THE UNIVERSITY OF CHICAGO

Core Formation in Asymmetric Deterring Groups

By

Jack Melin

August 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts Degree
in the Master of Arts Program of the Social Sciences

Faculty Advisor: Roger Myerson
Preceptor: Ryan Yuhao Fang

TABLE OF CONTENTS

ABSTRACT

This paper studies a cooperative model of multilateral deterrence with a focus on how parametric heterogeneities in the deterring group affect core formation. Alongside basic analysis, two notably contrasting results are derived. First, it is shown that no parametrically symmetric deterring group of size $n \geq 4$ can support a core allocation. Second, a sufficient condition for core existence is derived for "single-differing-agent" groups (groups in which all agents except one outlier are parametrically identical up to a positive scale factor) which does not depend upon the specific nature of the individual non-outliers, and thus holds for groups of arbitrarily large size. Finally, computational analysis of this condition is used to show that it holds in many cases in which the outlier is sufficiently dominant in either retaliatory capability or incentive to deter.

1 INTRODUCTION

The basic binary-action deterrence game has been well-understood since the inception of game theory; it is discussed as an example in the essential *Theory of Games and Economic Behavior* (Morganstern & Von Neumann 1944). This formulation is a sequential one in which an aggressor chooses whether or not to act, and then a deterrer chooses whether or not to retaliate. The effect of including potential costly ex-ante commitment to retaliation by the deterrer is also well-understood by Morganstern & Von Neumann; for perfect- and complete-information constructions of this commitment step, this game is very easily understood and its sequential equilibria arise clearly.

Extensions of this bilateral model involving the complication of information structures or action sets abound, particularly in the game-theoretic international relations literature on war and deterrence—"Rationalist Explanations for War" (Fearon 1995) and "The Reasons for War: An Updated Survey" (Jackson & Morelli 2009) are useful references which catalog many of these complications. However, empirical inspirations for deterrence theory often involve multiple deterrers acting in concert, and the study of these more complex systems is less fleshed out. Most of this study is highly empirical, studying burden-sharing within existing alliances such as NATO—"Economics of Alliances: The Lessons for Collective Action" (Sandler & Hartley 2001) is a useful survey. The empirical bent of this literature allows it to focus on descriptive and highly context-specific international relations questions—for instance, how different types of weapons systems are shared differently—but makes it ill-equipped to answer more comparative questions about the relationship between a given coalition's attributes and its ability to effectively cooperate. Small sample size, impossibly complicated historical context, and other econometric difficulties exacerbate this issue.

In this paper, we attempt to address this absence from the modeling end: by building and studying the behavior of a cooperative model resembling a familiar contribution game. Members of a potential coalition of deterrers each choose levels of commitment to retaliate against aggression by an exogenous aggressor; aggression is less likely to occur with each member's increasing commitment, but it is never impossible, and retaliation is individually costly. The positive externality generated by increasing individual actions (agent A's increased commitment decreases the probability of aggression without costing agent B anything more if aggression happens) creates a contribution game dynamic in which cooperation leads to better overall outcomes. However, the existence of a core allocation which can support said cooperation is unclear and therefore of pressing concern.

This paper abstracts out the deterree's action in order to focus on the conditions under which the deterrers can successfully cooperate with one another. Of particular focus is the effect on core existence of the structure of the deterring group: how does group behavior change as the group's deterrent capability is subdivided among more and more members? How does this change if members are asymmetric in capability and/or incentives? The crux research question of this paper can be summarized as follows: how does the (potentially heterogeneous) makeup of a deterring group affect its ability to cooperate?

The results of this type of comparative analysis are undoubtedly limited in their direct predictive power, given their intentional separation from empirical data. However, they can inform a more sophisticated qualitative understanding of the outcomes of these types of games by allowing for direct study of the relationship between parametric cause and effect. The two centerpiece results of this paper demonstrate this: we show that increased coalition size begets instability, while the existence of a single dominant member in a coalition begets stability. While these results take more explicit quantitative forms in the context of the model studied, the nuances of these forms are necessarily dependent upon the model's exact specification. Thus, this paper aspires to approach the problem of understanding multilateral deterrent behavior from a different direction, but it does so in the knowledge that this direction is also limited in its capacity. Ideally, this paper's model and its conclusions will be integrated with empirics in future work.

2 MODEL SPECIFICATION & BASIC ANALYSIS

2.1 Game Structure

Consider a cooperative game with finite agent set N, where N represents a set of agents working to deter aggression by some exogenous aggressor. Each agent $i \in N$ chooses an action $x_i \in \mathbb{R}^+$ which represents the quantity of retaliatory force they commit to use if aggression occurs—the probability of aggression occurring decreases as the group's total retaliatory commitment increases, but increasing individual commitment to retaliate is costly in the event that aggression occurs.

Each agent $i \in N$ is parametrically characterized by two real values, c_i and r_i.

- c_i is the fixed cost incurred by i if aggression occurs.
- r_i determines how costly it is for i to use retaliatory force. If aggression occurs and i is committed to retaliate with x_i amount of force, then i incurs a cost $r_i(e^{x_i/r_i} - 1)$[1].

For notational simplicity, we write $x \doteq \sum_{i \in N} x_i$, as well as $c \doteq \sum_{i \in N} c_i$ and $r \doteq \sum_{i \in N} r_i$. Under an action profile $\vec{x} \doteq \{x_i | i \in N\}$, aggression occurs with probability $e^{-\sum_{i \in N} x}$; this can be rationalized as an exogenous aggressor comparing the force of retaliation x against a random private non-negative value of aggression V with $F_V(v) = 1 - e^{-v}$. Agents only derive payoffs from the expected costs of aggression. Then, under action profile $\vec{x}$, each $i \in N$ experiences a payoff of

$$u_i(\vec{x}) = (\text{Prob. of Aggression}) \times (\text{Cost of Aggression}) = -e^{-x}(c_i + r_i(e^{x_i/r_i} - 1))$$

In order to get good behavior from this specification, a few constraints on the parametric values $\vec{c} = \{c_i | i \in N\}$ and $\vec{r} = \{r_i | i \in N\}$ are necessary:

---

[1] This functional form, along with the constraints described later in Section 2.1, is chosen to satisfy useful shape conditions such as convexity in x_i while retaining algebraic workability.

- $\forall i \in N : c_i > r_i > 0$. This constraint ensures that agents are in fact averse to aggression and will not pursue negative retaliation values.
- $1 > r$, which implies $\forall i \in N : 1 > r_i$. If this constraint is violated, then the problem of deterrence is trivialized because the coalition can achieve expected payoffs arbitrarily close to the first-best of 0 by committing to sufficiently large retaliatory force. In a sense, the coalition becomes "too strong" and the problem is no longer interesting.

2.2 Non-Cooperative Nash Equilibrium Benchmark

Although this paper will deal with the aforementioned game in a primarily cooperative framework, it is easy and instructive to address it non-cooperatively as well. This is helpful as a sort of benchmark for behavior in the absence of core-stable cooperation.

