# Nash Bargaining in Coalitional Games

Rajiv Vohra ⓡ Debraj Ray[†]

April 2025

**Abstract.** We revisit Nash's axiomatic bargaining theory when both individuals and coalitions have outside options. As in Nash, our solution maximizes a (possibly weighted) product of payoffs net of individual disagreements, but coalitional threats appear as conventional constraints that are not netted out. We embed this solution into a setting with externalities, and develop a "Nash-in-Nash" theory of viable coalitional structures. Every coalition follows its coalitional Nash solution but interacts noncooperatively with other coalitions. We discuss applications to public goods, R&D coalitions, and cartels. Finally, for transferable utility games, we connect the coalitional Nash solution to a notion of "pragmatic egalitarianism."

**JEL classification:** C70, D60, D63

**Keywords:** Nash bargaining, coalitional threats, strategic externalities

## 1. Introduction

Nash (1950) introduced and axiomatically derived his celebrated bargaining solution, which describes a procedure for choosing a payoff division from a feasible set of $n$-person payoff outcomes.[1] As is well known, the procedure involves first subtracting individual disagreement payoffs from each available payoff vector, thereby generating *net* payoffs, and then maximizing the product of net individual payoffs over the feasible set.

The goal of this paper is to extend the Nash bargaining solution to situations in which *coalitions*, not just individuals, can walk away from the bargaining table. To us, an understanding of Nash bargaining with coalitional outside options seems both worthwhile and challenging from a variety of perspectives.

---

[†]Ray: New York University and University of Warwick; debraj.ray@nyu.edu. Vohra: Brown University; rajiv_vohra@brown.edu. We are grateful to Nemanja Antic, Piotr Dworczak, Marco Mariotti, David Pearce, Ariel Rubinstein and Roberto Serrano for insightful comments. Ray acknowledges support under National Science Foundation Grant SES-2315720.

[1]Nash established his solution for $n = 2$, but it is easily seen to extend to an arbitrary number of individuals. See Thomson (1994, 2022), for surveys of the extensive literature that followed Nash (1950).

First, we do not have a widely applicable account of how a group of agents would divide a surplus, when sub-coalitions of *two or more* are in a position to thwart or "block" various agreements. We seek a general axiomatic solution analogous to the classical Nash solution. Currently, an entire applied literature — on which more below — is restricted to the study of agreements under the shadow of individual disagreements alone. We are interested in a theory that goes beyond this restriction.

Second, such a theory should ideally be embeddable into settings with pervasive cross-coalitional externalities. For instance, suppose that four firms are deciding whether or not to form a monopolistic cartel, and a "subcartel" of two firms threatens to abandon the negotiations. Their coalitional outside options are mediated by their interactions with the other two firms, and in particular depends on whether those firms (and indeed, the original deviating pair) will remain together or not. These interactions in turn will depend on how payoffs are generated *and* allocated in the subcartel-subcartel duopoly.

Third, "split-the-difference" is a fundamental property of Nash's solution, with individual disagreement payoffs netted out before the surplus is divided. This feature emerges notably from the presumption that any solution must be invariant to affine transforms of individual expected utility payoffs. In applications and even in related theory, split-the-difference has assumed a distinctly moral significance[2] — bargaining "should" set aside what we can get for ourselves, and "should" focus on dividing what remains. Is this a robust philosophical interpretation of the Nash bargaining solution? Might split-the-difference extend to the "subtraction" of coalitional threats, whatever that might mean in this vector-valued context?

Our tasks in this paper can therefore be summarized as:

(a) To characterize an analogue of the bargaining solution in a setting with individual and coalitional threats;

(b) To embed that solution into a setting with strategic externalities across coalitions, allowing us to derive viable coalition structures and the payoff allocations for coalitions in those structures in one unified exercise; and

(c) To explore our extended Nash solution from a philosophical perspective; in particular, to examine if it lines up with an ethic that asks for outside options to be "subtracted" before dividing the remaining surplus in some way.

[2] See, e.g., the "disputed garment principle" in Aumann and Maschler (1985), or the foundations of the pre-kernel in Davis and Maschler (1965).

We study *games with coalitional threats*. Each game is a triple $G = (F, \mathbf{\Theta}, d)$, where $F$ is a set of feasible payoff vectors for the grand coalition of all players, $\mathbf{\Theta}$ collects sets $\Theta(S)$ of payoff vectors (or threats) available to each coalition $S$, and $d$ is a vector of individual disagreement payoffs. We generally follow Nash's original setup, with the distinct exception of a new axiom that pertains to coalitions. In Section 2, we prove that a solution $\sigma(G)$ satisfies our axioms if and only if there exist strictly positive weights $(\gamma_1, \ldots, \gamma_n)$ such that

$$\sigma(G) = \arg\max \prod_{j \in N} [x_j - d_j]^{\gamma_i} \tag{1}$$

over the set of all allocations $x \in F$ that are unblocked by some coalitional threat $(S, y)$, where $S$ is a strict subset of $N$ and $y \in \Theta(S)$. See Theorem 1 for a precise statement, and Corollary 1 for the reduction to equal weights under a symmetry axiom.

