riders and not contributing is a competitive equilibrium. However, as we will describe, stable coalitions have more than a single element when $\alpha > 1/2$.

In this work, we follow the approach in d'Aspremont *et* al. [7] in the context of price collusion. We consider a two-stage game, where a coalition formation stage is followed by a public goods game involving the choices of efforts or contributions made to the public good by the agents.

We first analyze the second stage public goods game after coalitions have been formed. Following some current literature ([6, 8, 9, 10, 12]), we consider the game where the formed coalition is one of the players, its utility being the aggregate utility of the coalition members, and the other players are the agents that do not belong to the coalition called *free-riders* relative to the coalition. As we will discuss, the free-riders might contribute or not to the public good. As considered in some of the previous works, we analyze two versions of the public goods game: Nash–Cournot and Stackelberg. In the Nash–Cournot game, all players choose their contributions simultaneously, while in the Stackelberg game, the coalition player is the leader and hence chooses its contribution first, followed by a Nash–Cournot game played by all the other agents. We prove in Lemmas 2 and 3 that the Nash–Cournot equilibria and the Stackelberg equilibria are of the following two types: i) *low-cooperation* strategies as described before (where free-riders may contribute to the public good); and ii) the *stand-alone* strategies of a coalition, *i.e.* the aggregate contribution maximizing the utility of the coalition, with all the free-riders contributing zero.

Secondly, we analyze the first stage coalition formation game. Following d'Aspremont *et* al. [7] we define the concepts of internal and external stability of coalitions, and hence of stable coalitions: a coalition $\mathcal{A}$ is stable if the members of the coalition do not have a utility advantage to become free-riders by leaving the coalition and *vice-versa*. We concentrate our analysis on *focal* stand-alone strategies that are both Nash–Cournot and Stackelberg equilibria for the second stage public goods games: a focal stand-alone coalition is a coalition with a stand-alone strategy of the coalition that is equally shared among coalition members because of the homogeneity of the agents. We prove in Theorem 1 that for each α and total number of players N, there is a stable (focal stand-alone) coalition $\mathcal{S}(\alpha; N)$, which is unique up to a permutation of agents. Furthermore, we prove that there is a *coalition cardinality* increasing step function $\ell(\alpha) \in \mathbb{N}$ such that: (i) $\ell(\alpha) = 1$, for $0 < \alpha \leq 1/2$, and (ii) $\ell(\alpha)$ tends to infinity, when α tends to 1, with the following property: the cardinality of the stable coalition $\mathcal{S}(\alpha; N)$ is the minimum between $\ell(\alpha)$ and the total number of agents N.

Inspired by the approach in Barrett [4], we will study the difficulties arising in the formation of stable coalitions and the effects of stable coalitions in preserving the public good. By Theorem 1, increasing α increases the size of stable coalitions, evolving from singleton coalitions with a single member, to the grand coalition $\mathcal{N}$ including all the agents. We observe that when the number of members of the stable coalitions increases, the utility of a free-rider over a member of the stable coalition utility also increases, making the formation of large coalitions much more difficult (except when the stable coalition becomes the grand coalition). In fact, the ratio of the utility of a free-rider against the utility of a member of the stable coalition equals $(1 - \alpha)^{-1}$, and so tends to infinity when α tends to 1.

As usual, the *welfare* $W(\alpha; \mathcal{A})$ associated with a stand-alone strategy of a coalition $\mathcal{A}$ consists of the aggregate utility of all N agents, including the free-riders. Since agents are homogeneous, the welfare will depend only upon the cardinality of the coalition and not on the agents that form the coalition nor on the way the costs are distributed among coalition members. Our goal is to compare the relative performance of the welfare associated with stable (focal stand-alone) coalitions $\mathcal{S}$ to *competitive* (focal stand-alone) singleton coalitions $\mathcal{C}$ formed by a single agent (corresponding to a competitive equilibrium), and the (focal stand-alone) grand coalition $\mathcal{N}$ formed by all agents.

The gap between full cooperation and the competitive scenario can be measured in terms of the relative welfare $W(\alpha;\mathcal{N})/W(\alpha;\mathcal{C})$ between the grand coalition and a competitive singleton coalition. We observe that this ratio tends to infinity, and so the gap increases when α tends to 1. This can be interpreted as that the need or urge to preserve the public good by the grand coalition increases with the elasticity α of the gross utility.

The relative welfare between stable coalitions $\mathcal{S}$ and a competitive singleton coalition $\mathcal{C}$ is the ratio $W(\alpha;\mathcal{S})/W(\alpha;\mathcal{C})$. Since stable coalitions are singleton coalitions for $0 < \alpha \leq 1/2$, the relative welfare equals 1. This fact was pointed out by Baliga and Maskin [3], for α in the boundary of this region: $\alpha = 1/2$. However, when α tends to 1: (i) the number of members of the stable coalitions grows up to N; and (ii) the relative welfare grows to $+\infty$. This can be interpreted as that the stable coalition achieves much higher welfare than the competitive singleton coalition when the elasticity of the gross utility α increases.

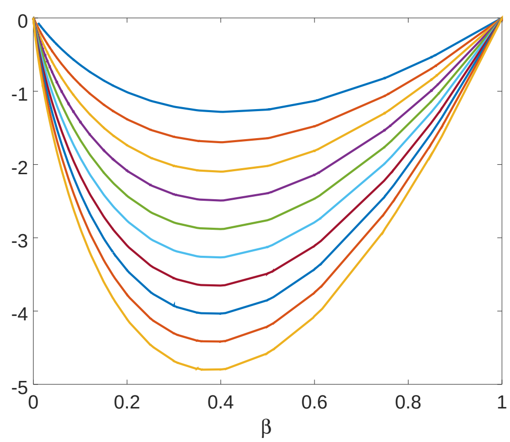

Figure 1: The function $\log\left(W(\alpha;\mathcal{S})/W(\alpha;\mathcal{N})\right)/N$ depending on $\beta \equiv \beta(\alpha;N) \equiv \ell(\alpha)/N$, for different values of $N = 10^3,\ldots,10^{12}$.

The relative welfare between stable coalitions $\mathcal{S}$ and the grand coalition $\mathcal{N}$ is the ratio $W(\alpha;\mathcal{S})/W(\alpha;\mathcal{N})$. By Theorem 3, for a large number of agents the relative welfare attains a low global minimum when the number of the members $\ell(\alpha)$ of the stable coalitions is close to one-third of the number of the members of the grand coalition N, more precisely, $\ell(\alpha) \approx e^{-1}N \approx 0.37N$ (see Figure 1). By Theorem 2 the welfare of stable coalitions differing from the grand coalition (*i.e.* in situations where there are free-riders) is very small compared to the welfare of the grand coalition. We observe that this observation enlarges the scope of Barrett's paradox of cooperation to the case where stable coalitions are large (for example when there is a single free-rider) and the gap between cooperation and no cooperation is also large. This case is seldom mentioned in the literature; usually, when the stable coalition is large then the gap between cooperation and no cooperation is small. For other types of utilities and emphasizing the role of diversity through asymmetric agents, Finus and McGinty [11] have shown related results.

Summarizing, when the membership of the stable coalitions grows up there is a much greater loss of the coalition members when compared with the free riders' utilities. This fact has two major drawbacks: firstly, it is the explanation and interpretation of Barrett's paradox of cooperation, since when a new free-rider enters the stable coalition, the jump of the welfare of the stable coalition has to become larger and larger. In particular, the welfare of the stable coalitions (when different from the grand coalition) has to be much smaller than the welfare of the grand coalition. Secondly, the formation of stable coalitions becomes more and more difficult since all agents much prefer to be free-riders rather than members of the stable coalition. These two facts are overcome only when the grand coalition becomes stable. Hence, to save the public good the stability of the grand coalition is of great importance and this only occurs for values of the elasticity of gross utility α large enough.

This paper is organized in the following way. In Section 2, we introduce the extended version of Baliga and Maskin's model of contributions for a public good. We study two variations of a public goods game regarding the contributions made after a coalition has been formed: Nash–Cournot, and Stackelberg, and we characterize equilibria in these public goods games. In Section 3, we study the first stage game regarding the formation of coalitions and characterize the stability of coalitions in terms of the elasticity α of the gross utility of the public good. In Section 4, we discuss the paradox of cooperation for the welfare of stable coalitions, in particular, the special case where the stable coalition is large but the gap between cooperation and the grand coalition is still large. In Section 5, we conclude and make some final remarks. In the appendices to this paper, we present the proofs of the results.

2 Public goods games

Baliga and Maskin [3] focus on the reduction of air pollution, but the model can be applied with suitable modifications to other situations involving collective social risk dilemmas on the preservation of public goods. Here, an extended version of Baliga and Maskin's model is presented by adding the parameter $0 < \alpha < 1$ ($\alpha = 1/2$ in [3]). In this section, we present this extended version of the model, and we study public good games regarding contributions to the public good after a certain coalition has been formed. We study two variations of the public goods game: Nash–Cournot, where the formed coalition and the remaining players choose their contributions simultaneously, and Stackelberg, where the formed coalition acts as the leader.

2.1 An extended Baliga–Maskin model

We consider that there are N homogeneous agents that can be countries, individuals, or, in game theory terms, players. They are indexed by $i \in \mathcal{N} = \{1, 2, \ldots, N\}$ who are users or consumers of a public good (which is the reduction of pollution in Baliga-Maskin's model). The contribution of agent j for the preservation of the public good is $r_j \geq 0$. The aggregate contribution of all agents except j is $r_{-j} = \sum_{i \in \mathcal{N} \setminus \{j\}} r_i$.

The *gross utility* or *benefit* function of an agent is

$$v(r_j + r_{-j}) = (r_j + r_{-j})^\alpha.$$

We observe that the gross benefit functions are strictly concave since $\alpha \in (0, 1)$ ($\alpha = 1/2$ in [3]). Hence,

$$\alpha = \lim_{\Delta r \to 0} \frac{\Delta v}{v(r)} / \frac{\Delta r}{r} = \frac{dv(r)}{dr} \frac{r}{v(r)},$$

and so the parameter α is the ratio of the percentage change in gross utility v to the percentage change in the aggregate effort r, that is, the elasticity of gross utility with respect to the aggregate contribution.

The *net utility* of agent j is the quasi-linear utility function

$$u_j(r_j; r_{-j}) = v(r_j + r_{-j}) - r_j = (r_j + r_{-j})^\alpha - r_j \ .$$

Therefore, the partial derivative of $u_j(r_j; r_{-j})$ with respect to r_j is

$$\frac{\partial u_j}{\partial r_j} = \alpha (r_j + r_{-j})^{\alpha - 1} - 1.$$

Note that $\partial u_j / \partial r_j = 0$ if and only if $r_j + r_{-j} = \underline{r} \equiv \alpha^{\frac{1}{1-\alpha}}$. Hence, we observe that $\underline{r}$ is the aggregate effort maximizing the net utility of agent j. We call $\underline{r}$ the *stand-alone effort* of a single agent since it is the optimal effort $r_j = \underline{r}$ of the agent when all the other agents do not contribute $r_{-j} = 0$.