Consider an agent i and an action profile $\vec{x}_{-i}$ over all $j \in N \backslash i$, with sum x_{-i}. i's payoff from choosing x_i is

$$u_i(x_i, \vec{x}_{-i}) = -e^{-x_i - x_{-i}}(c_i + r_i(e^{x_i/r_i} - 1))$$

Differentiating with respect to x_i,

$$\frac{\partial u_i(x_i, \vec{x}_{-i})}{\partial x_i} = e^{-x_i - x_{-i}}(c_i - r_i + (r_i - 1)e^{x_i/r_i})$$

Differentiating again with respect to x_i,

$$\frac{\partial^2 u_i(x_i, \vec{x}_{-i})}{\partial x_i^2} = -e^{-x_i - x_{-i}}(c_i - r_i + \frac{(r_i - 1)^2}{r_i} e^{x_i/r_i})$$

In conjunction with the relevant parametric constraints, this implies that u_i is convex. Then, we apply the usual first-order condition for convex optimization:

$$c_i - r_i + (r_i - 1)e^{x_i^*/r_i} = 0 \therefore e^{x_i^*/r_i} = \frac{c_i - r_i}{1 - r_i} \therefore x_i^* = r_i \ln(\frac{c_i - r_i}{1 - r_i})$$

Recalling the constraints $c_i > r_i$ and $1 > r$, it is easy to see that this value is indeed positive and therefore a valid choice of x_i. Thus, this value of x_i induces the unique global maximum payoff, and is therefore the unique best-response x_i.

Theorem 2.2.1. *In a non-cooperative version of the specified deterrence game, there exists a unique Nash equilibrium. In this Nash equilibrium, each agent $i \in N$ chooses*

$$x_i = r_i \ln(\frac{c_i - r_i}{1 - r_i})$$

This conclusion raises a few points which bear mentioning.

- The Nash equilibrium action of each agent does not depend upon the actions or characteristics of the other agents. This is despite the fact that each agent's *payoff* is indeed affected by others' actions.

- The Nash equilibrium action of each agent is strictly increasing in its c_i value. This makes intuitive sense, because increasing c_i implies greater incentive to avoid deterrence failure.

- Because $c_i > 1$, the Nash equilibrium action of each agent is also strictly increasing in its r_i value:
$$\frac{\partial}{\partial r_i} r_i \ln(\frac{c_i - r_i}{1 - r_i}) = \ln(\frac{c_i - r_i}{1 - r_i}) + \frac{r_i(c_i - 1)}{(1 - r_i)(c_i - r_i)} > 0$$
This also makes intuitive sense, because greater r_i implies less costly deterrence.

This means that in Nash equilibrium, each agent $i \in N$ gets payoff

$$-e^{-\sum_{j \in N} r_j \ln(\frac{c_j - r_j}{1 - r_j})}(c_i + r_i(e^{-r_i \ln(\frac{c_i - r_i}{1 - r_i})/r_i} - 1)) = -\prod_{j \in N}(\frac{c_j - r_j}{1 - r_j})^{r_j} \frac{2 - r_i}{1 - r_i}(c_i - r_i)$$

so the sum payoff across all agents in Nash equilibrium—the non-cooperative analog to the grand coalition's value—is

$$-\prod_{j\in N}(\frac{c_j - r_j}{1-r_j})^{r_j}\sum_{i\in N}\frac{2-r_i}{1-r_i}(c_i - r_i)$$

2.3 Characteristic Function

In order to address the specified deterrence game in a cooperative framework, a characteristic function $\mathcal{I}: 2^N \to \mathbb{R}$ is necessary. For the total coalition N, the characteristic function can be straightforwardly solved as a payoff-sum maximization problem with the actions of all $i \in N$ as control parameters. For any defecting proper sub-coalition $K \subset N$, a problem arises—in order to pin down $\mathcal{I}(K)$ by summing the payoffs of each $i \in K$, it is necessary to know the action profile pursued by the remaining coalition $N\backslash K$. This action profile is outside of the cooperative control of K, but nonetheless meaningful. Numerous solution concepts exist to deal with this issue, from straightforward maximin style concepts to more sophisticated higher-order belief systems; see "Cores of Games with Positive Externalities" (Chander 2010) and "The Core of Aggregative Cooperative Games with Externalities" (Stamatopoulos 2014) for examples.

For our purposes, a two-player Nash equilibrium approach is ideal in terms of matching tractability with realism[2]. We assume that K and the complementary sub-coalition $N\backslash K$ coordinate cooperatively within themselves but do not cooperate with one another. The expected outcome of K's defection is thereby well-defined if there is a unique Nash equilibrium of the following synthetic two-player game:

- The players are K and $N\backslash K$.
- K's action set is the set of possible action profiles $\vec{x}_K \doteq \{x_i | i \in K\}$ from the original

---

[2] 2. An ambitious robustness checker might be inclined to attempt other solution concepts such as those discussed in Chander and Stamatopoulos to see whether the results of this paper hold up. This is definitely out of scope here, however.

game. $N\backslash K$'s action set is defined likewise.

- K's payoff under action profile $(\vec{x}_K, \vec{x}_{N\backslash K})$ is

  $$u_K(\vec{x}_K, \vec{x}_{N\backslash K}) = \sum_{i \in K} u_i(\vec{x}_K, \vec{x}_{N\backslash K})$$

  i.e. the sum of the payoffs of the members of K in the original game under the action profile which corresponds to $(\vec{x}_K, \vec{x}_{N\backslash K})$. $N\backslash K$'s action set is defined likewise.

It is an important feature of the specified deterrence game that this synthetic game will always have a unique Nash equilibrium, and thus the characteristic function is well-defined on 2^N. This feature is demonstrated by construction in the following calculation of the closed-form value of $\mathcal{I}$.

The sum payoff of agents in a sub-coalition $K \subseteq N$ under action profile $(\vec{x}_K, \vec{x}_{N\backslash K})$ is

$$u_K(\vec{x}_K, \vec{x}_{N\backslash K}) = \sum_{i \in K} u_i((\vec{x}_K, \vec{x}_{N\backslash K})) = -e^{-x_K - x_{N\backslash K}}(c_K - r_K + \sum_{i \in K} r_i e^{x_i/r_i})$$

$$= e^{x_{N\backslash K}}(-e^{-x_K}(c_K - r_K + \sum_{i \in K} r_i e^{x_i/r_i}))$$

where it is emphasized that the effect on $u_K(\vec{x})$ of the actions taken by agents in $N\backslash K$ can be reduced to a positive leading factor $e^{-x_{N\backslash K}}$. This implies that K maximizes its value with the same action regardless of $N\backslash K$'s action—then, the problem of identifying Nash equilibria reduces to a straightforward optimization problem. Moreover, the aforementioned leading factor can be entirely ignored during this optimization process, although it will need to be reintroduced after K's value-optimizing action is identified.

K solves the maximization problem

$$\vec{x}_K^* \in \underset{\vec{x}_K}{\operatorname{argmax}} -e^{-x_K}(c_K - r_K + \sum_{i \in K} r_i e^{x_i/r_i})$$

Mirroring the non-cooperative case, we differentiate with respect to each $x_i \in \vec{x}_K$:

$$\forall i \in K : \frac{\partial}{\partial x_i} - e^{-x_K}(c_K - r_K + \sum_{j \in K} r_j e^{x_j/r_j}) = e^{-x_K}(c_K - r_K + \sum_{j \in K} r_j e^{x_j/r_j} - e^{x_i/r_i})$$

and once again by the same x_i in each case:

$$\forall i \in K : \frac{\partial^2}{\partial x_i^2} - e^{-x_K}(c_K - r_K + \sum_{j \in K} r_j e^{x_j/r_j}) = -e^{-x_K}(c_K - r_K + \sum_{j \in K \setminus \{i\}} r_j e^{x_j/r_j} + \frac{(1-r_i)^2}{r_i} e^{x_i/r_i})$$

This establishes convexity in each x_i. Then, applying first-order conditions on each x_i:

$$\forall i \in K : x_i^*/r_i = \ln(c_K - r_K + \sum_{j \in K} r_j e^{x_j^*/r_j})$$

It follows that x_i^*/r_i is constant across $i \in K$. Then, $\forall i \in K : \sum_{j \in K} r_j e^{x_j^*/r_j} = r_K e^{x_i^*/r_i}$. Substituting in:

$$\forall i \in K : x_i^*/r_i = \ln(c_K - r_K + r_K e^{x_i^*/r_i})$$

$$e^{x_i^*/r_i} = c_K - r_K + r_K e^{x_i^*/r_i}$$

$$x_i^* = r_i \ln(\frac{c_K - r_K}{1 - r_K})$$

We once again find that all of these x_i^* values are valid choices, so we have found the global optimum.