This derivation uncovers a central point. While individual disagreement payoffs continue to be subtracted from overall payoffs, as they are in Nash's theorem, coalitional threats appear as standard constraints *that bind in the conventional way*. A threat $y$ affects the solution $x$ only when the solution lies on the boundary of the unblocked set; that is, $x_i = y_i$ for some $i$. In contrast, a change in the disagreement vector $d$ affects the solution $x$ even when $x \gg d$.[3]

In Section 3, we embed our solution into games in strategic form, in which coalitions partition players into a coalition *structure* $\pi$. A coalition interacts non-cooperatively with other coalitions in $\pi$, but seeks cooperation within itself while contending with threats from potential breakaway subcoalitions. A *Nash-in-Nash equilibrium* for $\pi$ is one in which every coalition in $\pi$ interacts strategically with other coalitions in $\pi$ while dividing its own surplus according to the coalitional Nash solution in (1), constrained by the potential threats from breakaways. A coalition structure $\pi$ is *viable* if it admits a Nash-in-Nash equilibrium.

Our theory combines cooperation within coalitions with non-cooperative play across them, as in the applied literature; e.g., Horn and Wolinsky (1988), Collard-Wexler, Gowrisankaran and Lee (2019), or Bagwell, Staiger and Yurukoglu (2020). But it has novel distinguishing features. First, the applied Nash-in-Nash literature only accommodates individual outside options, which is hardly surprising given the absence of a Nash bargaining theory with coalitional threats. In our theory, the coalitional Nash solution is invoked within every coalition, with an eye on subcoalitional threats. Second, those threats are determined in a

[3]A possibly confusing but ultimately irrelevant detail concerns the connection between the disagreement payoff $d_i$ and an individual threat, as embodied in $\Theta(\{i\})$. Section 2.5 discusses this point.

consistent manner, using parallel concepts of viability for *sub*structures of $\pi$, so that the entire edifice of solutions is built recursively from the finest coalition structure upwards.[4]

The theory shows how no viable structure might be efficient, because potential breakaway threats might overpower what some coalitions in any efficient structure can achieve. More than that, it sets up a recursive structure that allows us to compute, in principle, the viable coalition structures as well as payoff allocations for any game that admits a (traditional) non-cooperative Nash equilibrium. In Section 4, we apply the model to three examples: global public goods, R&D coalitions, and cartels in oligopolistic competition.

In Section 5, we show that our recursive coalitional Nash solution simplifies significantly in the absence of externalities. If all feasible payoff sets are convex, our solution is equivalent to one that maximizes (1) over all allocations unblocked using the *unconstrained* Nash outcome for each coalition (Theorem 2). That is, this artificial but easy-to-solve problem delivers the "correct" answer for the recursively defined solution. Theorem 3 shows that in transferable utility (TU) games satisfying grand coalition superadditivity, there is an efficient coalitional Nash solution (under a mild consistency condition on bargaining weights). This is an unexpected finding, because it is well-known that core of superadditive games may be empty, unless stronger restrictions are imposed. As we explain, Theorem 3 deeply relies on our use of the coalitional Nash solution within all coalitions, which restricts credible threats for breakaways to a degree that permits efficiency with no further assumptions.

Theorem 2 also allows us to examine the question of whether split-the-difference of the symmetric Nash solution can be viewed as a plausible moral imperative. Our solution, in which coalitional constraints are conventional, suggests a different "pragmatic ethic": move as close as possible to equality in *overall* payoffs (not net of outside options), unless prevented from doing so by the possibility that some coalition will walk away. There is a close connection between the symmetric Nash-in-Nash solution and the egalitarian solution introduced in Dutta and Ray (1989, 1991). Their notion of "constrained egalitarianism" combines a commitment to equity with the pragmatics of coalitional participation. The idea is to apply a social norm (egalitarianism) to the greatest extent possible, while remaining bound by the need to seek individual and coalitional buy-ins. Specifically, Dutta and Ray

[4] In this respect, the construction resembles refinements of Nash equilibrium such as coalition-proof Nash equilibrium studied in Bernheim, Peleg and Whinston (1987), or renegotiation-proof equilibrium (Bernheim and Ray (1989), Farrell and Maskin (1989)) except that our solution allows for binding agreements and are far from refinements of Nash equilibrium. It also resembles equilibrium binding agreements (EBA) studied in Ray and Vohra (1997), except that our solution inserts a norm — the coalitional Nash solution — into every coalition that forms. EBAs have no within-coalition norm built in.

(1989, 1991) argued that the grand coalition would choose unmajorized or Lorenz-maximal elements from its set of unblocked allocations. At the same time, because every coalition is also presumed to subscribe to egalitarianism, any credible block would also need to be egalitarian (for the coalition doing the blocking), just as we ask of the grand coalition. This yields a "pragmatic" or "constrained" *egalitarian solution*.

Theorem 4 asserts that our coalitional Nash solution is a subset of the egalitarian solution.

We end the Introduction with two remarks. Theorem 1 relies on a new "expansion axiom" that we append to the otherwise standard arsenal of the Nash axioms. An axiomatic characterization is appealing if on the one hand, the axioms appear to be reasonable or intuitive, while at the same time their mathematically equivalent outcome is surprising in that it delivers new intuition. The reader must judge whether the expansion axiom is satisfactory in this sense, while retaining the simplicity of Nash's original characterization.