2.2 Low-cooperation and stand-alone strategies

A *low-cooperation* strategy is a vector of individual efforts satisfying the aggregate effort $r_{\mathcal{N}}$ of all agents is equal to the stand-alone effort of a single agent

$$r_{\mathcal{N}} = \underline{r}.$$

A *coalition* $\mathcal{A} \subset \mathcal{N}$ is a subset of agents that are willing to participate in a cooperation agreement for the conservation or provision of a public good. The agents in $\mathcal{N} \setminus \mathcal{A}$ are called the *free-riders* relative to $\mathcal{A}$. Two examples of coalitions are the *singleton coalitions* $\mathcal{C} = \{i\}$, where $i \in \mathcal{N}$, constituted by a single agent, and the *grand coalition* $\mathcal{N}$ constituted by all agents.

The *aggregate effort* $r_{\mathcal{A}} = \sum_{j \in \mathcal{A}} r_j$ of the coalition $\mathcal{A}$ is the sum of the efforts of all the agents in coalition $\mathcal{A}$. The quantity $r_{-\mathcal{A}} = r_{\mathcal{N}} - r_{\mathcal{A}}$ is the aggregate contribution of all the free-riders relative to $\mathcal{A}$. In particular, $r_{\mathcal{N}}$ is the aggregate contribution of all agents and so

$$r_{\mathcal{N}} = r_{\mathcal{A}} + r_{-\mathcal{A}} = r_j + r_{-j}.$$

The *aggregate net utility* of coalition $\mathcal{A}$

$$u_{\mathcal{A}}(r_{\mathcal{A}}; r_{-\mathcal{A}}) = \sum_{j \in \mathcal{A}} u_j(r_j; r_{-j}) = \#\mathcal{A}(r_{\mathcal{A}} + r_{-\mathcal{A}})^{\alpha} - r_{\mathcal{A}}$$

is the sum of the net utilities of all the agents of the coalition $\mathcal{A}$. Therefore, the partial derivative of $u_{\mathcal{A}}(r_{\mathcal{A}}; r_{-\mathcal{A}})$ with respect to $r_{\mathcal{A}}$ is

$$\frac{\partial u_{\mathcal{A}}}{\partial r_{\mathcal{A}}} = \alpha\#\mathcal{A}(r_{\mathcal{A}} + r_{-\mathcal{A}})^{\alpha-1} - 1.$$

Note that $\partial u_{\mathcal{A}} / \partial r_{\mathcal{A}} = 0$ if and only if $r_{\mathcal{A}} + r_{-\mathcal{A}} = \overline{r}_{\mathcal{A}} \equiv (\alpha\#\mathcal{A})^{\frac{1}{1-\alpha}}$. Hence, we observe that $\overline{r}_{\mathcal{A}}$is the aggregate effort maximizing the aggregate net utility of coalition $\mathcal{A}$. We call $\overline{r}_{\mathcal{A}}$the *stand-alone effort* of coalition $\mathcal{A}$ since it is the optimal effort $r_{\mathcal{A}} = \overline{r}_{\mathcal{A}}$ of coalition $\mathcal{A}$ when all the other agents do not contribute $r_{-\mathcal{A}} = 0$. Hence, the stand-alone effort $\overline{r}_{\mathcal{A}}$ is the (sub)-optimal aggregate effort for a given coalition $\mathcal{A}$, when the free-riders do not contribute $r_{-\mathcal{A}} = 0$, being optimal when $\mathcal{A} = \mathcal{N}$.

We observe that for every coalition $\mathcal{A}$ we have the following

$$\underline{r} \leq \overline{r}_{\mathcal{A}} \ \leq \overline{r}_{\mathcal{N}}.$$

We have that $\underline{r} = \overline{r}_{\mathcal{A}}$ if and only if $\#\mathcal{A} = 1$, and that $\overline{r}_{\mathcal{A}} = \overline{r}_{\mathcal{N}}$ if and only if $\mathcal{A}$ is the grand coalition $\mathcal{N}$.

A *stand-alone* strategy of a coalition $\mathcal{A}$ is a vector of individual efforts satisfying the following two properties: i) the aggregate effort $r_{\mathcal{A}}$ of coalition $\mathcal{A}$ is equal to its stand-alone effort $\mathcal{A}$

$$r_{\mathcal{A}} = \overline{r}_{\mathcal{A}};$$

and ii) all the free-riders do not contribute

$$r_{-\mathcal{A}} = 0.$$

We observe that for stand-alone strategies, the strategies of free-riders are determined (equal to zero), while for coalition members they are not, only its aggregate effort being equal to the stand-alone effort of the coalition.

In particular, a stand-alone strategy is a low-cooperation strategy only when coalition $\mathcal{A}$ is a singleton.

2.3 Nash–Cournot and Stackelberg games

In this section we are going to consider simultaneous and sequential games where the players are the following: (i) a (formed) coalition $\mathcal{F}$; and (ii) all the free-riders $j \in \mathcal{N} \setminus \mathcal{F}$. We will consider two types of games: Nash–Cournot and Stackelberg. In the Nash–Cournot game, all players play simultaneously. In the Stackelberg game, the formed coalition player $\mathcal{F}$ is the leader and plays first, followed by all the free-riders $j \in \mathcal{N} \setminus \mathcal{F}$ playing a Nash–Cournot subgame by choosing simultaneously.

The following lemma provides the best response functions of the players.

Lemma 1. *The best response function of the coalition player $\mathcal{F}$ is*

$$r^*_{\mathcal{F}}(r_{-\mathcal{F}}) = \begin{cases} 0 & \text{if } r_{-\mathcal{F}} \geq \overline{r}_{\mathcal{F}} \\ \overline{r}_{\mathcal{F}} - r_{-\mathcal{F}} & \text{if } r_{-\mathcal{F}} < \overline{r}_{\mathcal{F}} \end{cases} .$$

The best response functions of the free-riders $j \in \mathcal{N} \setminus \mathcal{F}$ are

$$r^*_j(r_{-j}) = \begin{cases} 0 & \text{if } r_{-j} \geq \underline{r} \\ \underline{r} - r_{-j} & \text{if } r_{-j} < \underline{r} \end{cases} .$$

We now fully characterize the Nash–Cournot equilibria of the game.

Lemma 2. *The Nash–Cournot equilibria are the following:*

1. *if $\#\mathcal{F} = 1$, every low-cooperation strategy;*
2. *if $\#\mathcal{F} \geq 2$, every stand-alone strategy of coalition $\mathcal{F}$.*

We observe that when $\#\mathcal{F} = 1$ the Nash–Cournot game is a competitive game between all the agents. Case 1 of Lemma 2 characterizes the competitive equilibria of this game.

We now fully characterize the Stackelberg equilibria of the game. In the trivial case where $\mathcal{F} = \mathcal{N}$, the equilibria are stand-alone strategies of the grand coalition by the previous lemma. From now on, we will assume that $\mathcal{F} \neq \mathcal{N}$ and so there are free-riders playing the second stage of the Stackelberg public goods game. Hence, we assume that there is at least one agent not belonging to the formed coalition in cases 1 and 2 of the following lemma.

Lemma 3. *The Stackelberg equilibria are the following:*

1. *if* $\#\mathcal{F} = 1$*, every low-cooperation strategies such that* $r_{\mathcal{F}} = 0$*.*

2. *if* $\#\mathcal{F} = 2$ *and* $\alpha \leq 1/2$*, every low-cooperation strategies such that* $r_{\mathcal{F}} = 0$*;*

3. *if* $\#\mathcal{F} = 2$ *and* $\alpha > 1/2$*, every stand-alone strategy of coalition* $\mathcal{F}$*;*

4. *if* $\#\mathcal{F} \geq 3$*, every stand-alone strategy of coalition* $\mathcal{F}$*.*

In cases 1 and 2 of Lemma 3, the coalition $\mathcal{F}$ uses the advantage of moving first for its members not to contribute, putting the burden of the aggregate contribution on the free-riders. We observe that when $\#\mathcal{F} = 1$ the Stackelberg game is equivalent to a competitive game between all the agents, where one of the agents is the leader and plays first, followed by a simultaneous game between the other agents.

3 Stability of focal stand-alone coalitions

Since agents are homogeneous, it is natural to assume that agents belonging to a coalition equally share the costs among themselves, as was considered in [3]. We call a coalition with this cost-sharing structure a focal coalition.

In this section, we will define focal stand-alone coalitions that are both Nash–Cournot and Stackelberg equilibria of the second stage public goods games. Restricted to the set of focal stand-alone coalitions we study their stability. We also study the stability of coalitions with corresponding low-cooperation strategies that are Nash–Cournot and Stackelberg equilibria of the public goods games of the previous section.

3.1 Distribution of costs and utilities

A coalition $\mathcal{A}$ is *focal* if all the members of the coalition have the same effort or contribution. A *focal stand-alone coalition* is a focal coalition $\mathcal{A}$ together with a stand-alone strategy associated with it., *i.e.* the stand-alone effort is equally shared among coalition members and free-riders do not contribute. Hence, the contribution $r(j/\mathcal{A})$ of every coalition member $j \in \mathcal{A}$ is

$$r(j/\mathcal{A}) = \frac{\overline{r}_{\mathcal{A}}}{\#\mathcal{A}} = (\alpha(\#\mathcal{A})^{\alpha})^{\frac{1}{1-\alpha}} = \alpha\overline{r}_{\mathcal{A}}^{\alpha} ,$$

and the contribution $r(j/\mathcal{A})$ of every free-rider $j \in \mathcal{N} \setminus \mathcal{A}$ is $r(j/\mathcal{A}) = 0$.

Therefore, the utility of each agent $j \in \mathcal{N}$ is

$$u(j/\mathcal{A}) = \begin{cases} (1-\alpha)\overline{r}_{\mathcal{A}}^{\alpha}, & \text{if } j \in \mathcal{A} \\ \overline{r}_{\mathcal{A}}^{\alpha}, & \text{if } j \notin \mathcal{A} \end{cases} . \qquad (1)$$

These utilities determine the valuation functions of each agent for a given coalition $\mathcal{A}$. We observe that both the utility of a coalition member and that of a free-rider increase when the number of coalition members increases.

Given a focal stand-alone coalition $\mathcal{A}$, we observe that the relative utility between a free-rider agent $i \notin \mathcal{A}$ and a member of the coalition $j \in \mathcal{A}$ is

$$\frac{u(i/\mathcal{A})}{u(j/\mathcal{A})} = (1-\alpha)^{-1}.$$

Hence, the above ratio tends to $+\infty$ when α tends to 1.