Lemma 2.3.1. *A sub-coalition $K \subseteq N$ optimizes its value when each $i \in K$ plays action*

$$x_i^* = r_i \ln(\frac{c_K - r_K}{1 - r_K})$$

It bears mentioning that this value is quite similar to the non-cooperative optimal value $r_i \ln(\frac{c_i - r_i}{1 - r_i})$, but with the argument of the logarithm having been replaced with the coalition's values at large. Moreover, when summed over the coalition's members, this leads to a total sum action $x^*_K = r_K \ln(\frac{c_K - r_K}{1-r_K})$, which is identical to the non-cooperative optimal action in a hypothetical case in which the coalition is a unitary actor with the sum parameters of its members. Related to this observation is the note that adding a member j to K always increases all of its agents' optimal x^*_i values—$c_j > r_j$ implies $c_K \cup \{j\} - r_{K\cup\{j\}} > c_K - r_K$, and $1 - r_{K\cup\{j\}} < 1 - r_K$, so the argument of the logarithm always increases. This makes intuitive sense, because increasing any x_i generates a positive externality for all other agents.

The complementary sub-coalition optimizes identically, substituting $N\setminus K$ for K. Then, there is a unique Nash equilibrium to the synthetic two-player game between the two sub-coalitions, so $\mathcal{I}$ is well-defined. Applying the sum-payoff formula for $\mathcal{I}$ to this Nash equilibrium, K's value $\mathcal{I}(K)$ can be computed as follows:

$$\mathcal{I}(K) = \sum_{i\in K} -e^{-x^*_K - x^*_{N\setminus K}}(c_i + r_i(e^{x^*_i/r_i} - 1)) = -e^{-x^*_K - x^*_{N\setminus K}}(c_K - r_K + \sum_{i\in K} r_i e^{x^*_i/r_i})$$

$$= -e^{-r_K \ln(\frac{c_K - r_K}{1-r_K}) - r_{N\setminus K} \ln(\frac{c_{N\setminus K} - r_{N\setminus K}}{1-r_{N\setminus K}})}(c_K - r_K + r_K e^{\ln(\frac{c_K - r_K}{1-r_K})})$$

$$= -(\frac{c_K - r_K}{1-r_K})^{-r_K}(\frac{c_{N\setminus K} - r_{N\setminus K}}{1-r_{N\setminus K}})^{-r_{N\setminus K}}(\frac{c_K - r_K}{1-r_K})$$

$$= -(\frac{1-r_{N\setminus K}}{c_{N\setminus K} - r_{N\setminus K}})^{r_{N\setminus K}}(\frac{c_K - r_K}{1-r_K})^{1-r_K}$$

Theorem 2.3.2. *A proper sub-coalition $K \subset N$ has characteristic value*

$$\mathcal{I}(K) = -(\frac{1-r_{N\setminus K}}{c_{N\setminus K} - r_{N\setminus K}})^{r_{N\setminus K}}(\frac{c_K - r_K}{1-r_K})^{1-r_K}$$

and the grand coalition N has characteristic value

$$\mathcal{I}(N) = -(\frac{c-r}{1-r})^{1-r}$$

2.4 A Note on Notation & Definition of the Core

This paper uses the notational convention that under an allocation $\vec{T}$, each agent i gets total payoff T_i. This is different than a transfer-based notation, in which T_i would refer to the transfer paid by agent i (such that i would get a payoff of $u_i + T_i$). This choice is purely in the interest of result elegance and salience.

We define the core as usual with respect to the characteristic function $\mathcal{I}$: an allocation $\vec{T}$ is in the core iff it satisfies the following conditions:

- Feasibility, i.e. $\sum_{i \in N} T_i \leq \mathcal{I}(N)$.
- Stability, i.e. $\forall K \subseteq N : \sum_{i \in K} T_i \geq \mathcal{I}(K)$.

Note that the combination of these two conditions implies $\sum_{i \in N} T_i = \mathcal{I}(N)$ in any core allocation.

2.5 Core Existence Partial Order Over Sub-Partitions

Consider an arbitrary coalition satisfying the specification constraints, $N = \{1, ..., n\}$, with associated parameter values c_i and r_i for each $i \in N$. Now consider a modified coalition N' in which some agent $m \in N$ is "split" into two agents m' and m'', such that $r_m = r_{m'} + r_{m''}$ and $c_m = c_{m'} + c_{m''}$. Suppose that a core allocation T' exists for N'. Consider the allocation T on N which has $T_i = T'_i$ for $i \neq m$ and $T_m = T'_{m'} + T'_{m''}$.

Suppose T is blocked by some coalition $K \subseteq N$. Consider the corresponding coalition K' in N', where m' and m'' are in K' iff m is in K. It is easy to see that $r_K = r_{K'}$, $c_K = c_{K'}$, $r_{N \setminus K} = r_{N' \setminus K'}$, and $c_{N \setminus K} = c_{N' \setminus K'}$; then, $\mathcal{I}(K) = \mathcal{I}(K')$. It is further easy to show that $\sum_{i \in K} T_i = \sum_{i' \in K'} T'_{i'}$. Then, if K blocks T from N's core, K' blocks T' from N''s core; contradiction. Lemma 2.5.1 follows:

Lemma 2.5.1. *If $\exists$ a core allocation in N', then $\exists$ a core allocation in N.*

Induction on this result leads to Theorem 2.5.2:

Theorem 2.5.2. *If coalitions N_1 and N_2 are such that N_2 can be formed by partitioning members of N_1, then core existence in N_2 implies core existence in N_1.*

2.6 Scale Invariance of Core Existence

Consider two n-member groups N and N' parameterized respectively by $(c_i, r_i)_{i\in\{1,\ldots,n\}}$ and $(c'_i, r'_i)_{i\in\{1,\ldots,n\}}$. Suppose that N''s parameter set is transformed from N's in the following way:

- $\forall i : r'_i = r_i$.
- Each c'_i is chosen so that the ratio $\frac{c'_i - r'_i}{c_i - r_i} = \frac{c'_i - r_i}{c_i - r_i}$ is some positive constant α across all i. Explicitly, $\forall i : c'_i = \alpha c_i + (1-\alpha) r_i$.