Second, as a historical footnote, von Neumann was appreciative of Nash's foray into cooperative bargaining; see the acknowledgements in Nash (1950). The same reaction was decidedly not forthcoming for Nash's particular notion of non-cooperative equilibrium.[5] As Bhattacharya (2022) notes in his biography of von Neumann, "The idea that people might not work together for mutual benefit was anathema to him . . . [and] ran counter to the spirit of his 'coalitional' conception of game theory." So while the notion of Nash equilibrium does not seem to have worked for von Neumann, it seems likely that he would have welcomed an attempt to go beyond two-person situations and connect the Nash solution to his "coalitional conception of game theory." Our paper is one such attempt.

## 2. Nash Bargaining With Coalitional Threats

An $n$-person *game with coalitional threats* is $G = (F, \mathbf{\Theta}, d)$, where $F \subset \mathbb{R}^N$ is a set of feasible payoff vectors for the grand coalition $N$,[6] $\mathbf{\Theta} = \{\Theta(S)_{S\subset N}\}$ are sets of *threats* in $\mathbb{R}^S$ for each subcoalition $S$, and $d \in \mathbb{R}^N$ is a vector of individual disagreement payoffs. For those familiar with cooperative game theory, $G$ can be interpreted as a characteristic function, with $\Theta(S)$ being some set of exogenous payoffs that coalition $S$ can "guarantee

[5]But a general sense of ambivalence is evident in von Neumann's letter to David Gale, dated November 5, 1949: "The idea of discussing $n$-person games 'without cooperation' is an attractive one that had occurred to me too, but I gave it up in favor of the coalition-type procedure because I didn't see how to exploit it." (von Neumann 1949). We thank Dirk Bergemann and Philipp Strack for bringing this letter to our attention.

[6]We use $\subseteq$ to refer to "subset" and $\subset$ to refer to "strict subset". A subcoalition ia a strict subset of $N$. For $S \subseteq N$, $\mathbb{R}^S$ denotes $|S|$-dimensional Euclidean space with coordinates indexed by the elements of $S$.

on its own." Indeed, in this Section we do regard $\boldsymbol{\Theta}$ as exogenous. However, in Section 3, we will endogenize $\boldsymbol{\Theta}$ and embed $G$ within a game in strategic form.

An allocation $x \in F$ is *blocked* by the threat $(S, y)$ if $y \in \Theta(S)$ and $y \gg x_S$.[7] It is *unblocked* if it is not blocked by any threat. For $G = (F, \boldsymbol{\Theta}, d)$, define its *unblocked set* by

$$U(G) \equiv \{x \in F | x \text{ is not blocked by any threat } (S, y)\}.$$

Under the characteristic function interpretation of $G$, $U(G)$ is the core of $G$.[8]

2.1. **Domain.** We consider the universe of all conceivable games $G = (F, \boldsymbol{\Theta}, d)$ such that:

[Dom 1] $F$ is nonempty and compact, and contains some $x \gg d$.

[Dom 2] For each subcoalition $S$, $\Theta(S)$ is nonempty and compact with $z \geqslant d_S$ for every $z \in \Theta(S)$. In particular $\zeta_i \equiv \max \Theta(\{i\})$ is well defined, and $\zeta_i \geqslant d_i$ for every $i$.

Apart from technicalities, [Dom 1] asks that *some* feasible payoff for the grand coalition strictly dominate individual disagreement payoffs. [Dom 2] states the obvious point that $d_S$ is always feasible for $S$. In particular, $\zeta_i = \max \Theta(\{i\}) \geqslant d_i$. There is no problem in reading the entire paper with $\zeta_i$ always *equal* to $d_i$, though the distinction between "inaction" (yielding $d_i$) and "outside option" (yielding $\zeta_i$) is discussed further in Section 2.5.

Coalition $S$ is *ineffective* if $\Theta(S) = \{d_S\}$, and *effective* otherwise. Let $\boldsymbol{\Theta}^0$ denote threats when all subcoalitions are ineffective. Each such game is a standard bargaining problem without coalitional threats. Nontrivial threats with $\boldsymbol{\Theta} \neq \boldsymbol{\Theta}^0$ are our point of departure.

2.2. **Coalitional Nash Solution.** A *solution* $\sigma$ assigns to every $G$ a nonempty subset $\sigma(G)$ of $U(G)$ when $U(G) \neq \emptyset$, and the empty set otherwise. A solution could be multi-valued. This is not driven by a desire for generality but by the fact that coalitional threats lead naturally to nonconvex unblocked sets and so to the possibility of multiple outcomes.[9] At the same time, an empty-valued solution does *not* mean that the game itself has no predictable outcome, but this discussion must await our endogenization of threat sets.

2.3. **Axioms.** Here are some standard Nash axioms naturally adapted for $n$ players and possibly multi-valued solutions. We begin with invariance. In Nash's original conception,

[7] For any $T \subset S$ and $x \in \mathbb{R}^S$, $x_T$ is the restriction of $x$ to $\mathbb{R}^T$. For $x$ and $y$ in $\mathbb{R}^m$, $x \geqslant y$ if $x_i \geqslant y_i$ for all $i$, $x > y$ if $x \geqslant y$ but $x \neq y$, and $x \gg y$ if $x_i > y_i$ for all $i$.

[8] However, this is not so when the threats $\Theta(S)$ are endogenous, even without externalities; see Section 5.

[9] Randomization does not help: $F$ would be convex then, but $U(G)$ would, in general, remain nonconvex.

payoffs are generated from outcomes that could include lotteries. Any expected utility representation of preferences must be invariant to affine transformations, which induces an affine transformation of the entire game $G$ in the obvious way.[10] The axiom asserts that:

[**Inv**] If $G'$ is an affine transform of $G$, then $\sigma(G')$ is the same affine transform of $\sigma(G)$.