Remark 1. *Therefore, a free-rider has a much higher utility than a member of a coalition when α is close to 1. Hence, when α is high, forming a coalition may become more complex, since free-riders have high utilities compared to coalition members, except in the case of the grand coalition where every agent is part of the coalition and there are no free-riders.*

3.2 Stable coalitions cardinality

After d'Aspremont *et al.* [7], a focal stand-alone coalition $\mathcal{A}$ is *internally stable* if all members $j \in \mathcal{A}$ of the coalition $\mathcal{A}$ prefer not to become free-riders, *i.e.*

$$u(j/\mathcal{A}) > u(j/(\mathcal{A} \setminus \{j\}));$$

and a focal stand-alone coalition coalition $\mathcal{A}$ is *externally stable* if all free-riders $j \notin \mathcal{A}$ prefer not to become members of $\mathcal{A}$, *i.e.*

$$u(j/\mathcal{A}) \geq u(j/(\mathcal{A} \cup \{j\})).$$

A focal stand-alone coalition $\mathcal{A}$ is *stable* if and only if it is both internally and externally stable.

Let $F : (0,1) \to \mathbb{R}^+$ be

$$F(\alpha) = 1 - (1-\alpha)^{\frac{1-\alpha}{\alpha}}.$$

Hence, F is a decreasing function whose range is $(0, (e-1)/e))$, $(e-1)/e \approx 0.63$. Let

$$G(\alpha) = \frac{1}{F(\alpha)}$$

which is an increasing function whose range is $(e/(e-1), +\infty)$, where $e/(e-1) \approx 1.58$.

Definition 1. *The* coalition cardinality $\ell(\alpha) \in \mathbb{N}$ *is the unique integer such that*

$$\ell(\alpha) \in [G(\alpha) - 1, G(\alpha)).$$

For all $i \in \mathbb{N}$, let $\alpha_1 = 0$ and let $\alpha_i \in (0,1)$ be such that $G(\alpha_i) = i$, for $i \geq 2$, and let $I_i = (\alpha_i, \alpha_{i+1}]$.

Hence, for every $i \in \mathbb{N}$ and $\alpha \in I_i$ we have

$$l(\alpha) = i.$$

From equation (10) in the appendix, we get

$$\lim_{\alpha \to 1} \frac{1/\ell(\alpha)}{-(1-\alpha)\log(1-\alpha)} = 1 \quad \text{and} \quad \lim_{\alpha \to 1} \frac{\log(\ell(\alpha))}{-\log(1-\alpha)} = 1. \tag{2}$$

We observe that: (i) $\ell(\alpha)$ is an increasing step function of α; (ii) $\ell(\alpha)$ is left-continuous, *i.e.* $\ell(\alpha_i) = G(\alpha_i) - 1$ and $\ell(\alpha_i^+) = G(\alpha_i) = \ell(\alpha_i) + 1$; (iii) $\ell(\alpha) = 1$ for $\alpha \leq \alpha_2 = 1/2$ (recall that $1/2$ is the value used by Baliga and Maskin); (iv) $\ell(\alpha)$ tends to $+\infty$, when α tends to 1 (see Figure 2).

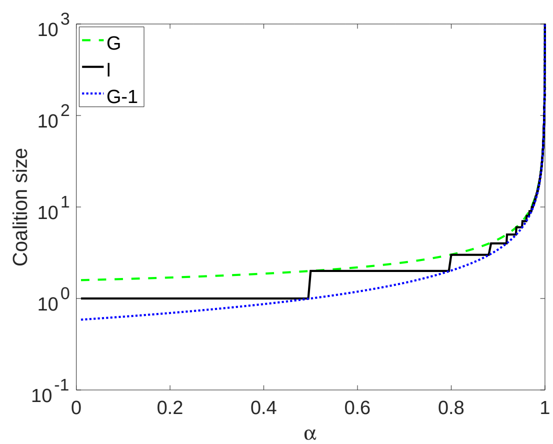

Figure 2: The coalition cardinality $\ell(\alpha)$ (black line) and the associated bounds from the definition: $G(\alpha) - 1$ (blue dots) and $G(\alpha)$ (green dashes). The vertical axis has logarithmic scale.

Let

$$\ell(\alpha; N) = \min\{\ell(\alpha), N\}.$$

Letting $J_N = (0, \alpha_N]$, we observe that: (i) $\ell(\alpha; N) = \ell(\alpha)$ if $\alpha \in J_N$; (ii) $\ell(\alpha; N) = N$ if $\alpha \in (0,1) \setminus J_N = (\alpha_N, 1)$.

Theorem 1 (Stability and cardinality)**.** *Let us consider that in the public goods game agents choose only to form Nash-Cournot equilibria that are focal stand-alone coalitions. A focal stand-alone coalition $\mathcal{A}$ is stable if and only if*

$$\#\mathcal{A} = \ell(\alpha; N).$$

The fact that only the cardinality of the coalition matters in terms of stability is due to the homogeneity of the agents.

In light of the previous theorem, in the remainder of this work, we will let $\mathcal{S} = \mathcal{S}(\alpha; N)$ denote a focal stand-alone stable coalition, *i.e.*, a coalition with

$$\#\mathcal{S}(\alpha; N) = \ell(\alpha; N).$$

3.3 A note on low-cooperation strategies

In this section, we study the external stability of coalitions when low-cooperation strategies are Nash–Cournot and Stackelberg equilibria of the public goods games in the previous section (case 1 of Lemmas 2 and 3).

In a low-cooperation strategy a free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ can have a positive contribution $r_i \geq 0$, since the only restriction of the strategy is that the aggregate contribution is $\underline{r}$. Let $r_M(\mathcal{F})$ be the largest contribution among free-riders

$$r_M(\mathcal{F}) = \max_{i \in \mathcal{N} \setminus \mathcal{F}} r_i.$$

By Lemma 2 a low-cooperation strategy is a Nash–Cournot equilibrium when the cardinality of the formed coalition $\mathcal{F}$ is 1. If a free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ joins $\mathcal{F} = \{j\}$, then the newly formed coalition $\mathcal{A}$ has 2 elements. For the strategy of the newly formed coalition $\mathcal{A}$ to be a Nash–Cournot equilibrium, by Lemma 2, it must be a stand-alone strategy of coalition $\mathcal{A}$. To be in the same context as in the previous section, we assume that $\mathcal{A}$ is a focal stand-alone coalition. We define the *relative contribution threshold* function

$$Z(\alpha) = \underline{r}^{\alpha-1} + 2^{\frac{\alpha}{1-\alpha}} \left(1 - \underline{r}^{\alpha-1}\right).$$

In Figure 3 We observe that Z is decreasing, $Z(1/2) = 0$ and that

$$\lim_{\alpha \to 0} Z(\alpha) = 1 - \log(2) \approx 0.3.$$

We have the following result.

Lemma 4. *Consider a Nash–Cournot equilibrium low-cooperation strategy for a formed coalition $\mathcal{F} = \{j\}$. A free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ prefers not to join coalition $\mathcal{F}$, forming a focal stand-alone coalition $\mathcal{A} = \{i, j\}$ if and only if $r_i/\underline{r} \leq Z(\alpha)$.*

Hence, the coalition $\mathcal{F} = \{j\}$ with a low-cooperation strategy is externally stable if and only if the largest contribution among free-riders $r_M(\mathcal{F})$ is such that $r_M(\mathcal{F})/\underline{r} \leq Z(\alpha)$. In particular, focal stand-alone coalitions $\mathcal{F}$ with cardinality 1 are externally stable when $\alpha \leq 1/2$ (and so $Z(\alpha) \geq 0$), in accordance with Theorem 1.

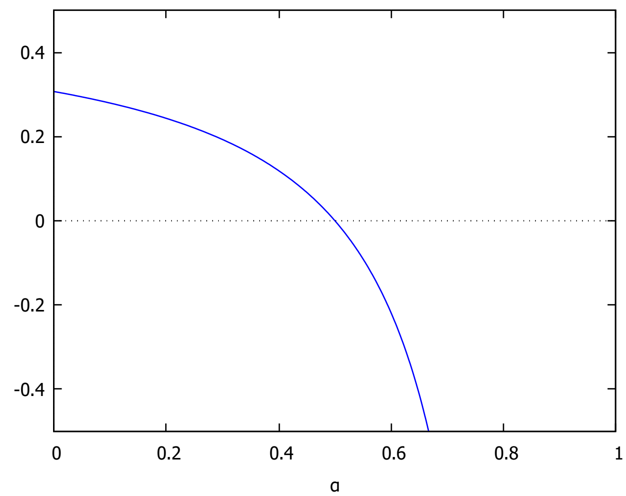

Figure 3: The function $Z(\alpha)$.

By Lemma 4, when $\alpha < 1/2$, if the relative contributions $r_i/\underline{r}$ of all the free-riders $i \neq j$ are small enough, the low-cooperation strategy is externally stable when compared to the focal stand-alone coalition formed by two agents $\mathcal{F} = \{i, j\}$. In particular, the relative contributions of free-riders $r_i/\underline{r}$ can increase up to approximately 0.3 when α tends to zero. However, $Z(\alpha) < 0$ for all $\alpha > 1/2$ gives the following conclusion.

Corollary 1. *A low-cooperation strategy of a formed coalition $\mathcal{F}$ is not externally stable for $\alpha > 1/2$.*

Hence, putting together Lemma 2 and Corollary 1, we conclude that when $\alpha \geq 1/2$ only stand-alone strategies of coalitions can be stable.

We now address the case when low-cooperation strategies are equilibria for the Stackelberg game. Recall from Lemma 3 that in case 1), $\#\mathcal{F} = 1$, or in case 2), $\#\mathcal{F} = 2$ and $\alpha \leq 1/2$ then a low-cooperation strategy with $r_{\mathcal{F}} = 0$ is a Stackelberg equilibrium. We have the following result.

Lemma 5. *Consider a Stackelberg equilibrium low-cooperation strategy for a formed coalition $\mathcal{F}$. A free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ always prefers to join coalition $\mathcal{F}$, and so $\mathcal{F}$ is not externally stable. Hence there are no stable Stackelberg low-cooperation equilibrium strategies.*

Recall that in the trivial case $\mathcal{F} = \mathcal{N}$, the Stackelberg equilibria is the stand-alone strategy of the grand coalition. By Lemma 3, we observe that for $\alpha \leq 1/2$, if $\#\mathcal{F} < N$, there are no focal stand-alone coalitions that are Stackelberg equilibria with cardinality 1 or 2. Hence, by Lemma 5 we obtain the following remark.

Remark 2. *Let us consider that in the public goods game agents choose only to form Stackelberg equilibria that are focal stand-alone coalitions or low-cooperation strategies. A focal stand-alone coalition $\mathcal{A}$ is stable if and only if*

1. *if* $\alpha \leq 1/2$, $\#\mathcal{A} = \min\{3, N\}$;
2. *if* $\alpha > 1/2$, $\#\mathcal{A} = \ell(\alpha; N)$.