It is clear that this scaling property extends to coalitions in N and N': for some subset $K \subseteq N$ with sum total values c_K and r_K, the corresponding coalition $K' \subseteq N'$ has sum total values

$$r'_{K'} = \sum_i r_i = r_K$$

and

$$c'_{K'} = \sum_i \alpha c_i + (1-\alpha) r_i = \alpha c_K + (1-\alpha) r_K$$

which implies $\frac{c'_{K'} - r'_{K'}}{c_K - r_K} = \alpha$. Plugging into the function $\mathcal{I}$ and simplifying, we find

$$\mathcal{I}(K') = \alpha^{1-r} \mathcal{I}(K)$$

This means that the transformation from N to N' entails only a constant positive scaling of $\mathcal{I}$. As such, the conditions for feasibility and stability apply to an allocation $\vec{T}$ in N if and only if they apply to a scaled allocation $\alpha^{1-r}\vec{T}$ in N'. Theorem 2.6.1 follows:

Theorem 2.6.1. *If N and N' are identical apart from their members' cost values c_i and c_i', and their members' c_i values are such that $\forall i : c_i' = \alpha c_i + (1-\alpha) r_i$ for some constant $\alpha > 0$, then a core allocation exists in N if and only iff a core allocation exists in N'.*

It is of note that as c values get large, this transformation quickly approaches a general scaling $c_i' \approx \alpha c_i$ because r_i is constrained below 1. The transient r_i term is an unfortunate byproduct of specification design which cannot be avoided without generating more significant inconveniences elsewhere[3]. In general, these transient effects are not particularly interesting or influential except in very extreme situations with tiny c_i values. These extreme situations are also particularly difficult to analyze via parametric variation because they are highly constrained by the lower bound $c_i > 1$. Thus, in Section 4.3, very large c scales are used while plotting for the dual purpose of approximating asymptotically invariant behavior and enabling higher-resolution plotting.

---

[3] Specifically, the r_i term can be removed by respecifying agents' utility functions from $-e^{-x}(c_i + r_i(e^{x_i/r_i} - 1))$ to $-e^{-x}(c_i + r_i e^{x_i/r_i})$. However, changes the cost of non-retaliation from 0 to r_i, which brings with it a host of issues.

3 UNILATERAL DEVIATION CONSTRAINT & SYMMETRIC CORE EMPTINESS

3.1 Unilateral Deviation Constraint

In general, the necessary and sufficient conditions for core formation comprise a system of inequalities with one for each potential sub-coalition. This is technically workable given the closed-form representation of $\mathcal{I}$ given in Theorem 2.3.2. However, it is difficult to form practical conclusions with this extended system. In this section, we consider a weaker necessary condition for core formation derived from this system, which identifies the potential for singleton sub-coalitions to deviate.

Summing the subset of stability condition inequalities stated in Section 2.4 wherein K is a singleton, we arrive at

$$\sum_{K\in\{\{i\}|i\in N\}}(\sum_{j\in K}T_j)\geq\sum_{K\in\{\{i\}|i\in N\}}\mathcal{I}(K)\therefore\sum_{i\in N}T_i\geq\sum_{i\in N}\mathcal{I}(\{i\})$$

In conjunction with the feasibility constraint, this implies the following lemma:

Lemma 3.1.1. *There exists a core allocation $\vec{T}$ only if*

$$\mathcal{I}(N)\geq\sum_{i\in N}\mathcal{I}(\{i\})$$

3.2 Symmetric Core Emptiness

Consider a special case in which $N=\{1,...,n\}$ and each deterrer i has $r_i=r/n$ and $c_i=c/n$ for some r and c such that r_i and c_i satisfy the necessary constraints. Each singleton sub-coalition has the same symmetric value

$$\mathcal{I}(\{i\})=-(\frac{1-\frac{n-1}{n}r}{\frac{n-1}{n}c-\frac{n-1}{n}r)})^{\frac{n-1}{n}r}(\frac{\frac{1}{n}c-\frac{1}{n}r}{1-\frac{1}{n}r})^{1-\frac{1}{n}r}=-(\frac{n}{n-1}-r)^{\frac{n-1}{n}r}(n-r)^{\frac{1}{n}r-1}(c-r)^{1-r}$$

so the sum over all such sub-coalitions is

$$\sum_{i \in N} \mathcal{I}(\{i\}) = -n(\frac{n}{n-1} - r)^{\frac{n-1}{n}r}(n-r)^{\frac{1}{n}r-1}(c-r)^{1-r}$$

Meanwhile, the value of the grand coalition is

$$\mathcal{I}(N) = -(\frac{c-r}{1-r})^{1-r}$$

Applying Lemma 3.1.1 and simplifying results in Corollary 3.2.0.1:

Corollary 3.2.0.1. *In a symmetric group of n deterrers each parameterized by common values c/n and r/n, there exists a core allocation $\vec{T}$ only if*

$$1 \leq n(1-r)^{1-r}(\frac{n}{n-1} - r)^{\frac{n-1}{n}r}(n-r)^{\frac{r}{n}-1}$$

We now go about showing that for sufficiently large n, this inequality cannot hold. The derivative of the right side of the inequality with respect to n is

$$\frac{r(1-r)^{1-r}(n-r)^{r/n}(\frac{n}{n-1} - r)^{\frac{n-1}{n}r}}{n(n-r)(n(1-r)+r)}((n(r-1)-r)\ln(\frac{n-r}{\frac{n}{n-1} - r}) - n)$$

Within the regime $0 < r < 1$, every term in the leading fractional term is positive, so it is positive. Then, note that $n(r-1) - r) < 0$, $\frac{n-r}{\frac{n}{n-1}-r} > 1 \therefore \ln(\frac{n-r}{\frac{n}{n-1}-r}) > 0$, and $-n < 0$. It follows that the latter term is the sum of two negative terms, so it is negative. Thus, the derivative is the product of a positive term and a negative term, so it is negative. Then, at any r, the right side of Corollary 3.2.0.1 is decreasing in n. Lemma 3.2.1 follows:

Lemma 3.2.1. *If the inequality in Corollary* 3.2.0.1 *is violated for some r and n, then it is violated at that r for any greater n as well.*

It can be computationally confirmed that the constraint is violated at all $r \in (0, 1)$ for

$n = 4$. Then, Theorem 3.2.2 follows:

Theorem 3.2.2. *No core allocation exists in any symmetric deterrer group of size* $n \geq 4$.

4 SINGLE-DIFFERING-ACTOR CORE

Consider the class of deterring groups $N = \{0, 1, ..., n\}$ in which one agent $i = 0$ has some parameters r_0 and c_0 and the rest of the agents $i = 1, ..., n$ split some r_{-0} and c_{-0} proportionally, but not necessarily equally—this is to say, for each $i \in 1, ..., n$, the ratio r_i/c_i is constant. In this section we will demonstrate sufficient conditions on r_0, c_0, r_{-0}, and c_{-0} such that a core allocation exists. It is of particular note that these conditions do not depend upon how the $i \in 1, ..., n$ group subdivides its total r_{-0} and c_{-0}, or moreover even how many members of the group there are. We will thus conclude that if the conditions hold, it can support core allocations in arbitrarily large coalitions[1], contrasting Theorem 3.2.2's impossibility result regarding core allocations in symmetric coalitions of size $n \geq 4$.

4.1 Characteristic Function in Single-Differing-Actor Groups

Consider an arbitrary deterring group N in the aforementioned class. Every sub-coalition $K \subseteq N$ either contains $i = 0$ or does not contain $i = 0$. By proportionality, if K contains $i = 0$, then it must have $r_K = r_0 + \alpha_K r_{-0}$ and $c_K = c_0 + \alpha_K c_{-0}$ for some $\alpha_K \in [0, 1]$. If K does not contain $i = 0$, then it must have $r_K = \alpha_K r_{-0}$ and $c_K = \alpha_K c_{-0}$ for some $\alpha_K \in [0, 1]$. The characteristic function $\mathcal{I} : 2^N \to \mathbb{R}$ can thus be reworked into an equivalent function $\mathcal{J} : \{0, 1\} \times [0, 1] \to \mathbb{R}$ wherein the input arguments (d_0, α_K) of $\mathcal{J}$ are produced from a given coalition $K \subseteq N$ as follows:

- d_0 is the indicator $\mathbf{1}_{0 \in K}$.
- α_K is the value $\frac{1}{r_{-0}} \sum_{i \in \{1,...,n\}} r_i \mathbf{1}_{i \in K}$.