Next, we suitably modify Nash's independence of irrelevant alternatives to allow for multi-valued solutions. We adapt Kaneko's (1980) version of this axiom in a minimal way to our framework. Our modification allows for changes in the feasible set $F$ but leaves coalitional threats $\mathbf{\Theta}$ unchanged. Of course, the fact that our adaptation is silent on what happens if $\mathbf{\Theta}$ changes makes the axiom weaker rather than stronger.

[**IIA**] If $G = (F, \mathbf{\Theta}, d)$ and $G' = (F', \mathbf{\Theta}, d)$ differ only in that $F' \subseteq F$, then $\sigma(G') = \sigma(G) \cap F'$ whenever this intersection is nonempty.

We impose upperhemicontinuity of the solution in game parameters, which is a mild restriction that's automatically satisfied in the original Nash problem.

[**UHC**] Let $G^k$ be a sequence such that $(F^k, \mathbf{\Theta}^k, d^k)$ converges in the (product) Hausdorff metric to $G = (F, \mathbf{\Theta}, d)$. Then $x^k \in \sigma(G^k)$ for all $k$ and $x^k \to x$ implies $x \in \sigma(G)$.

Kaneko (1980) attributes the modified versions [IIA] and [UHC] to an informal note of Nash; see Shapley and Shubik (1974).

For any two vectors $\lambda$ and $x$ both in $\mathbb{R}^m$, write $\lambda \circ x \equiv (\lambda_1 x_1, \ldots, \lambda_m x_m)$, and say that $\lambda$ is an *expansion* if $\lambda > 1$. It is a *strict expansion* if $\lambda \gg 1$.

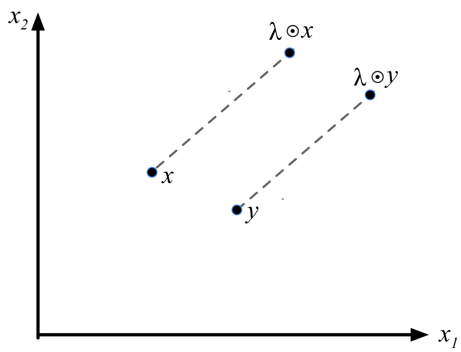
[**PO**] Suppose that $G = (F, \mathbf{\Theta}, d)$ is such that $F = \{x, \lambda \circ x\}$ for some $x \gg d$ and expansion $\lambda$. Then it cannot be that $x \in \sigma(G)$.

This is an innocuous Pareto-optimality requirement placed just on two-point feasible sets.

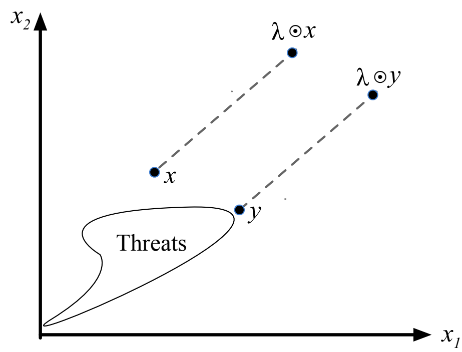
The above axioms are all standard conditions, minimally modified for the coalitional setting. But the next (and last) condition we impose is completely new, and we invoke it to deal with non-trivial coalitional threats.

[**Exp**] Suppose $G = (F, \mathbf{\Theta}, d)$ is such that $F = \{x, y\}$, neither of which is blocked by any threat from $\mathbf{\Theta}$, and suppose that $x \in \sigma(\{x, y\}, \mathbf{\Theta}, d)$. Then there exists an strict expansion $\lambda \gg 1$ such that $\lambda \circ x \in \sigma(\{\lambda \circ x, \lambda \circ y\}, \mathbf{\Theta}, d)$.

[10] $G' = (F', \mathbf{\Theta}', d')$ is an affine transform of $G = (F, \mathbf{\Theta}, d)$ if $G'$ is the result of converting every payoff $x_i$ of every player $i$ in $G$ to $y_i$ by some affine transform $y_i = \alpha_i + \beta_i x_i$, where $\alpha_i \in \mathbb{R}$ and $\beta_i > 0$.




**(a)** No threats




**(b)** With threats

**Figure 1.** The expansion axiom.

To understand [Exp], consult Figure 1, which depicts a two-point feasible set $\{x, y\}$ in each panel, with disagreement payoffs are normalized to zero by invariance. Panel (a) illustrates a pure bargaining problem; every subcoalition is ineffective. Suppose that $x$ is in the solution of this problem. The invariance axiom tells us that *no matter* which expansion vector $\lambda$ we use to scale $x$ and $y$, $\lambda \circ x$ will be in the solution of the new bargaining problem with feasible payoff vectors $\{\lambda \circ x, \lambda \circ y\}$. [Exp] is automatically implied in this case.

Panel (b) illustrates the same two-point feasible set. This time there are nontrivial coalitional threats, and [Exp] is no longer implied by Invariance. However, if $x$ is the *only* outcome to start with, [UHC] guarantees that for some strict expansion $\lambda$, $\lambda \circ x$ will be a solution in the feasible set $\{\lambda \circ x, \lambda \circ y\}$, but with the old threats. [Exp] is again redundant.