4 The paradox of cooperation

We now compare the gains and losses of the welfare for the three most relevant coalitions: competitive singleton coalitions, stable coalitions, and the grand coalition, and we relate our results to the paradox of cooperation.

The *welfare* associated to a strategy $(r_1, \ldots, r_N)$ is

$$W(r_1, \ldots, r_N) = \sum_{j \in \mathcal{N}} u_j(r_j; r_{-j})$$

is the sum of the net utilities of all agents.

In the remainder of this paper, we will consider stand-alone strategies. Recall that this means that $r_{\mathcal{A}} = \overline{r}_{\mathcal{A}}$ and that $r_{-\mathcal{A}} = 0$.

The *welfare* W associated to a stand-alone strategy of a coalition $\mathcal{A}$ is

$$W(\alpha; \mathcal{A}) = \overline{r}_{\mathcal{A}} \left(\frac{N}{\alpha \# \mathcal{A}} - 1 \right) = \overline{r}_{\mathcal{A}}^{\alpha} \left(N - \alpha \# \mathcal{A} \right). \tag{3}$$

We observe that the welfare does not depend on the way the costs are distributed, and so the welfare is the same as for the focal stand-alone strategy of coalition $\mathcal{A}$.

The *relative welfare* $W(\alpha; \mathcal{A}/\mathcal{B})$ between coalition $\mathcal{A}$ and coalition $\mathcal{B}$ is

$$W(\alpha; \mathcal{A}/\mathcal{B}) = \frac{W(\alpha; \mathcal{A})}{W(\alpha; \mathcal{B})} = \left(\frac{\# \mathcal{A}}{\# \mathcal{B}} \right)^{\frac{\alpha}{1-\alpha}} \left(\frac{N - \alpha \# \mathcal{A}}{N - \alpha \# \mathcal{B}} \right). \tag{4}$$

4.1 Competitive versus full cooperation

The *gap* between the competitive and the full cooperation scenarios can be measured in terms of the relative welfare between the grand coalition and the competitive singleton coalition given by

$$W(\alpha; \mathcal{N}/\mathcal{C}) = N^{\frac{1}{1-\alpha}} \left(\frac{1 - \alpha}{N - \alpha} \right).$$

Hence, for $N > 1$, we get

$$\lim_{\alpha \to 0} W(\alpha; \mathcal{N}/\mathcal{C}) = 1 \quad \text{and} \quad \lim_{\alpha \to 1} W(\alpha; \mathcal{N}/\mathcal{C}) = +\infty.$$

Hence, the gap measured in terms of the relative welfare $W(\alpha;\mathcal{N}/\mathcal{C})$ between the grand coalition and the competitive singleton coalition increases with α. This can be interpreted as that the need or urge to preserve the public good by the grand coalition increases with the elasticity α of the gross utility. We recall that the paradox of cooperation says that when the above gap is higher it is when the relative welfare between the stable coalition and the grand coalition is smaller. We are going to analyze this in subsection 4.3.

4.2 Stable coalitions versus competitive scenario

The relative welfare between the stable coalition and the competitive singleton coalition is given by

$$W_{\mathcal{C}}(\alpha;N) \equiv W(\alpha;\mathcal{S}(\alpha;N)/\mathcal{C}) = \ell(\alpha;N)^{\frac{\alpha}{1-\alpha}}\left(\frac{N-\alpha\ell(\alpha;N)}{N-\alpha}\right).$$

For $N>1$, we observe that

$$\lim_{\alpha\to 0} W_{\mathcal{C}}(\alpha;N) = 1 \quad \text{and} \quad \lim_{\alpha\to 1} W_{\mathcal{C}}(\alpha;N) = +\infty.$$

Since the relative welfare $W_{\mathcal{C}}(\alpha;N)$ increases when α increases to 1, this means stable coalitions achieve a much higher welfare compared to the competitive scenario of a singleton coalition when the elasticity of the gross utility α increases. To analyse the status of Barrett's paradox of cooperation we have to compare stable coalitions with the full cooperation scenario, which we will do in the next section.

4.3 Stable coalitions versus full cooperation

The relative welfare $W_{\mathcal{N}}(\alpha;N)$ between the stable coalitions $\mathcal{S}(\alpha;N)$ and the grand coalition $\mathcal{N}$ is given by

$$W_{\mathcal{N}}(\alpha;N) \equiv W(\alpha;\mathcal{S}(\alpha;N)/\mathcal{N}) = \left(\frac{\ell(\alpha;N)}{N}\right)^{\frac{\alpha}{1-\alpha}}\left(\frac{N-\alpha\ell(\alpha;N)}{N(1-\alpha)}\right). \tag{5}$$

Recalling that $J_N=(0,\alpha_N]$, we observe that: (i) $W_{\mathcal{N}}(\alpha;N)<1$ if $\alpha\in J_N$; and (ii) $W_{\mathcal{N}}(\alpha;N)=1$ if $\alpha\in(0,1)\setminus J_N$, *i.e.*, when $\alpha>\alpha_N$. Hence, from now on, we restrict our analysis to the interval J_N where the stable coalitions $\mathcal{S}(\alpha;N)$ are not the grand coalition $\mathcal{N}$ and so, it is the set of interest to study the status of the paradox of cooperation.

The relative welfare $W_{\mathcal{N}}(\alpha;N)$ as a function of the parameter α has the following properties (see Figure 4): (i) for each interval $I_i=(\alpha_i,\alpha_{i+1}]$, $W_{\mathcal{N}}(\alpha;N)$ is smooth and strictly decreasing; and (ii) $W_{\mathcal{N}}(\alpha;N)$ is left-continuous.

Now, let us introduce the following *jump* functions:

$$W_L(\alpha) \equiv W_L(\alpha;N) = \left(\frac{G(\alpha)-1}{N}\right)^{\frac{\alpha}{1-\alpha}}\left(\frac{N-\alpha(G(\alpha)-1)}{N(1-\alpha)}\right); \tag{6}$$

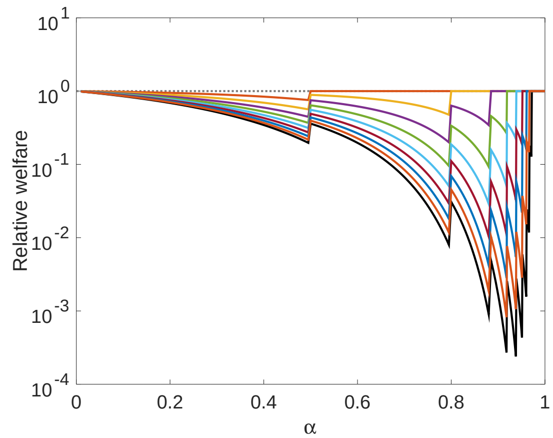

Figure 4: The relative welfare $W_{\mathcal{N}}(\alpha; N)$ between the stable and the grand coalition for values of $N = 2, \ldots, 10$. The vertical axis has logarithmic scale.

and

$$W_H(\alpha) \equiv W_H(\alpha; N) = \left(\frac{G(\alpha)}{N}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{N - \alpha G(\alpha)}{N(1-\alpha)}\right). \tag{7}$$

Numerically, we observe that the jump functions $W_L(\alpha)$ and $W_H(\alpha)$ are smooth functions having a unique minimum point in J_N. For $N > 1$, from (5) we observe that the bounds on $\ell(\alpha)$ imply that, for every $\alpha_i \leq \alpha_N$,

$$W_L(\alpha_i) = W_{\mathcal{N}}(\alpha_i; N) \quad \text{and} \quad W_H(\alpha_i) = W_{\mathcal{N}}(\alpha_i^+; N). \tag{8}$$

By (8) and (9) in the appendix, the relative jump of the relative welfare function $W_{\mathcal{N}}(\alpha; N)$ at the discontinuity $\alpha_i \leq \alpha_N$ is

$$\frac{W_{\mathcal{N}}(\alpha_i^+; N)}{W_{\mathcal{N}}(\alpha_i; N)} = \frac{W_H(\alpha_i)}{W_L(\alpha_i)} = \left(\frac{i}{i-1}\right)^{\frac{\alpha_i}{1-\alpha_i}} \left(\frac{N - i\alpha_i}{N - i\alpha_i + \alpha_i}\right) = (1-\alpha_i)^{-1} \left(\frac{N - i\alpha_i}{N - i\alpha_i + \alpha_i}\right) .$$

Hence, the jumps $W_{\mathcal{N}}(\alpha_i^+; N)/W_{\mathcal{N}}(\alpha_i; N)$ tend to $+\infty$ when α_i tends to 1 (*i.e.* when i and N tend to $+\infty$).

Remark 3. *In particular, noting that $W_{\mathcal{N}}(\alpha_N^+; N) = 1$, we have that*

$$W_{\mathcal{N}}(\alpha_N; N) = \frac{W_{\mathcal{N}}(\alpha_N; N)}{W_{\mathcal{N}}(\alpha_N^+; N)} = (1 - \alpha_N) + \frac{\alpha_N}{N} .$$

tends to 0 *when* N *tends to* $+\infty$. *This shows the strong impact of the entrance of the last free-rider when the grand coalition becomes stable.*

4.3.1 Global supremum of the relative welfare

Firstly, we study the relative welfare between the largest stable coalition different from the grand coalition and the grand coalition in terms of the parameter α. For every $N > 1$ and every $\alpha \in I_{N-1}$, the cardinality of the stable coalition is $\ell(\alpha) = N - 1$. Since the grand coalition $\mathcal{N}(\alpha)$ has only one element more than the stable coalition, we have

$$N(\alpha) = 1 + \ell(\alpha) = N.$$

Thus, the relative welfare achieved by the stable coalitions $\mathcal{S}(\alpha; N(\alpha))$ relative to the grand coalition $\mathcal{N}(\alpha)$ is

$$W_{\mathcal{N}(\alpha)}(\alpha; N(\alpha)) = \left(\frac{N(\alpha) - 1}{N(\alpha)}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{N(\alpha)(1-\alpha) + \alpha}{N(\alpha)(1-\alpha)}\right).$$

Theorem 2. *The following limits hold*

$$\lim_{\alpha\to 1} \frac{\log(W_{\mathcal{N}(\alpha)}(\alpha^+_{N(\alpha)-1}; N(\alpha)))}{\log(W_{\mathcal{N}(\alpha)}((1/2)^+; N(\alpha)))} = \lim_{\alpha\to 1} \frac{\log(W_{\mathcal{N}(\alpha)}(\alpha^+_{N(\alpha)-1}; N(\alpha)))}{\log(1-\alpha)} = 1 \ .$$

In particular, this implies that

$$\lim_{\alpha\to 1} W_{\mathcal{N}(\alpha)}(\alpha^+_{N(\alpha)-1}; N(\alpha)) = \lim_{N\to+\infty} W_{\mathcal{N}}(\alpha^+_{N-1}; N) = 0 \ .$$

This last assertion implies that the welfare that is achieved by the stable coalition with only one member less than the grand coalition becomes very low when compared to the welfare of the full-cooperation scenario of the grand coalition when the total number of agents is large.