---

[1] This size *is* technically constrained by the $c_i > 1$ condition, which implies that c_1 can only be subdivided across at most $\lfloor c_1 \rfloor$ agents. However, c_1 can be made arbitrarily large.

Then, $\mathcal{J}$ may be written in closed form as

$$\mathcal{J}(d_0, \alpha_K) = \begin{cases} (\frac{1-r_0-(1-\alpha_K)r_{-0}}{c_0-r_0+(1-\alpha_K)(c_{-0}-r_{-0})})^{r_0+(1-\alpha_K)r_{-0}}(\frac{\alpha_K(c_{-0}-r_{-0})}{1-\alpha_K r_{-0}})^{1-\alpha_K r_{-0}} & d_0 = 0 \\ (\frac{1-(1-\alpha_K)r_{-0}}{(1-\alpha_K)(c_{-0}-r_{-0})})^{(1-\alpha_K)r_{-0}}(\frac{c_0-r_0+\alpha_K(c_{-0}-r_{-0})}{1-r_0-\alpha_K r_{-0}})^{1-r_0-\alpha_K r_{-0}} & d_0 = 1 \end{cases}$$

4.2 Single-Differing-Actor Core Existence Conditions

It is of particular note that the value of the grand coalition $\mathcal{I}(N) = \mathcal{J}(1, 1)$ only depends upon c_0, c_{-0}, r_0, and r_{-0}, but does not depend upon the way in which c_{-0} and r_{-0} are subdivided. We now derive conditions under which a core allocation scheme can be designed purely as a function of these parameters—clearly, these conditions are also sufficient for general core existence.

Consider a quasi-proportional allocation $\vec{T}(c_0, c_{-0}, r_0, r_{-0})$ with T_0 going to agent 0 and a total of T_{-0} split among the remaining agents in the same proportionality as the splitting of c_{-0} and r_{-0}. Then, any coalition characterized by (d_0, α_K) has sum total allocation $d_0 T_0 + \alpha_K T_{-0}$. To be in the core, $T_0 + T_{-0} = \mathcal{I}(N)$.

Any single-differing-actor N has some fixed set of α_K values in the interval $[0, 1]$ which can be formed by its sub-coalitions. Then, if $\vec{T}(c_0, c_{-0}, r_0, r_{-0})$ is stable with respect to *all* α_K values in the interval $[0, 1]$, it must be stable with respect to every single-differing-actor N constructed with those parameters. This general-stability condition can be written

$$\forall (d_0, \alpha_K) \in \{0, 1\} \times [0, 1] : \mathcal{J}(d_0, \alpha_K) \leq d_0 T_0 + \alpha_K T_{-0}$$

At $(d_0, \alpha_K) = (1, 1)$, this condition is trivially satisfied because both sides are $\mathcal{I}(N)$. At $(d_0, \alpha_K) = (0, 0)$, this condition is trivially satisfied because both sides are zero[2]. The remaining cases form two meaningful sets of conditions on T:

---

[2] These edge cases are noted because they prevent division by $1 - \alpha_K$ and α_K in the algebraic simplification of the conditions which follow.

- If $(d_0, \alpha_K) \in \{0\} \times (0, 1]$, then

$$\frac{\mathcal{J}(0, \alpha_K)}{\alpha_K} \leq T_{-0}$$

- If $(d_0, \alpha_K) \in \{1\} \times [0, 1)$, then

$$T_{-0} \leq \frac{\mathcal{I}(N) - \mathcal{J}(1, \alpha_K)}{1 - \alpha_K}$$

If T_{-0} satisfies these inequalities everywhere on the relevant domains, then it satisfies the conditions for core formation. Thus, if such a T_{-0} exists, then a core allocation exists. Theorem 4.2.1 follows:

Theorem 4.2.1. *Consider a deterring group* $N = \{0, 1, ..., n\}$ *such that*

$$\forall i, j \in \{1, ..., n\} : \frac{c_i}{c_j} = \frac{r_i}{r_j}$$

If

$$\sup_{\alpha_K \in (0,1]} \frac{\mathcal{J}(0, \alpha_K)}{\alpha_K} \leq \inf_{\alpha_K \in [0,1)} \frac{\mathcal{I}(N) - \mathcal{J}(1, \alpha_K)}{1 - \alpha_K}$$

then a core allocation exists. Moreover, this allocation exhibits the property that

$$\forall i, j \in \{1, ..., n\} : \frac{T_i}{T_j} = \frac{r_i}{r_j} = \frac{c_i}{c_j}$$

4.3 Plots Related to Single-Differing-Actor Core

Despite its mathematical convenience, the condition described in Theorem 4.2.1 is intuitively incomprehensible. In order to address this, we briefly turn now to a computational approach.

First, it is helpful to understand the sorts of shapes taken by the bounding functions $\frac{\mathcal{J}(0,\alpha_K)}{\alpha_K}$ and $\frac{\mathcal{I}(N)-\mathcal{J}(1,\alpha_K)}{1-\alpha_K}$. Figure 4.1 charts these functions, as well as their relevant suprema and infima, on the interval $[0,1]$ for three parametric setups. In the leftmost setup, the condition in Theorem 4.2.1 holds, because the red dotted line is above the blue dotted line. In the middle setup, the condition narrowly fails. In the rightmost setup, the condition fails by a comparatively large margin. The rightmost setup is also notable insofar as it illustrates the non-monotonicity of both functions on the interval: the red curve reaches a local minimum at approximately $\alpha_K = .1$, while the blue curve reaches a local minimum at approximately $\alpha_K = .8$. In the absence of this non-monotonicity, the condition in Theorem 4.2.1 could be reduced to comparison at the $\alpha_K = 0$ point, i.e.

$$\lim_{\alpha_K \to 0^+} \frac{\mathcal{J}(0, \alpha_K)}{\alpha_K} \leq \mathcal{I}(N) - \mathcal{J}(1, 0)$$

That being said, for reasons unknown to the author, this reduced condition still seems to be equivalent to the original across all cases tested. This is an intriguing observation, albeit one which must be left as a conjecture until further investigation.