[Exp] therefore has extra bite only when $x$ and $y$ are both in the solution to begin with. Let us provide an intuitive motivation for the axiom in this situation. Think of $x$ and $y$ as two different resource vectors for society. Then there is *something* about either alternative that makes it valuable. The allocation $x$ favors one of the players (2, in Figure 1), and the additional payoff to 2 presumably just about compensates for the extra payoff to player 1 under $y$. If there are "diminishing returns," an equiproportionate increase in $x$ and $y$ that places a lot of weight on 1's payoff should retain $x$'s stature as a solution. With "agglomeration" instead, an equiproportionate increase in $x$ and $y$ with high weight on 2's payoff should retain $x$ as a solution. But one way or the other, there should be *some* scaling that continues to (weakly) favor $x$. That's what [Exp] assumes.

All things considered, then, we conclude that the expansion axiom is natural and weak. As we shall see, however, its implications are striking.

2.4. **Characterization.** The assumed restrictions on the bargaining solution lead to:

**Theorem 1.** *A solution $\sigma$ satisfies Axioms* [Inv], [IIA], [UHC], [PO] *and* [Exp] *on the domain $\mathcal{G}$ if and only if it maximizes the (possibly asymmetric) product of the payoffs in excess of the individual disagreement payoffs within the set of unblocked allocations $U(G)$:*

$$\sigma(G) = \arg\max_{x \in U(G)} \prod_{j \in N} [x_j - d_j]^{\gamma_j}, \tag{2}$$

*for some $\gamma \gg 0$. We shall refer to $\sigma$ as the coalitional Nash solution. Unlike disagreement payoffs which are subtracted, all coalitional threats act as conventional constraints in (2).*

This is the non-symmetric version of the Nash bargaining solution with coalitional threats. If all players are presumed to be symmetric, as in the Nash's original two-player setting, the weights $\gamma$ must be taken to be equal. We need to place our symmetry assumption only on bargaining problems in which all coalitions are ineffective:

[**Sym**] Suppose that every coalition is ineffective and $d_i = d_j$ for all $i, j \in N$. If, for every permutation $\pi$ of $N$, $y(\pi) \equiv (y_{\pi(1)}, \ldots, y_{\pi(n)}) \in F$ for any $y \in F$, then $x(\pi) \equiv (x_{\pi(1)}, \ldots, x_{\pi(n)}) \in \sigma(G)$ for any $x \in \sigma(G)$.

**Corollary 1.** *A solution $\sigma^{\text{sym}}$ satisfies all the axioms of Theorem 1 and* [Sym] *if and only if*

$$\sigma^{\text{sym}}(G) = \arg\max_{x \in U(G)} \prod_{j \in N} [x_j - d_j]. \tag{3}$$

*We shall refer to $\sigma^{\text{sym}}$ as the symmetric coalitional Nash solution.*

2.5. **Discussion.** Theorem 1 and Corollary 1 uncover a central point. While individual disagreement payoffs $d$ are subtracted from overall payoffs as as in Nash's theorem, coalitional threats appear as standard constraints *that bind in the conventional way*. A change in $d$ affects the solution $x$ even when $x \gg d$. But a changed coalitional threat $y$ only affects the solution $x$ when it is binding to begin with; i.e., $x_i = y_i$ for some $i$.

Compte and Jehiel (2010) (and before them Chatterjee, Dutta, Ray and Sengupta 1993 for convex characteristic functions) arrive at the same solution as (3), though our solution will depart from theirs when we endogenize threats.[11] Binmore, Shaked and Sutton (1989)

[11]Binmore, Rubinstein and Wolinsky (1986) show how the traditional Nash solution could also arise in a non-cooperative setting, when there is some exogenous probability that the game ends, with individuals holding their outside options. While this is an interpretation that we choose not to pursue here, the approach has potentially interesting implications in a coalitional setting.

experimentally study *individual* outside options, and to our knowledge is the first paper that makes the case for viewing individual threats as conventional constraints:

> "The attraction of split-the-difference lies in the fact that a larger outside option seems to confer greater bargaining power . . . When is such a threat credible? Only when dealing himself out gives the bargainer a bigger payoff than dealing himself in. It follows that the agreement that would be reached without outside options is immune to deal-me-out threats, *unless the deal assigns one of the bargainers less than he can get elsewhere*" [emphasis ours].

We agree, though our Theorem 1 is fundamentally different in that it generates these findings axiomatically, and not in a noncooperative bargaining game with a specific protocol.

The noncooperative setting also permits a distinction to be drawn between disagreement payoffs $d_i$ and the individual outside option $\zeta_i$. The former stems from inaction; for instance, it is the zero payoff that occurs in a binding round with no agreement. In contrast, $\zeta_i$ is obtained by actively exercising the outside option — setting up an individual firm, say, or accepting a job offer. In other settings, there may be no distinction between inaction and an individual outside option. Distinction or not, it is really a detail, and Theorem 1 is proved in exactly the same way. Without the distinction, the individual threat $\zeta(= d)$ must be subtracted from payoffs as a necessary implication of invariance to affine transformations. But there is no parallel latitude for *nonsingleton* threats.

Kaneko (1980), Zhou (1997), Mariotti (1998) and Xu and Yoshihara (2020) extend the Nash bargaining solution to nonconvex feasible sets with $n$ players, where $n \geqslant 2$. They axiomatically characterize solutions of the form (3) when there are no coalitional threats.