Finally, let us study the *global supremum* of $W_{\mathcal{N}}(\alpha; N)|J_N$, for $N > 1$. We will study separately the *supremum* at the intervals I_1 and $J_N \setminus I_1$ since the *supremum* at I_1 is 1:

$$W_{\mathcal{N}}(\alpha; N)|I_1 = \left(\frac{1}{N}\right)^{\frac{\alpha}{1-\alpha}} \left(\frac{N-\alpha}{N(1-\alpha)}\right)$$

is a decreasing function of α with $W_{\mathcal{N}}(0^+; N) = 1$ and

$$W_{\mathcal{N}}(1/2; N) = \frac{2 - 1/N}{N} < \frac{2}{N}.$$

Now, the *supremum* of $W_{\mathcal{N}}(\alpha; N)|J_N \setminus I_1$ is attained: (i) at $\alpha = (1/2)^+$, where

$$W_{\mathcal{N}}((1/2)^+; N) = \left(\frac{4}{N}\right)\left(\frac{N-1}{N}\right);$$

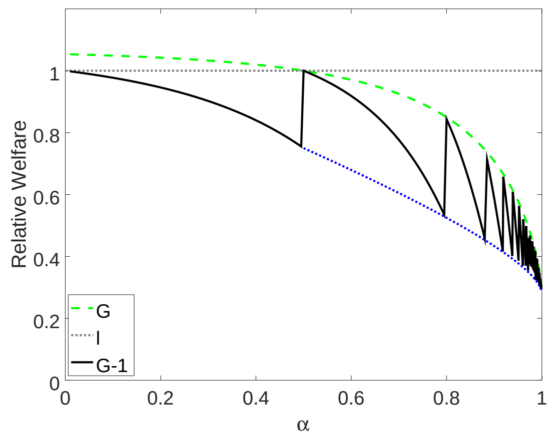

Figure 5: The function $W_{\mathcal{N}(\alpha)}(\alpha; N(\alpha))/W_{\mathcal{N}(\alpha)}((1/2)^+; N(\alpha))$.

or (ii) at the last discontinuity α^+_{N-1}, where

$$W_{\mathcal{N}}(\alpha^+_{N-1}; N) = \left(\frac{N-1}{N}\right)^{\frac{\alpha_{N-1}}{1-\alpha_{N-1}}} \left(\frac{N(1-\alpha_{N-1}) + \alpha_{N-1}}{N(1-\alpha_{N-1})}\right).$$

Thus, the *supremum* of $W_{\mathcal{N}}(\alpha; N)|J_N \setminus I_1$ is

$$\sup_{J_N \setminus I_1} W_{\mathcal{N}}(\alpha; N) = \max\left\{\left(\frac{4}{N}\right)\left(\frac{N-1}{N}\right), \left(\frac{N-1}{N}\right)^{\frac{\alpha_{N-1}}{1-\alpha_{N-1}}} \left(\frac{N(1-\alpha_{N-1}) + \alpha_{N-1}}{N(1-\alpha_{N-1})}\right)\right\}.$$

In Figure 5, for $\alpha \in J_N \setminus I_1$, we observe that

$$\frac{W_{\mathcal{N}(\alpha)}(\alpha^+_{N(\alpha)-1}; N(\alpha))}{W_{\mathcal{N}(\alpha)}((1/2)^+; N(\alpha))} \leq 1,$$

and so

$$\sup_{J_N \setminus I_1} W_{\mathcal{N}}(\alpha; N) = \left(\frac{4}{N}\right)\left(\frac{N-1}{N}\right) < \frac{4}{N}.$$

Remark 4. *For $N > 1$ by increasing $\alpha \in J_N$ the cardinality of the stable coalition $\mathcal{S}(\alpha; N)$ increases. However, increasing α does not increase the relative welfare between the stable coalitions $\mathcal{S}(\alpha; N)$ and the grand coalition $\mathcal{N}$, when compared to the relative welfare between the stable coalitions $\mathcal{S}(\alpha; N)$ and the grand coalition $\mathcal{N}$ at $\alpha = (1/2)^+$. This is a result along the same lines as the paradox of cooperation. Nonetheless, when $\alpha > \alpha_N$, the grand coalition becomes stable, the full cooperation scenario is attained, and the paradox no longer holds.*

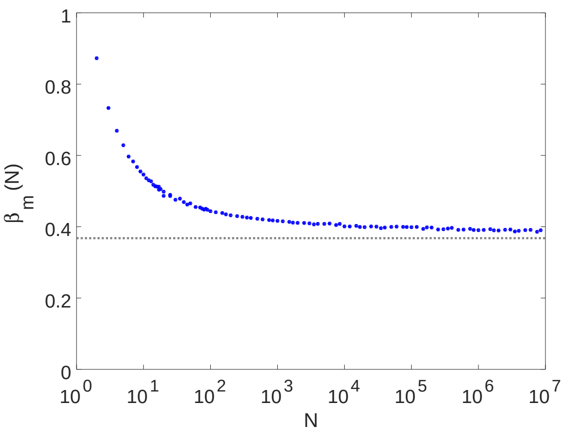

Figure 6: The sequence $\beta_m(N) \equiv \ell(\alpha_m(N))/N$ tends to $e^{-1} \approx 0.37$ (horizontal grey dots) when N tends to $+\infty$.

4.3.2 Global minimum of the relative welfare

Now we will characterize the values of α corresponding to the global minimum of the relative welfare $W_{\mathcal{N}}(\alpha; N)$ when N is large (see figure 1).

Let $\alpha_m(N)$ be such that $W'_L(\alpha_m(N); N) = 0$. Let $i(N) \in \mathbb{N}$ be such that $\alpha_{i(N)-1} < \alpha_m(N) \leq \alpha_{i(N)}$. Hence, the global minimum of $W_{\mathcal{N}}(\alpha; N)$ is attained at $\alpha_{i(N)-1}$ or at $\alpha_{i(N)}$. We recall that $\ell(\alpha_{i(N)}) = \ell(\alpha_m(N))$ and $\ell(\alpha_{i(N)-1}) = \ell(\alpha_m(N)) - 1$ (recall figure 2 and the definition of $l(\alpha)$). The following theorem characterizes these when N is large (see also figure 6).

Theorem 3. *The following limits hold*

$$\lim_{N\to+\infty} \frac{\ell(\alpha_{i(N)-1})}{N} = \lim_{N\to+\infty} \frac{\ell(\alpha_{i(N)})}{N} = \lim_{N\to+\infty} \frac{\ell(\alpha_m(N))}{N} = e^{-1}.$$

For N large enough, from (5) the minimum of the relative welfare is

$$W_{\mathcal{N}}(\alpha_m(N); N) \approx e^{-\frac{1}{1-\alpha_m(N)}} \left(\frac{e - \alpha_m(N)}{1 - \alpha_m(N)} \right).$$

From Theorem 3 we obtain that the minimum $\alpha_m(N)$ goes to 1 when N increases. This implies that the minimum relative welfare becomes very small when N is large enough, which can be observed numerically in Figure 1.

5 Conclusions

The antagonism between individual and collective rationality takes on a more dramatic and socially influential nuance when it comes to public or common goods whose nonexcludability usually results in problems of under-provision or exhaustion, which has been an issue in economics for a very long time. The main results from this work are the following: we show that there are stable coalitions different from the competitive singleton coalition consisting of only one agent, and so with more than one agent willing to bear the costs of preserving the public good, even with the issue of the existence of free-riders. Furthermore, the number of agents belonging to the stable coalition grows when there is an increase in the elasticity of the gross utility α and so a greater need for the preservation of the public good. Hence, contrarily to the paradox of cooperation, when the urge to preserve the good is high enough (α high enough), the grand coalition becomes stable. However, we have reinterpreted the paradox of cooperation for a case not much considered in the literature, where stable coalitions are large (but not the grand coalition) and the gap between cooperation and no cooperation is still large. The preservation of the public good by these large stable coalitions is far better than by the competitive coalitions, but also far worse than the full cooperative scenario corresponding to the grand coalition. Hence, the existence of even a few free-riders may undermine the preservation of the public good by the stable coalitions.

As future developments, one can study the repercussions of heterogeneity among agents in the formation of stable coalitions. Another future development would be to identify mechanisms that make the grand coalition stable through a system of rewards, transfers, and/or cost allocations. Another possibility is to study uncertainty and thresholds regarding climate catastrophes (see Barrett [5]).

A Proof of Lemmas 1, 2 and 3

Note that $\partial u_{\mathcal{A}}(r_{\mathcal{A}}; r_{-\mathcal{A}})/\partial r_{\mathcal{A}} = 0$ if and only if $r_{\mathcal{A}} + r_{-\mathcal{A}} = \overline{r}_{\mathcal{A}} = (\alpha \# \mathcal{A})^{\frac{1}{1-\alpha}}$. moreover, $\partial u_{\mathcal{A}}(r_{\mathcal{A}}; r_{-\mathcal{A}})/\partial r_{\mathcal{A}} < 0$ if and only if $r_{\mathcal{A}} + r_{-\mathcal{A}} > \overline{r}_{\mathcal{A}}$. Hence, the best-response function of player $\mathcal{A}$ against the effort $r_{-\mathcal{A}}$ of other players is

$$r^*_{\mathcal{A}}(r_{-\mathcal{A}}) = \begin{cases} 0, & \text{if } r_{-\mathcal{A}} \geq \overline{r}_{\mathcal{A}} \\ \overline{r}_{\mathcal{A}} - r_{-\mathcal{A}}, & \text{if } r_{-\mathcal{A}} < \overline{r}_{\mathcal{A}} \end{cases} .$$

Note that $\partial u_j(r_j; r_{-j})/\partial r_j = 0$ if and only if $r_j + r_{-j} = \underline{r} = \alpha^{\frac{1}{1-\alpha}}$. Moreover $\partial u_j(r_j; r_{-j})/\partial r_j < 0$ if and only if $r_j + r_{-j} > \underline{r}$. Hence, the best-response function of players $j \in \mathcal{N} \setminus \mathcal{A}$ against the effort r_{-j} of other players is

$$r^*_j(r_{-j}) = \begin{cases} 0, & \text{if } \ r_{-j} \geq \underline{r} \\ \underline{r} - r_{-j}, & \text{if } \ r_{-j} < \underline{r} \end{cases} .$$