Conjecture 4.3.1. *The inequality condition*

$$\sup_{\alpha_K \in (0,1]} \frac{\mathcal{J}(0, \alpha_K)}{\alpha_K} \leq \inf_{\alpha_K \in [0,1)} \frac{\mathcal{I}(N) - \mathcal{J}(1, \alpha_K)}{1 - \alpha_K}$$

in Theorem 4.2.1 is equivalent to the reduced condition

$$\lim_{\alpha_K \to 0^+} \frac{\mathcal{J}(0, \alpha_K)}{\alpha_K} \leq \mathcal{I}(N) - \mathcal{J}(1, 0)$$

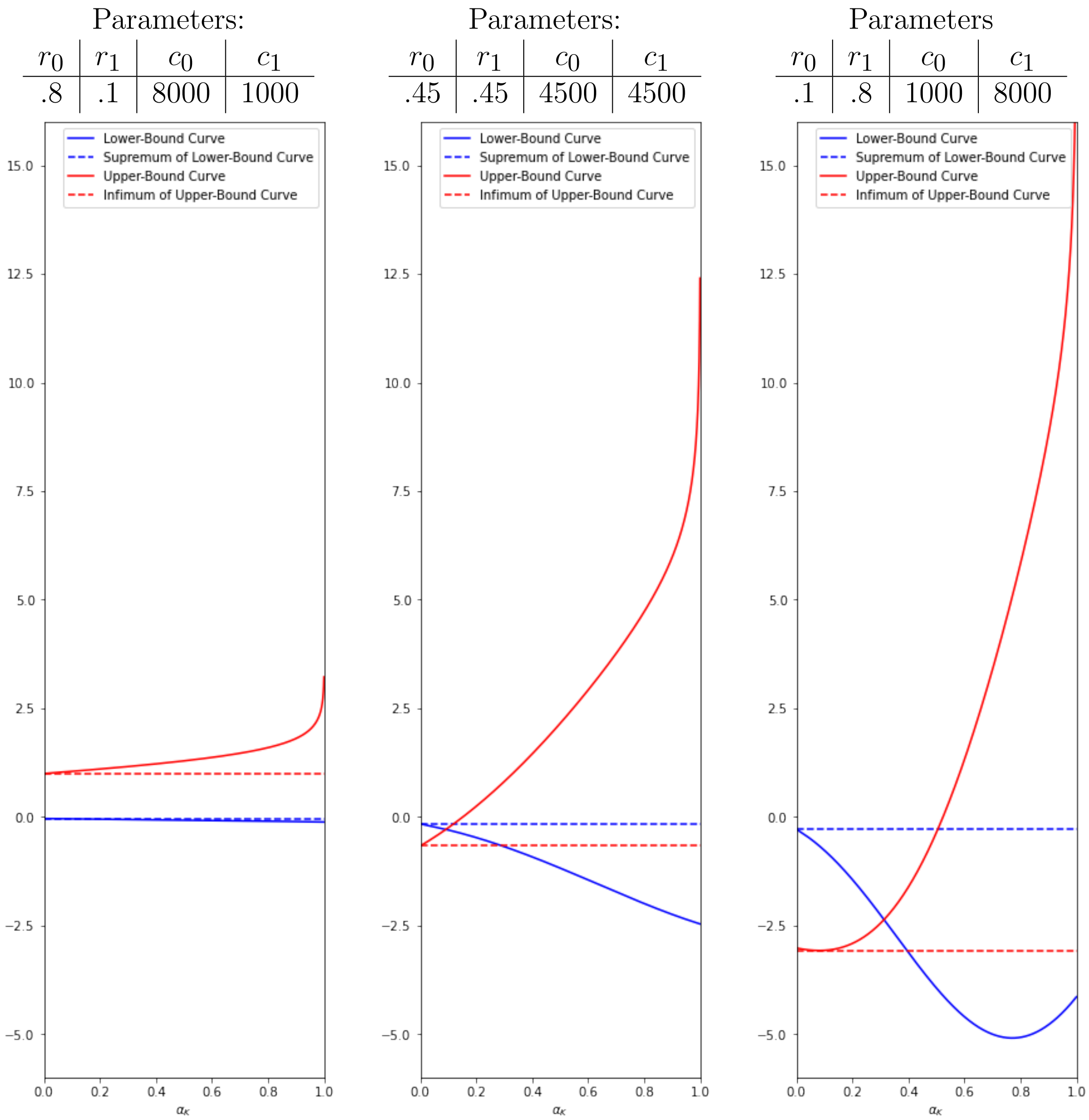
Figure 4.1 identifies the existence of a parametric setup which satisfies the condition for single-differing-actor core existence, and two setups which do not. However, with only three setups, it is impossible to develop a deeper understanding of what characterizes a parametric setup which supports a single-differing-actor core. In Figure 4.2, a much larger number of setups are examined on the continuum of possible divisions of a fixed total $r = .9$

and $c = 100000$ value, with the horizontal coordinate representing the single differing actor's share of the group's total r and the vertical coordinate representing the single differing actor's share of the group's total c. Points are colored in if their corresponding parametric setup supports a single-differing-actor core.

This representation begins to address the intuition gap. We see that core existence is broadly dependent on the single differing actor being sufficiently parametrically "dominant" in the coalition, except at extremely high r shares (on the right edge of the graph). This notion of a dominant actor being necessary for stability complements the symmetric impossibility result in Theorem 3.2.2; in a symmetric coalition, no agent is dominant by definition.

The shape of Figure 4.2 is, of course, only the outcome of a one specific choice of r and c. In order to draw more general results, it is important to consider other choices. As discussed in Section 2.6, the choice of c is not particularly significant unless it is very small, and the transient effects at small c are likely quirks of specification as opposed to products of realism. However, the effect of the choice of r on core existence is not yet well understood. To address this question, Figure 4.3 plots the same shape plotted in Figure 4.2, but across all r in the set $\{.999999, .9, .7, .5, .3, .1, .000001\}$.

Figure 4.3 makes it clear that the choice of r matters a great deal. At higher r values, i.e. when the coalition is overall comparatively capable of effective retaliation, the single-differing-actor core tends to depend on a combination of dominance in both r and c share. This is highlighted by the asymptotic case approximated with $r = .999999$, where the single-differing-actor core condition boils down to roughly $\frac{r_0}{r_0+r_1} + \frac{c_0}{c_0+c_1} \geq 1$. Meanwhile, at lower r values, i.e. when the coalition is overall comparatively incapable of meaningful retaliation, the single-differing-actor core starts depending more on pure c dominance, with very low r values even punishing r dominance slightly. This is highlighted in the asymptotic case approximated by $r = .000001$, where the frontier of core existence is actually slightly upward-sloped. Finally, it is notable that in all cases but the asymptotic $r = .999999$ case, the characteristic upward swoop at the right edge of Figure 4.2 persists. It is likely that the swoop exists even in this extreme case, but that the plotting resolution is simply too low to display it correctly. This swoop is difficult to explain intuitively in the overall useful language of parametric dominance. However, its persistence suggests that it merits further investigation.



Parameters:

r_0	r_1	c_0	c_1
.8	.1	8000	1000

Parameters:

r_0	r_1	c_0	c_1
.45	.45	4500	4500

Parameters

r_0	r_1	c_0	c_1
.1	.8	1000	8000

Figure 4.1: Three example plots which illustrate application of Theorem 4.2.1's condition for single-differing-agent core existence. Horizontal axis is the value α_K. Solid blue and red lines are respectively the lower-bound function $\frac{\mathcal{J}(0,\alpha_K)}{\alpha_K}$ and the upper-bound function $\frac{\mathcal{I}(N)-\mathcal{J}(1,\alpha_K)}{1-\alpha_K}$. Dashed lines identify the relevant extreme values necessary to evaluate core existence: Theorem 4.2.1 guarantees single-differing-agent core existence if the red dashed line is above the blue dashed line.