Burguet and Caminal (2016) suggest an extension of the Nash solution to TU games, where every coalition $S$ has a scalar "worth" of $v(S)$. Outside options $\zeta_j^S$ for $j \in S$ are subtracted and the rest divided equally, so that the payoff for any $i \in S$ is

$$u_i^S = \zeta_i^S + \frac{1}{|S|}\left[v(S) - \sum_{j \in S} \zeta_i^S\right]. \tag{4}$$

In turn, $\zeta_i^S$ is given by $i$'s expected payoff from joining other coalitions. An endogenous probability distribution governs the formation of such coalitions, with payoffs calculated just as in (4), so the vector $\{u_i^S\}$ is an equilibrium object that runs over all players and coalitions. Coalitional threats are fully subsumed in these individual payoffs and are *assumed* not to be conventional. A related approach that collapses coalitional threats into individual values can be found in Isbell (1960), whose work draws in turn on Harsanyi (1959).

In similar vein, Serrano and Shimomura (1998) study an efficient payoff profile $x$ that precipitates a particular feasible set on any pair $\{ij\}$: all allocations in $F$ that respect the granting of $x_k$ to any $k \neq i, j$, The outside option of $i$ is defined to be the maximum that she can get by joining some coalition $T$ that doesn't include $j$, again paying off $x_k$ to each $k \in T$. Serrano and Shimomura (1998) impose a consistency condition that asks $(x_i, x_j)$ to be the Nash bargaining solution for the induced two-player game, for every pair $\{ij\}$. In TU games, this approach characterizes the pre-kernel of Davis and Maschler (1965).

We have already argued that [Exp] is an intuitive restriction. Solutions such as the one in Kalai and Smorodinsky (1975) adapted to our context satisfy it (see Supplementary Notes). Yet, along with the other axioms it is central to Theorem 1. To illustrate, we construct a solution which satisfies all our axioms barring [Exp] (which establishes *en passant* the independence of [Exp].[12]) This solution $\phi(G)$ for $G \in \mathcal{G}$ satisfies [Par], [Inv], [Sym], [IIA] and [UHC], but differs from the coalitional Nash solution $\sigma^{\text{sym}}(G)$ of Corollary 1.

For every $S \subset N$, let $\tilde{\Theta}(S)$ denote the Pareto frontier of $\Theta(S)$. For each $i \in S$, let

$$m_i(S) \equiv \frac{\min_{x \in \tilde{\Theta}(S)} x_i + \max_{y \in \tilde{\Theta}(S)} y_i}{2}$$

be $i$'s mean over her worst and best payoffs in $\tilde{\Theta}(S)$. The average of these mean payoffs over all subcoalitions containing $i$ is $a_i \equiv [\sum_{S \ni i} m_i(S)]/[2^{n-1} - 1]$, where $2^{n-1} - 1$ is the number of such coalitions). Now, for any game $G$ with $U(G) \neq \emptyset$, define a solution by

$$\phi(G) = \underset{x \in U(G), x \geqslant \beta a + (1-\beta)d}{\arg\max} \prod_{j \in N} [x_j - \beta a_j - (1-\beta)d_j],$$

where $\beta$ is the largest number in $[0, 1]$ such that $x \geqslant \beta a + (1-\beta)d$ for some $x \in U(G)$.[13] Clearly, $\phi$ satisfies [Inv], [IIA], [UHC] and [PO]. Under $\phi$, coalitional constraints are clearly not "conventional." *However*, [Exp] *eliminates this solution*, as we shall now see.

**Example 1.** Consider a three-person game with $d = (0, 0, 0)$, in which all coalitions except $N$ and $\{12\}$ are ineffective.[14] Assume that:

$$\Theta(\{12\}) = \{(6, 0), (0, 6)\} \text{ and } F = \{y, z\},$$

---

[12] The Supplementary Notes show that the other axioms are also independent.

[13] If $a \equiv (a_1, \ldots, a_n)$ is itself in $U(G)$, then $\beta = 1$. Otherwise, $\beta \in [0, 1]$ is a scaling factor, and given that $U(G)$ is nonempty and [Dom1] and [Dom2] hold, a unique $\beta$ satisfying the condition in the text must exist.

[14] We use the notation $\{ij\}$, $\{ijk\}$ etc. to denote specific coalitions, instead of $\{i, j\}$, $\{i, j, k\}$ etc.

where $y = (3, 3, 1)$ and $z = (2, 2, 4)$. Observe that $a = (1, 1, 0)$ and $\beta = 1$. So, because $y$ and $z$ are both unblocked,

$$\phi(G) = \arg\max_{x\in\{y,z\}}(x_1 - 1)(x_2 - 1)x_3 = \{y, z\}, \tag{5}$$

where the second equality follows from the fact that $(y_1 - 1)(y_2 - 1)y_3 = 4 = (z_1 - 1)(z_2 - 1)z_3$. We claim that $\phi$ does not satisfy [Exp]. To this end, consider any expansion $\lambda = (\lambda_1, \lambda_2, \lambda_3) \gg (1, 1, 1)$. We shall show that for the game $G' = (\{\lambda \odot y, \lambda \odot z\}, \Theta, 0)$,

$$\phi(G') = \arg\max_{x\in\{\lambda\odot y,\lambda\odot z\}} (x_1 - 1)(x_2 - 1)x_3 = \{\lambda \odot z\}, \tag{6}$$

which, given $y \in \phi(G)$ by (5), violates [Exp]. To show (6), let $\delta_i \equiv \lambda_i - 1$ for each $i$. Then:

$$\lambda \odot y = (3[1 + \delta_1], 3[1 + \delta_2], 1 + \delta_3) \text{ and } \lambda \odot z = (2[1 + \delta_1], 2[1 + \delta_2], 4[1 + \delta_3]).$$