Hence, Lemma 1 holds. □

Let us now analyse the Nash–Cournot equilibria.
Case $\#\mathcal{A} = 1$: noting that $\overline{r}_{\mathcal{A}} = \underline{r}$, the best-response of all players is optimal if and only if

$$r_{\mathcal{N}} = r_{\mathcal{A}} + \sum_{j \in \mathcal{N} \setminus \mathcal{A}} r_j = \underline{r}.$$

Hence, the Nash–Cournot equilibria are all the low-cooperation coalitions of $\mathcal{N}$.
Case $\#\mathcal{A} > 1$*:* noting that $\overline{r}_{\mathcal{A}} > \underline{r}$, the best-response of player $\mathcal{A}$ is optimal if and only if

$$r_{\mathcal{A}} + \sum_{j \in \mathcal{N} \setminus \mathcal{A}} r_j = \overline{r}_{\mathcal{A}}.$$

Since $\overline{r}_{\mathcal{A}} > \underline{r}$, the best-response of player $j \in \mathcal{N} \setminus \mathcal{A}$ is optimal if and only if $r_j = 0$ and so $r_{-\mathcal{A}} = 0$. Hence, $r_{\mathcal{A}} = \overline{r}_{\mathcal{A}}$ and the equilibria are stand-alone strategies of coalition $\mathcal{A}$. Therefore, Lemma 2 holds. □

Let us now analyse the Stackelberg equilibria.
Leader chooses $r_{\mathcal{A}} \geq \underline{r}$: in this case the best response r_j of the followers j is 0 and so $r_{-\mathcal{A}} = 0$. Hence, the best choice for the leader is $r_{\mathcal{A}} = \overline{r}_{\mathcal{A}}$. Furthermore,

$$u_{\mathcal{A}}(\overline{r}_{\mathcal{A}}; 0) = \alpha^{\frac{\alpha}{1-\alpha}} \#\mathcal{A}^{\frac{1}{1-\alpha}} (1-\alpha).$$

Leader chooses $r_{\mathcal{A}} < \underline{r}$: in this case, the best response r_j of the followers j is such that

$$r_{\mathcal{A}} + \sum_{j \in \mathcal{N} \setminus \mathcal{A}} r_j = \underline{r}.$$

Hence, the best choice for the leader is $r_{\mathcal{A}} = 0$. Furthermore,

$$u_{\mathcal{A}}(0; \underline{r}) = \#\mathcal{A} \alpha^{\frac{\alpha}{1-\alpha}}.$$

Let us compare the utility of the leader in both cases:

$$\frac{u_{\mathcal{A}}(\overline{r}_{\mathcal{A}}; 0)}{u_{\mathcal{A}}(0; \underline{r})} = \#\mathcal{A}^{\frac{\alpha}{1-\alpha}} (1-\alpha) \geq 1$$

which is equivalent to

$$\#\mathcal{A} \geq (1-\alpha)^{\frac{\alpha-1}{\alpha}}.$$

We observe that:

- $2 \leq (1-\alpha)^{\frac{\alpha-1}{\alpha}} < e < 3$, if $\alpha \leq 1/2$; and
- $1 < (1-\alpha)^{\frac{\alpha-1}{\alpha}} < 2$, if $\alpha > 1/2$.

Hence, case 1 of Lemma 3 only holds if: (i) $\#\mathcal{A} = 1$ or (ii) $\#\mathcal{A} = 2$ and $\alpha \leq 1/2$; and case 2 of Lemma 3 only holds if: (i) $\#\mathcal{A} > 2$ or (ii) $\#\mathcal{A} = 2$ and $\alpha > 1/2$. Therefore, Lemma 3 follows. □

B Proof of Theorem 1

Consider a coalition $\mathcal{S}$. Using (1), the coalition $\mathcal{S}$ is internally stable if for $j \in \mathcal{S}$, we have

$$(1-\alpha)\overline{r}_{\mathcal{S}}^{\alpha} > \overline{r}_{\mathcal{S}\setminus\{j\}}^{\alpha}.$$

Analogously, the coalition $\mathcal{S}$ is externally stable if for $j \notin \mathcal{S}$ we have

$$\overline{r}_{\mathcal{S}}^{\alpha} \geq (1-\alpha)\overline{r}_{\mathcal{S}\cup\{j\}}^{\alpha}.$$

Manipulating these inequalities using the stand-alone contributions, we obtain, respectively,

$$\#\mathcal{S} < \left(1-(1-\alpha)^{\frac{1-\alpha}{\alpha}}\right)^{-1} = G(\alpha),$$

and

$$\#\mathcal{S} \geq \left(1-(1-\alpha)^{\frac{1-\alpha}{\alpha}}\right)^{-1} - 1 = G(\alpha)-1.$$

So, the coalition $\mathcal{S}$ is stable if and only if

$$G(\alpha)-1 \leq \#\mathcal{S} < G(\alpha).$$

We have defined $\ell(\alpha)$ as the unique integer in this interval, so we have that $\mathcal{S}$ is stable if and only if $\#\mathcal{S} = \ell(\alpha)$, when $\ell(\alpha) < N$, since N is the maximum coalition cardinality, and $\#\mathcal{S} = N$ when $\ell(\alpha) \geq N$. □

C Proof of Lemmas 4 and 5

In a low-cooperation strategy of a formed coalition $\mathcal{F} = \{j\}$, a free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ contributing r_i has utility $\underline{r}^{\alpha} - r_i$. If $\#\mathcal{A} = 2$ we have that $r_{\mathcal{A}} = 2^{\frac{1}{1-\alpha}}\underline{r}$. If the free-rider i joins coalition $\mathcal{F}$, then for the newly formed focal stand-alone coalition $\mathcal{A}$ we have that $u(i/\mathcal{A}) = r_{\mathcal{A}}^{\alpha} - r_{\mathcal{A}}/2 = 2^{\frac{\alpha}{1-\alpha}}\underline{r}^{\alpha} - 2^{\frac{\alpha}{1-\alpha}}\underline{r}$. Hence, the free-rider prefers not to join coalition $\mathcal{F}$ if and only if

$$r_i \leq Z_{NC}(\alpha) = \underline{r}^{\alpha} + 2^{\frac{\alpha}{1-\alpha}}\left(\underline{r} - \underline{r}^{\alpha}\right),$$

which proves Lemma 4. □

Consider a low-cooperation strategy of a formed coalition $\mathcal{F} = \{j\}$ with $r_j = 0$ and a free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ contributing r_i, hence having utility $\underline{r}^{\alpha} - r_i$. When i joins coalition $\mathcal{F}$, then according to Lemma 3 there are two possibilities for the strategy of the newly formed coalition $\mathcal{A} = \{i, j\}$ to remain a Stackelberg equilibrium, given by cases 2 and 3. In case 2), occurring if $\alpha < 1/2$, i has utility $\underline{r}^{\alpha}$ after joining $\mathcal{F}$, which is always greater than his/her utility when a free-rider, and so i always joins, and $\mathcal{F}$ is not externally stable. In case 3), occurring if $\alpha \geq 1/2$, then the strategy of the newly formed coalition must be a stand-alone strategy, in which case we assume it to be focal. Then the same computation as in the previous lemma applies, giving $r_i \geq 0 > Z_{NC}(\alpha)$, and hence i always joins and $\mathcal{F}$ is not externally stable.

Now consider a low-cooperation strategy of a formed coalition $\mathcal{F} = \{j, l\}$, with $r_j = r_l = 0$, and a free-rider $i \in \mathcal{N} \setminus \mathcal{F}$ contributing r_i. When i joins coalition $\mathcal{F}$, then according to Lemma 3 there only possibility for the strategy of the newly formed coalition $\mathcal{A} = \{i, j\}$ to remain a Stackelberg equilibrium, is given by case 4, *i.e.* a stand-alone strategy, in which case we assume it to be focal. The utility of i as a free-rider is $\underline{r}^\alpha - r_i$. If $\#\mathcal{A} = 3$ we have that $r_{\mathcal{A}} = 3^{\frac{1}{1-\alpha}}\underline{r}$, and we have that $u(i/\mathcal{A}) = 3^{\frac{\alpha}{1-\alpha}}\underline{r}^\alpha - 3^{\frac{\alpha}{1-\alpha}}\underline{r}$. Hence, the free-rider prefers not to join coalition $\mathcal{F}$ if and only if

$$r_i \leq Z_S(\alpha) = \underline{r}^\alpha + 3^{\frac{\alpha}{1-\alpha}}\left(\underline{r} - \underline{r}^\alpha\right),$$

which never holds since $Z_S(\alpha) < 0$ for every α, and so $\mathcal{F}$ is not externally stable, which finishes the proof of Lemma 5. □

D Formulas for F

We start by showing some useful properties of the function $F(\alpha) = 1/G(\alpha)$.

For every $i > 1$, we have

$$(1 - \alpha_i)^{\frac{1-\alpha_i}{\alpha_i}} = 1 - F(\alpha_i) = 1 - 1/i = \frac{i-1}{i}. \tag{9}$$

The function $F(\alpha)$ satisfies the following limits

$$\lim_{\alpha \to 1} \frac{F(\alpha)}{-(1-\alpha)\log(1-\alpha)} = 1 \quad \text{and} \quad \lim_{\alpha \to 1} \frac{\log(\mathrm{F}(\alpha))}{\log(1-\alpha)} = 1. \tag{10}$$

Therefore,

$$\lim_{\alpha \to 1} \frac{\log(1 + F(\alpha))}{-(1-\alpha)\log(1-\alpha)} = 1. \tag{11}$$

The first derivative $F'(\alpha)$ of $F(\alpha)$ satisfies the following limit

$$\lim_{\alpha \to 1} \frac{F'(\alpha)}{\log(1-\alpha)} = 1. \tag{12}$$

Therefore

$$\lim_{\alpha \to 1} \frac{(1-\alpha)F'(\alpha)}{F(\alpha)} = -1. \tag{13}$$

D.1 Proofs of the above limits

Let

$$T(\alpha) = 1 - F(\alpha) = (1-\alpha)^{\frac{1-\alpha}{\alpha}}. \tag{14}$$

We observe that $\lim_{\alpha\to 1} F(\alpha) = \lim_{\alpha\to 1}(1 - T(\alpha)) = 0$. Hence, $F'(\alpha) = -T'(\alpha)$ and

$$\log T(\alpha) = \frac{1-\alpha}{\alpha} \log{(1-\alpha)}.$$

Let

$$V(\alpha) = \frac{F'(\alpha)}{1 - F(\alpha)} \tag{15}$$

Therefore,

$$V(\alpha) = -(\log T(\alpha))' = \frac{1}{\alpha^2} \log{(1-\alpha)} + \frac{1}{\alpha}.$$

Hence,

$$\lim_{\alpha\to 1} \frac{V(\alpha)}{F'(\alpha)} = \lim_{\alpha\to 1} \frac{1}{T(\alpha)} = 1, \tag{16}$$

because $\lim_{\alpha\to 1} F(\alpha) = 0$. Therefore, equation (12) holds. In particular,

$$\lim_{\alpha\to 1} F'(\alpha) = \lim_{\alpha\to 1} V(\alpha) = -\infty \quad \text{and} \quad \lim_{\alpha\to 1}(1-\alpha)F'(\alpha) = \lim_{\alpha\to 1}(1-\alpha)V(\alpha) = 0^- \ . \tag{17}$$

By L'Hopital's rule and using equation (12),

$$\lim_{\alpha\to 1} \frac{F(\alpha)}{-(1-\alpha)\log(1-\alpha)} = \lim_{\alpha\to 1} \frac{F'(\alpha)}{1+\log(1-\alpha)} = \lim_{\alpha\to 1} \frac{F'(\alpha)}{\log(1-\alpha)} = 1,$$

proving the left-side of (10).