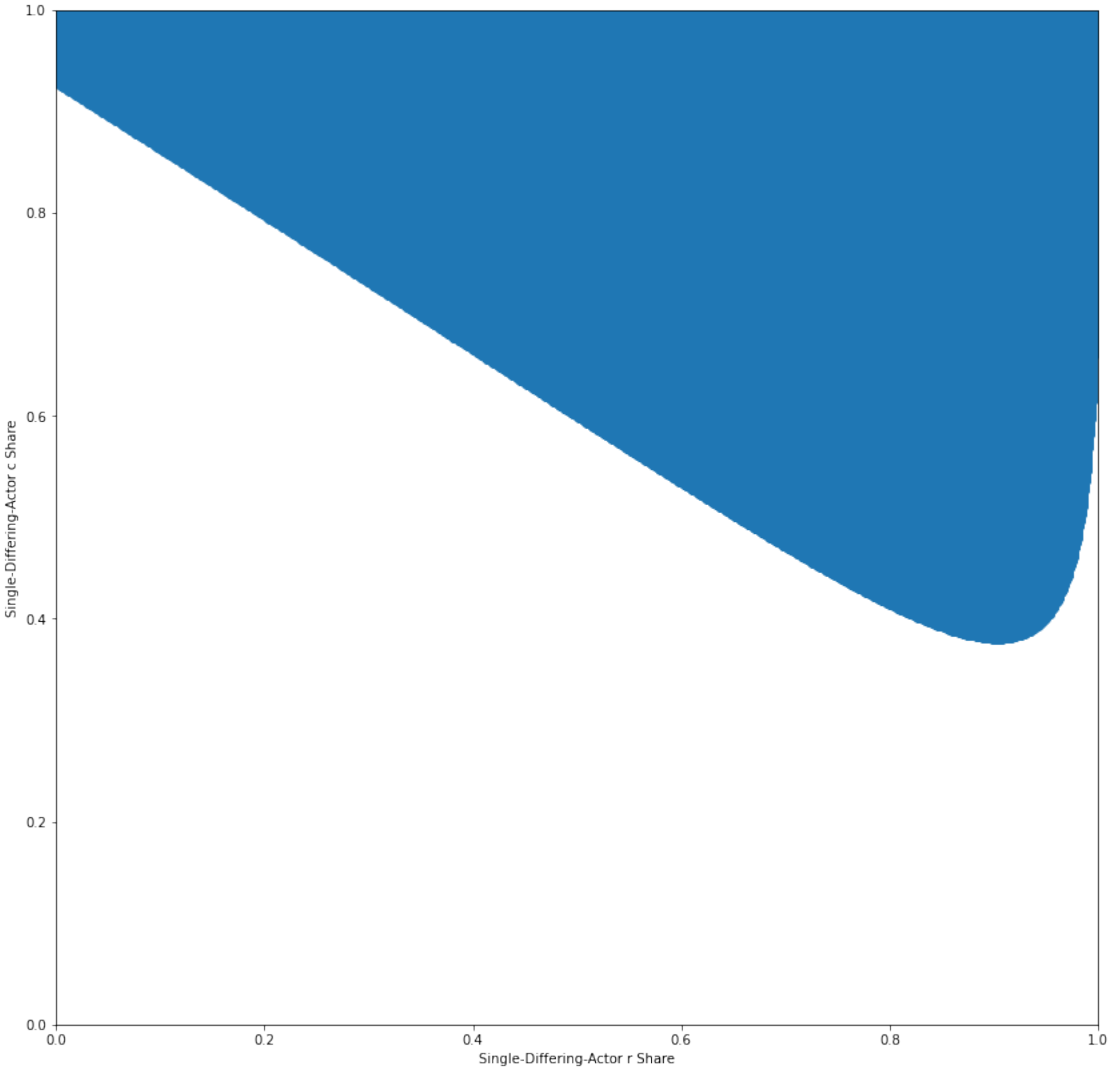


Figure 4.2: Parametric area which supports single-differing-actor core existence for total $r = .9$ and $c = 1000000$. Horizontal coordinates correspond to r_0's share of the group's total r, and vertical coordinates correspond to c_0's share of the group's total c. Thus, the point (x, y) refers to a single-differing-actor group with $(r_0, r_1, c_0, c_1) = (xr, (1-x)r, yc, (1-y)c)$. Points are filled if they support single-differing-actor core existence based on Theorem 4.2.1. Large c value chosen to support sufficient subdivision to produce high-resolution plot; recall that asymptotic scaling of c values produces negligible effects on core existence as discussed in Section 2.6.

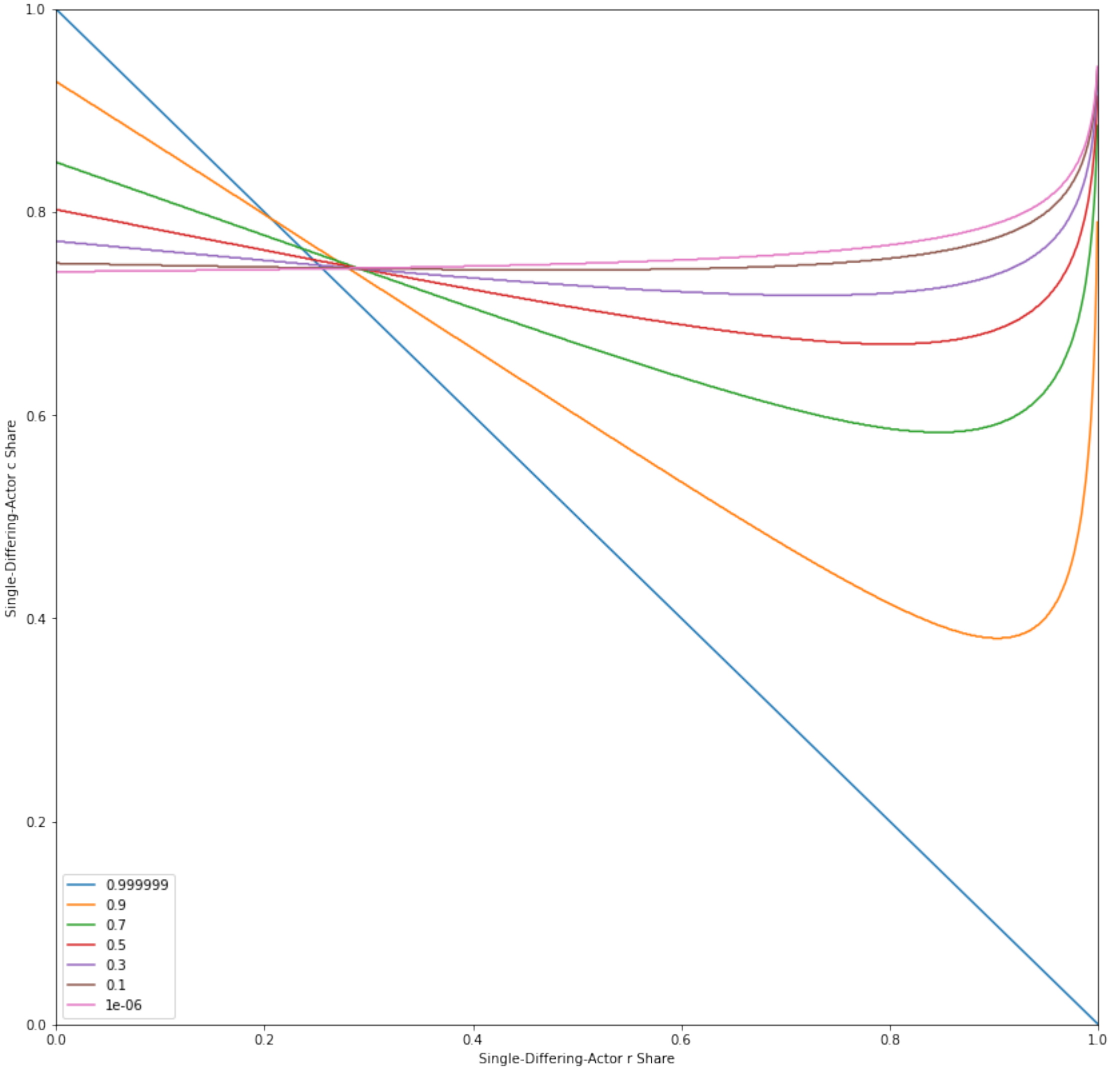


Figure 4.3: Frontiers of parametric areas which support single-differing-actor core existence for total $r \in \{.999999, .9, .7, .5, .3, .1, .000001\}$ and $c = 1000000$. Areas left unfilled to improve legibility; in all cases, the area in question is that which lies above the relevant frontier. All other plotting methodology is identical to that employed in Figure 4.2.