Let the maximand in $\phi$ be denoted by $f(x) = (x_1 - 1)(x_2 - 1)x_3$. Then we have:

$$f(\lambda \odot y) = (2 + 3\delta_1)(2 + 3\delta_2)(1 + \delta_3) = (4 + 6\delta_1 + 6\delta_2 + 9\delta_1\delta_2)(1 + \delta_3),$$

and

$$f(\lambda \odot z) = (1 + 2\delta_1)(1 + 2\delta_2)[4(1 + \delta_3)] = (4 + 8\delta_1 + 8\delta_2 + 16\delta_1\delta_2)(1 + \delta_3).$$

Clearly, $f(\lambda \odot z) > f(\lambda \odot y)$, which implies (6). ◇

## 3. Embedding $G$ in a Coalitional Game

To many, the object $G = (F, \boldsymbol{\Theta}, d)$ is a relic from an era that has long fallen out of favor in modern research because it ignores externalities. In this section, we show how $G$ can be embedded within an ambient environment in which externalities and non-cooperative interaction are present. To do so, we endogenize both the coalition structure and the payoffs to all coalitions in that structure. The coalitional Nash solution identified in Theorem 1 will play a central (and dual) role in both these tasks: one, as a solution for a coalition, and two, as a threat for its "parent" coalitions attempting to come to solutions of their own.

3.1. **Nash-in-Nash.** A *coalition structure* is a partition $\pi$ of the player set $N$ into coalitions. For each coalition $S$, let $A(S)$ be its set of feasible (joint) actions; these could be coalition-dependent. For instance, transfers of goods or money or certain within-group activities may become available once a coalition forms. Define $A(\pi) \equiv \bigtimes_{S\in\pi} A(S)$, and $A \equiv \cup_\pi A(\pi)$. Each player $i$ has a payoff function $f_i(a)$ on $a \in A$. Standard games are a special case.[15] An "inaction action" is always available to every player $i$, with payoff $d_i$ normalized to zero. Finally, a set of bargaining weights $\gamma(S)$ is available to each coalition $S$. Their coalitional bargaining solution will employ these weights.

A coalition interacts non-cooperatively with other coalitions in $\pi$ but attempts to find a cooperative agreement within itself. As it does so, it contends with "internal threats" from potential *breakaways*: subcoalitions $T$ of coalitions $S$ in $\pi$. Suppose that every $T \subset S \in \pi$ has a set of threat payoffs $\Theta(T, \pi)$, soon to be endogenized. Let $\boldsymbol{\Theta}(\pi) \equiv \{\Theta(T, \pi)\}_{T \subset S \in \pi}$. Drawing on the characterization in Theorem 1 and recalling that $d = 0$, say that $a$ is a *Nash-in-Nash equilibrium* for $\pi$ relative to $\boldsymbol{\Theta}(\pi)$ if for every $S \in \pi$,

$$(7) \quad a_S \in \underset{a'_S \in A(S)}{\arg\max} \prod_{i \in S} f_i(a'_S, a_{-S})^{\gamma_i(S)}, \text{ subject to } f_S(a'_S, a_{-S}) \text{ unblocked by any } T \subset S;$$

that is, $f_S(a)$ is a coalitional Nash solution relative to $\boldsymbol{\Theta}(\pi)$, given the weights $\gamma(S) \gg 0$. Structure $\pi$ is *viable* relative to $\boldsymbol{\Theta}(\pi)$ if it admits a Nash-in-Nash equilibrium relative to $\boldsymbol{\Theta}(\pi)$. This definition begins to connect non-cooperative interaction across coalitions and cooperative play within them, the latter in the spirit of Theorem 1. The connection will be made complete as we endogenize $\boldsymbol{\Theta}$. For more remarks, see the Supplementary Notes.

Viability isn't guaranteed. But the possible lack of it is a feature, not a bug. A central implication that we will develop is that not all structures are viable, even the efficient ones.

Cooperation within and competition across coalitions is a hallmark of the literature on applications of Nash-in-Nash (Horn and Wolinsky 1988, Ho and Lee 2017, 2019, Collard-Wexler, Gowrisankaran and Lee 2019, or Bagwell, Staiger and Yurukoglu 2020). But this literature focuses exclusively on bilateral bargaining with *individual* outside options.[16] That isn't surprising, given the absence of a Nash bargaining theory with coalitional threats.[17] Our model does allow for them. We now turn to their endogenous determination.

3.2. **Viable Structures.** First, we initialize the recursion: a Nash-in-Nash equilibrium for the singleton structure $\pi^s$ is just a Nash equilibrium; assume it exists and declare $\pi^s$ to be viable. Consider any structure $\pi^{ij} \equiv \{\{ij\}, \{k\}, \{\ell\}, \ldots\}$, with one doubleton $\{ij\}$ and

[15] In standard games, action sets $A_i$ are defined for each individual $i$ in $S$, whereupon $A(S) = \bigtimes_{i \in S} A_i$.

[16] Outside options are sometimes netted out; see, e.g., Lee and Fong (2013). Sometimes, they are treated as conventional constraints, as in Ho and Lee (2019) or Collard-Wexler, Gowrisankaran and Lee (2019).