Furthermore,

$$V'(\alpha) = \frac{F''(\alpha)(1-F(\alpha)) + (F'(\alpha))^2}{1-F(\alpha)} = -\frac{2}{\alpha^3} \log{(1-\alpha)} - \frac{1}{(1-\alpha)\alpha^2} - \frac{1}{\alpha^2}. \tag{18}$$

Thus,

$$\lim_{\alpha\to 1}(1-\alpha)V'(\alpha) = -1 \quad \text{and so} \quad \lim_{\alpha\to 1} \frac{(1-\alpha)V'(\alpha)}{F'(\alpha)} = 0^+. \tag{19}$$

Using L'Hôpital's rule and (16) and (19),

$$\lim_{\alpha\to 1} \frac{(1-\alpha)V(\alpha)}{F(\alpha)} = \lim_{\alpha\to 1} \frac{-V(\alpha) + (1-\alpha)V'(\alpha)}{F'(\alpha)} = -1. \tag{20}$$

By (15),

$$\frac{(1-\alpha)F'(\alpha)}{F(\alpha)} = \frac{(1-\alpha)V(\alpha)(1-F(\alpha))}{F(\alpha)} = \frac{(1-\alpha)V(\alpha)}{F(\alpha)} - \frac{(1-\alpha)V(\alpha)F(\alpha)}{F(\alpha)}$$

Hence, by (20),

$$\lim_{\alpha\to 1}\frac{(1-\alpha)F'(\alpha)}{F(\alpha)} = \lim_{\alpha\to 1}\frac{(1-\alpha)V(\alpha)}{F(\alpha)} - \lim_{\alpha\to 1}\frac{(1-\alpha)V(\alpha)F(\alpha)}{F(\alpha)} = -1$$

and so (13) holds. Using L'Hopital's rule,

$$\lim_{\alpha\to 1}\frac{\log(F(\alpha))}{\log(1-\alpha)} = \lim_{\alpha\to 1} -\frac{F'(\alpha)(1-\alpha)}{F(\alpha)} = 1$$

and so the right-side of (10) holds. Using L'Hopital's rule,

$$\lim_{\alpha\to 1}\frac{\log(1+F(\alpha))}{-(1-\alpha)\log(1-\alpha)} = \lim_{\alpha\to 1}\frac{F'(\alpha)}{(1+F(\alpha))(1+\log(1-\alpha))} = \lim_{\alpha\to 1}\frac{F'(\alpha)}{\log(1-\alpha)} = 1$$

and so equation (11) holds.

E Proof of Theorem 2

Noting that $\ell(\alpha) = N(\alpha) - 1$ and by the definition of $\alpha_{N(\alpha)-1}$, we get

$$\ell(\alpha) = N(\alpha) - 1 = G(\alpha_{N(\alpha)-1}) = (F(\alpha_{N(\alpha)-1}))^{-1}.$$

Hence,

$$N(\alpha) = \frac{1+F(\alpha_{N(\alpha)-1})}{F(\alpha_{N(\alpha)-1})} \quad \text{and} \quad \frac{N(\alpha)-1}{N(\alpha)} = \frac{1}{1+F(\alpha_{N(\alpha)-1})}.$$

Letting

$$v(\alpha) = (1+F(\alpha))^{-\frac{\alpha}{1-\alpha}}\left(\frac{(1+F(\alpha))(1-\alpha)+\alpha F(\alpha)}{(1+F(\alpha))(1-\alpha)}\right),$$

we get

$$\begin{aligned} v(\alpha_{N(\alpha)-1}) &= \left(\frac{N(\alpha)-1}{N(\alpha)}\right)^{\frac{\alpha_{N(\alpha)-1}}{1-\alpha_{N(\alpha)-1}}}\left(\frac{N(\alpha)(1-\alpha_{N(\alpha)-1})+\alpha_{N(\alpha)-1}}{N(\alpha)(1-\alpha_{N(\alpha)-1})}\right) \\ &= W_{\mathcal{N}(\alpha)}(\alpha^{+}_{N(\alpha)-1}; N(\alpha)). \end{aligned}$$

Let $z(\alpha) = \log(v(\alpha)) = z_1(\alpha) + \log(1+z_2(\alpha))$, where

$$z_1(\alpha) = -\frac{\alpha}{1-\alpha}\log(1+F(\alpha)) \quad \text{and} \quad z_2(\alpha) = \frac{\alpha F(\alpha)}{(1+F(\alpha))(1-\alpha)}.$$

By equation (11), we get

$$\lim_{\alpha\to 1}\frac{z_1(\alpha)}{\log(1-\alpha)} = 1.$$

By equation (12) and using L'Hopital's rule, we obtain

$$\begin{aligned}\lim_{\alpha\to 1}\frac{-z_2(\alpha)}{\log(1-\alpha)} &= \lim_{\alpha\to 1}\frac{(F(\alpha)+\alpha F'(\alpha)(1+F(\alpha))-\alpha F(\alpha)(1+F'(\alpha))}{(1+F(\alpha))^2(1+\log(1-\alpha))}\\ &= \lim_{\alpha\to 1}\frac{\alpha F'(\alpha)}{\log(1-\alpha)} = 1\end{aligned}$$

Hence,

$$\begin{aligned}\lim_{\alpha\to 1} z(\alpha) &= \lim_{\alpha\to 1}\left(z_1(\alpha)+\log(-\log(1-\alpha))+\log\left(\frac{1+z_2(\alpha)}{-\log(1-\alpha)}\right)\right) =\\ &= \lim_{\alpha\to 1}\left(z_1(\alpha)+\log(-\log(1-\alpha))\right).\end{aligned}$$

Therefore, we have that

$$\lim_{\alpha\to 1}\frac{z(\alpha)}{\log(1-\alpha)} = \lim_{\alpha\to 1}\frac{z_1(\alpha)+\log(-\log(1-\alpha))}{\log(1-\alpha)} = \lim_{\alpha\to 1}\frac{z_1(\alpha)}{\log(1-\alpha)} = 1.$$

which proves the second equality in the first assertion in the theorem. In particular, we obtain that

$$\lim_{\alpha\to 1} W_{\mathcal{N}(\alpha)}(\alpha^+_{N(\alpha)-1}; N(\alpha)) = \lim_{\alpha\to 1} v(\alpha_{N(\alpha)-1}) = 0.$$

Finally, by equation (10), we observe that

$$\lim_{\alpha\to 1}\frac{\log(W_{\mathcal{N}(\alpha)}((1/2)^+; N(\alpha)))}{\log(1-\alpha)} = 1$$

which gives the first equality and ends the proof. □

F Proof of Theorem 3

For $(\alpha, x)\in(0,1)^2$, define the function $\beta(\alpha)$ by

$$\beta(\alpha)\equiv\beta(\alpha;x) = \frac{x}{F(\alpha)}. \tag{21}$$

The motivation to define $\beta(\alpha)$ is that

$$\beta(\alpha_i; 1/N) = \frac{\ell(\alpha_i^+)}{N}.$$

Therefore,

$$\beta'(\alpha) = -\frac{\beta(\alpha)F'(\alpha)}{F(\alpha)} \quad \text{and} \quad (\log \beta(\alpha))' = \frac{\beta'(\alpha)}{\beta(\alpha)} = -\frac{\mathrm{F}'(\alpha)}{\mathrm{F}(\alpha)} \tag{22}$$

Hence, by (13),

$$\lim_{\alpha\to 1} \frac{(1-\alpha)\beta'(\alpha)}{\beta(\alpha)} = \lim_{\alpha\to 1} -\frac{(1-\alpha)F'(\alpha)}{F(\alpha)} = 1. \tag{23}$$

Given a function $V_T(\alpha; x)$ with a unique interior minimum point $\alpha_T(x)$, we have that $V'(\alpha_T(x); x) = 0$. Assume that $\alpha_T(x)$ is invertible and that $x_T(\alpha)$ is its inverse, *i.e.*

$$\alpha_T(x_T(\alpha)) = \alpha \quad \text{and} \quad V_T'(\alpha; x_T(\alpha)) = 0.$$

Let

$$\beta_T(\alpha) = \beta(\alpha; x_T(\alpha)) = x_T(\alpha)/F(\alpha).$$

Let

$$V_{WH}(\alpha) = \beta(\alpha)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\alpha\beta(\alpha)}{1-\alpha}\right).$$

We observe that for every $N \in \mathbb{N}$, $V_{WH}(\alpha; 1/N)$ coincides with $W_H(\alpha; N)$ (see equations (6) and (7)). Let $\alpha_{WH}(x)$ be such that $V'(\alpha_{WH}(x); x) = 0$, and let $x_{WH}(\alpha)$ be its inverse, *i.e.*

$$\alpha_T(x_{WH}(\alpha)) = \alpha \quad \text{and} \quad V_{WH}'(\alpha; x_{WH}(\alpha)) = 0.$$

Let

$$\beta_{WH}(\alpha) = \beta(\alpha; x_{WH}(\alpha)) = x_{WH}(\alpha)/F(\alpha).$$

Lemma 6. *We have* $\beta_{WH}(1^-) = e^{-1}$.