5 CONCLUSION

In this paper, we use a parametric model of multilateral deterrence to study the relationship between parametric features of the deterring group and core existence. We begin by comparing non-cooperative equilibrium actions with cooperative value-maximizing actions: cooperative deterrence levels are always greater than non-cooperative levels in any non-singleton group, which makes sense because deterrence generates a positive externality for other agents. Next, we establish a partial order on groups which can be formed by parametric partition, and show that this partial order implies core existence. We also identify a core-existence equivalence relation between groups whose c_i values are scaled appropriately. This scaling is asymptotically linear but not precisely so, due to transient effects of model specification; said asymptotic linearity suggests a rough scale invariance at large c values.

We then discuss two contrasting phenomena which arise in this model. First, we show that no symmetric coalition of size $n \geq 4$ can support a core allocation. Next, we construct a notion of a single-differing-actor group, and demonstrate conditions in such a group which guarantee core existence; application of these conditions leads to the notional conclusion that a single dominant member in a group can support a core allocation regardless of how the rest of the group is subdivided, so long as this subdivision is proportional across parameters. The curious contrast between the instability of even moderately-sized symmetric groups and the stability of arbitrarily large groups with a single dominant member forms the critical intuitive takeaway of this paper.

Applied to international relations, this conclusion suggests that alliances are comparatively stable when dominated by a leading member, and comparatively unstable when members are on roughly even footing. The separate partial-order conclusion also implies that alliances are more stable when their resources are subdivided among fewer members; of course, the extreme case which demonstrates this is an alliance which constitutes a single unitary deterrer, which is trivially stable. We therefore might expect to see more consistent deterrent cooperation from international organizations resembling NATO (in which the United States is indisputably a dominant member) than the European Union, which otherwise resembles NATO but conspicuously excludes the United States. We also might expect to see particularly consistent cooperation in less numerous alliances, such as that which exists between the U.S. and Israel.

Beyond international relations, multilateral deterrence arises in a few other arenas. Market entry deterrence is a well-studied empirical example, in which firms collude to deter new entrants. Mapped onto this topic, the conclusions of this paper would suggest that market entry deterrence is more likely in markets which are highly oligopolistic or even captured largely by one quasi-monopolist. While this result confirms obvious priors about market power, it should be cited with reservation due to this paper's model's comparatively unsophisticated retaliatory mechanism; better models of market manipulation exist. Still, integration of those models with the cooperative framework studied here might yield interesting results. Organized labor union strikes—or threats thereof—are another potentially interesting application of this model, where agents balance their personal cost of striking with the opportunity cost of not enforcing negotiating leverage against an employer. A reframing of this labor situation might also be useful to address political jockeying between a unitary policymaker and a popular movement threatening revolt as well. In both cases, this paper's core takeaways can be straightforwardly applied to the cooperative stability of the deterring group (the labor union or the popular movement).

In a mathematical sense, the analysis in this paper is ripe to be expanded upon. While the focus of this paper is the question of core existence, a notion of welfare based on negotiating ranges in nonempty-core cases might prove interesting as an object of investigation—for instance, in single-differing-actor core cases with a dominant actor, is the dominant actor paying the others on net, or being paid by them? Does this pay increase or decrease as the differing actor becomes more dominant? More generally, how much do actors benefit in a welfare sense from cooperation, and how does this change with group asymmetry? These questions are secondary in the context of this paper, but the machinery for addressing them is still largely fleshed out during the investigation of core existence. They are also of particular applied importance, as a tool for validating or invalidating this model as a representation of reality: if study of single-agent-dominated alliances such as NATO or the late Warsaw Pact reveals empirical transfer behavior which differs markedly from expected core transfer behavior, we may conclude that there is a meaningful feature at play in reality which this model does not accurately capture.

Another venue for potential future work is model refinement. While discussing scale invariance, we briefly mention that an alternate agent utility function

$$u_i(\vec{x}) = -e^{-x}(c_i + r_i e^{x_i/r_i})$$

would produce truly c-scale-invariant core existence results at the expense of making zero commitment have positive fixed cost. Amid model development, a best-of-both-worlds solution to this specification dilemma was identified but not pursued due to realism concerns: use the modified utility function listed above, but unbind x_i from non-negativity (that is, let each agent choose $x_i \in \mathbb{R}$). In this form, $-\infty$ becomes the new "zero commitment", and the only parametric constraints necessary to guarantee good behavior are $r_i > 0$, $c_i > 0$, and $r < 1$. The realism issue which emerges is that negative x values can arise, which implies a greater-than-one probability of aggression. This can be rationalized away in any specific case with a sufficiently small leading multiplier on the probability value e^{-x}; however, it still undeniably sacrifices realism for mathematical elegance, which it has in spades. A theoretical investigation using this modified form would likely reveal substantially simpler and clearer results.

Beyond refinement of the original model, more substantive adjustments can be made which may yield interesting new results. For instance, deterrence is usually not actually free in the case of non-aggression—to address this, a cost dependent upon x_i but not upon the aggression probability e^{-x} might be a warranted and interesting inclusion into the agents' utility functions. However, when including one or more such modifications, care must be taken to avoid over-parameterizing. The key feature of this paper's model which makes it workable is its relatively few degrees of parametric freedom per actor (two, and one of them is asymptotically scale-invariant). This simplifies the problem and allows for more precise questions about heterogeneity to be addressed.

One final substantive adjustment which might yield interesting and more realistic results deals with information structure. The ability of the agents to cooperate and enact a mutually beneficial core transfer scheme relies on the group's common knowledge of all parameters. In practice, this is a significant assumption. A natural conjecture is that implementing private parameter values with noisy public signals would decrease effective core size by impeding negotiation, due to a version of the phenomenon which impedes efficient allocation in the Myerson-Satterthwaite bilateral trading model (Myerson & Satterthwaite, 1983). The existence and extent of this decrease is an interesting potential point of future analysis.

REFERENCES

Parkash Chander. Cores of games with positive externalities. *Center for Operations Research and Econometrics*, 2010.

James D. Fearon. Rationalist explanations for war. *International Organization*, 49(3):379—414, 1995.

Todd Sandler & Keith Hartley. Economics of alliances: The lessons for collective action. *Journal of Economic Literature*, 39(3):869–896, 2001.

Matthew O. Jackson & Massimo Morelli. The reasons for war—an updated survey. *Handbook on the Political Economy of War*, 2009.

Oskar Morganstern & John Von Neumann. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Roger Myerson & Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281, 1983.

Giorgos Stamotopoulos. The core of aggregative cooperative games with externalities. *The B.E. Journal of Theoretical Economics*, 2014.