Proof. We have

$$\log V_{WH}(\alpha) = \frac{\alpha}{1-\alpha} \log \beta(\alpha) + \log(1-\alpha\beta(\alpha)) - \log(1-\alpha).$$

Therefore,

$$\begin{aligned}\frac{V_{WH}'(\alpha)}{V_{WH}(\alpha)} &= \frac{1}{(1-\alpha)^2}\log\beta(\alpha) + \frac{\alpha}{1-\alpha}\frac{\beta'(\alpha)}{\beta(\alpha)} - \frac{\beta(\alpha)+\alpha\beta'(\alpha)}{1-\alpha\beta(\alpha)} + \frac{1}{1-\alpha} \\ &= \frac{1}{(1-\alpha)^2}\log\beta(\alpha) + \frac{(1-2\alpha\beta(\alpha)+\beta(\alpha))\alpha}{1-\alpha\beta(\alpha)}\frac{\beta'(\alpha)}{(1-\alpha)\beta(\alpha)} - \frac{\beta(\alpha)}{1-\alpha\beta(\alpha)} + \frac{1}{1-\alpha}\end{aligned}$$

Hence, by equation (23),

$$0 = \lim_{\alpha\to 1} \frac{(1-\alpha)^2 V_{WH}'(\alpha)}{V_{WH}(\alpha)} = 1 + \log \beta_{WH}(1^-). \tag{24}$$

□

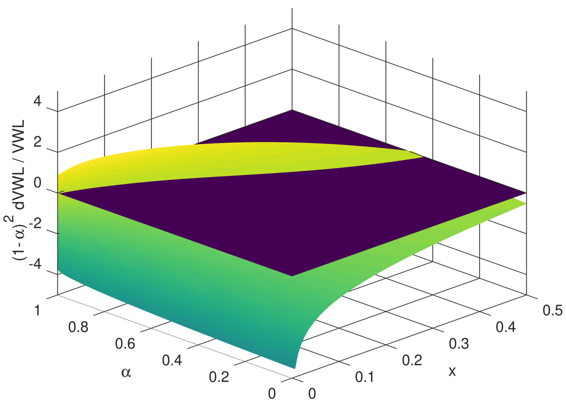

Figure 7: The function $\alpha_{WL}(x)$ and its inverse $x_{WL}(\alpha)$ are the intersection of the surface with the horizontal plane.

Let

$$V_{WL}(\alpha) = (1-\alpha)\left(1 + \frac{\alpha\beta F(\alpha)}{1-\alpha\beta}\right) V_{WH}(\alpha).$$

Noting that $(1-F)^{\alpha/(1-\alpha)} = 1-\alpha$, for every $N \in \mathbb{N}$, $V_{WL}(\alpha; 1/N)$ coincides with $W_L(\alpha; N)$ (see equations (6) and (7)). Let $\alpha_{WL}(x)$ be such that $V'(\alpha_{WL}(x); x) = 0$, and let $x_{WL}(\alpha)$ be its inverse (see Figure 7). Let (see Figure 8)

$$\beta_{WL}(\alpha) = \beta(\alpha; x_{WL}(\alpha)) = x_{WL}(\alpha)/F(\alpha).$$

Lemma 7. *We have* $\beta_{WL}(1^-) = e^{-1}$.

Proof. We have that

$$\log V_{WL}(\alpha) = \log(1-\alpha) + \log\left(1 + \frac{\alpha\beta F(\alpha)}{1-\alpha\beta}\right) + \log V_{WH}(\alpha).$$

Hence, after some algebraic manipulations we get

$$\frac{V'_{WL}(\alpha)}{V_{WL}(\alpha)} = \frac{V'_{WH}(\alpha)}{V_{WH}(\alpha)} - \frac{1}{1-\alpha} + \frac{F(\alpha)\beta}{1-\alpha\beta(1-F(\alpha))}\frac{1}{1-\alpha\beta} - \frac{\alpha\beta F'(\alpha)}{1-\alpha\beta(1-F(\alpha))}.$$

Hence, by (24),

$$\lim_{\alpha\to 1}\frac{(1-\alpha)^2 V'_{WL}(\alpha)}{V_{WL}(\alpha)} = \lim_{\alpha\to 1}\frac{(1-\alpha)^2 V'_{WH}(\alpha)}{V_{WH}(\alpha)} = 0 \qquad (25)$$

Therefore, V_{WL} and V_{WH} have the same critical values $\beta_{WL}(1^-) = \beta_{WH}(1^-) = e^{-1}$. □

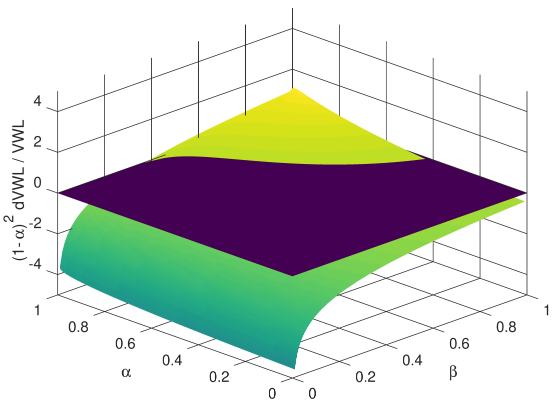

Figure 8: The function $\beta_{WL}(\alpha)$ is the intersection of the surface with the horizontal plane.

Lemma 8. *The following limits hold*

$$\lim_{N\to+\infty} \alpha_{WL}(1/N) = 1 \quad \textit{and so} \quad \lim_{N\to+\infty} \beta_{WL}(\alpha_{WL}(1/N)) = e^{-1}.$$

Proof. By Lemma 7 and continuity of $\beta_{WL}(\alpha)$, for any $0 < \beta_0 < e^{-1}$, there is $0 < \alpha_0 < 1$, such that $\beta_{WL}(\alpha) \geq \beta_0$, for all $\alpha \geq \alpha_0$. In particular, $\beta_{WL}(\alpha_{WL}(x)) \geq \beta_0$, for every $x \leq \beta(\alpha_0)F(\alpha_0)$. Hence,

$$\lim_{x\to 0} G(\alpha_{WL}(x)) = \lim_{x\to 0} \frac{\beta_{WL}(\alpha_{WL}(x))}{x} \geq \lim_{x\to 0} \frac{\beta_0}{x} = +\infty.$$

Therefore, $\lim_{x\to 0} \alpha_{WL}(x) = 1$. In particular,

$$\lim_{N\to+\infty} \alpha_{WL}(1/N) = 1 \quad \text{and so} \quad \lim_{N\to+\infty} \beta_{WL}(\alpha_{WL}(1/N)) = e^{-1}.$$

□

We observe that for every $N \in \mathbb{N}$, $V_{WL}(\alpha; 1/N)$ coincides with $W_L(\alpha; N)$. Hence,

$$W_L'(\alpha_{WL}(1/N); N) = V_{WL}'(\alpha_{WL}(1/N); 1/N) = 0.$$

Since

$$\beta_{WL}(\alpha_{WL}(1/N)) - \frac{2}{N} \leq \frac{\ell(\alpha_{WL}(1/N) - 1)}{N} < \frac{\ell(\alpha_{WL}(1/N))}{N} < \beta_{WL}(\alpha_{WL}(1/N)),$$

from Lemma 8, we obtain

$$\lim_{N\to+\infty} \frac{\ell(\alpha_{WL}(1/N)) - 1}{N} = \lim_{N\to+\infty} \frac{\ell(\alpha_{WL}(1/N))}{N} = e^{-1},$$

which ends the proof of Theorem 3. □

Acknowledgements

This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020, and within project "Modelling, Dynamics and Games" with reference PTDC/MAT-APL/31753/2017.

Elvio Accinelli wishes to thank CUMEX (Consortium of Mexican Universities) and the Ibero-American Postgraduate University Association (AUIP) for the support granted for his stay in Portugal between July 10 and August 24, 2021 and also the Department of Mathematics of University of Porto for their hospitality.

Atefeh Afsar would like to thank the financial support of FCT through a PhD. grant of the MAP-PDMA program with reference PD/BD/142886/2018.

Filipe Martins was partially supported by CMUP, member of LASI, which is financed by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the project with reference UIDB/00144/2020.

References

[1] Elvio Accinelli, Filipe Martins, and Alberto A. Pinto. Evolutionary dynamics for the generalized Baliga–Maskin public good model. *Chaos, Solitons & Fractals*, page 109496, 2019. URL: `http://www.sciencedirect.com/science/article/pii/S0960077919304424`, `doi:https://doi.org/10.1016/j.chaos.2019.109496`.

[2] Elvio Accinelli, Filipe Martins, and Alberto A. Pinto. The basins of attraction in the generalized Baliga–Maskin public good model. *Journal of Evolutionary Economics*, 2021. URL: `https://link.springer.com/article/10.1007/s00191-021-00758-z`, `doi:10.1007/s00191-021-00758-z`.

[3] Sandeep Baliga and Eric Maskin. Mechanism design for the environment. In K.G. Mäler and J. Vincent, editors, *Handbook of environmental economics*, pages 306–324. Elsevier Science/North Holland, 2003. URL: `https://www.sciencedirect.com/science/article/abs/pii/S157400990301012X`, `doi:10.1016/S1574-0099(03)01012-X`.

[4] Scott Barrett. Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46:878–894, 1994. URL: `http://www.jstor.org/stable/2663505`.

[5] Scott Barrett. Climate treaties and approaching catastrophes. *Journal of Environmental Economics and Management*, 66(2):235–250, 2013. URL: `https://www.sciencedirect.com/science/article/abs/pii/S0095069612001222`, `doi:10.1016/j.jeem.2012.12.004`.

[6] Basak Bayramoglu, Michael Finus, and Jean-François Jacques. Climate agreements in a mitigation-adaptation game. *Journal of Public Economics*, 165:101–113, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0047272718301300, doi:https://doi.org/10.1016/j.jpubeco.2018.07.005.

[7] Claude d'Aspremont, Alexis Jacquemin, Jean Gabszewicz, and John Weymark. On the stability of collusive price leadership. *Canadian Journal of Economics*, 16(1):17–25, 1983. URL: https://EconPapers.repec.org/RePEc:cje:issued:v:16:y:1983:i:1:p:17-25.

[8] Effrosyni Diamantoudi and Eftichios S Sartzetakis. Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory*, 8(2):247–263, 2006. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9779.2006.00262.x, doi:10.1111/j.1467-9779.2006.00262.x.

[9] Klaus Eisenack and Leonhard Kähler. Adaptation to climate change can support unilateral emission reductions. *Oxford Economic Papers*, 68(1):258–278, 2016. URL: https://academic.oup.com/oep/article-abstract/68/1/258/2362411, doi:10.1093/oep/gpv057.

[10] Michael Finus, Francesco Furini, and Anna Viktoria Rohrer. The efficacy of international environmental agreements when adaptation matters: Nash-Cournot vs Stackelberg leadership. *Journal of Environmental Economics and Management*, 109:102461, 2021. URL: https://www.sciencedirect.com/science/article/pii/S0095069621000449, doi:https://doi.org/10.1016/j.jeem.2021.102461.

[11] Michael Finus and Matthew McGinty. The anti-paradox of cooperation: Diversity may pay! *Journal of Economic Behavior & Organization*, 157:541 – 559, 2019. URL: http://www.sciencedirect.com/science/article/pii/S0167268118302890, doi:https://doi.org/10.1016/j.jebo.2018.10.015.

[12] Santiago J Rubio and Alistair Ulph. Self-enforcing international environmental agreements revisited. *Oxford economic papers*, 58(2):233–263, 2006. URL: https://academic.oup.com/oep/article-abstract/58/2/233/2361959, doi:10.1093/oep/gpl002.