[17] However, this literature permits an individual to enter into several (bilateral) agreements, a feature that our use of coalition structures does not accommodate. While $G$ could be easily embedded into a network setting, with multilateral bargaining permitted over all connected components, we leave this task for future research.

$n-2$ singletons. It has two potential breakaways $\{i\}$ and $\{j\}$. Each induces $\pi^s$. Define:

$$\Theta^*(\{k\},\pi^{ij}) \equiv \{f_i(a)|\ a \text{ is Nash-in-Nash for } \pi^s\}$$

and $\boldsymbol{\Theta}^*(\pi^{ij}) \equiv \{\Theta^*(k,\pi^{ij})\}_{k=i,j}$.[18] Say that $\pi^{ij}$ is *viable* if it is viable relative to $\boldsymbol{\Theta}^*(\pi^{ij})$.

Next, the recursive step: for any $m \in \{1,\ldots,n-2\}$ and any structure $\pi'$ of cardinality $m+1$ or greater, suppose that we know if $\pi'$ is viable or not, and also all the threats to $\pi'$: $\boldsymbol{\Theta}^*(\pi') = \{\Theta^*(T',\pi')\}_{T'\subset T\in\pi'}$. Pick any $\pi$ of cardinality $m$. For any breakaway $T\subset S\in\pi$, let $\pi^T \equiv \pi_{-S}\cup\{T,S-T\}$. If $\pi^T$ is viable relative to $\boldsymbol{\Theta}^*(\pi^T)$, let:

$$\Theta^*(T,\pi) \equiv \{f_T(a)|a \text{ is Nash-in-Nash for } \pi^T \text{ relative to } \boldsymbol{\Theta}^*(\pi^T)\}. \tag{8}$$

If $\pi^T$ is not viable, define $\mathcal{P}^T$ to be the collection of all the *coarsest* viable coalition structures $\pi'$ that are *refinements* of $\pi^T$.[19] Then define

$$\Theta^*(T,\pi) \equiv \{f_T(a)|a \text{ is Nash-in-Nash for some } \pi'\in\mathcal{P}^T \text{ relative to } \boldsymbol{\Theta}^*(\pi')\} \tag{9}$$

and collect these threats: $\boldsymbol{\Theta}^*(\pi) \equiv \{\Theta^*(T,\pi)\}_{T\subset S\in\pi'}$. Say that $\pi$ is *viable* if it is viable relative to $\boldsymbol{\Theta}^*(\pi)$, as in equation (7).

This recursion is intuitive, especially when $\pi^T$ is viable. When $\pi^T$ is not viable, our construction uses the (correct) prediction by $T$ that the structure will crumble further. Just where it will crumble *to* is unclear. Our construction asserts that it will stop at one of the coarsest structures that refine $\pi^T$, presumably taken there by other breakaways, including subcoalitions of $T$ itself. In all cases, the members of $T$ forecast where they end up, so that a threat payoff is well defined. These threats are then collected to define $\Theta^*(T,\pi)$.[20]

Once $\boldsymbol{\Theta}^*(\pi) \equiv \{\Theta^*(T,\pi)\}$ is defined for every $\pi$, we focus on the *coarsest* viable coalition structures and the payoffs they generate — these are the Nash-in-Nash solutions for the overall game (we drop the "relative to" clause as the threats are now recursively specified). Our focus reflects our implicit viewpoint that an attempt to write agreements starts with the grand coalition, and then breaks down to the least degree possible. This idea is easily extended to cases in which some $\bar{\pi}$ is the parent structure, "above" which no coalition can form. Then our structures of interest would be the coarsest viable substructures of $\bar{\pi}$.

[18]Endogenous threats, obtained recursively, will be identified with an asterisk as in $\boldsymbol{\Theta}^*$. Our definition presumes that $k$ can "choose" their threats if there are multiple equilibria. Other alternatives are possible but are not quite central to the analysis or the applications, so we will stay with this approach.

[19]Obviously, $\mathcal{P}^T \neq \emptyset$ because $\pi^s$ is a viable structure that refines $\pi^T$.

[20]Our definition asks that the breakaway members have common beliefs about what will happen following a deviation. These may rely on their assessment of the beliefs that future breakaways these may not be aligned with the beliefs that $T$ expects other breakaway coalitions to have, were $\pi^T$ to disintegrate further.

Our definition does not allow for breakaways to combine with other players who are not in their parent coalition. The Supplementary Notes contain a discussion of the issues that could arise in this context. Note that the grand coalition is at the apex of our recursion, so any conceivable coalition structure can potentially form with or without our restriction.

The viability of a structure could fail for two reasons. One is that breakaway threats overpower what some coalitions in the structure can achieve. The other is a technical failure of existence, such as openness or the lack of continuity. We are not really interested in these latter failures. The real test of viability comes from the threats posed by breakaways.

3.3. **Transfers and Viability.** In many settings, a coalitional allocation is achieved by first maximizing aggregate coalitional payoff in the strategic game and then distributing that payoff among members *with no additional externalities*. TU games are a leading instance. There is a distinguished good $m$, freely transferable across coalition members, with individual payoffs quasi-linear in $m$. It is convenient not to include these actions in the set $A_S$. Then, for every $a \in A$ and every person $i$ who receives $m_i$ (positive or negative):

$$\text{Payoff to } i \ = f_i(a) + m_i.$$

A *coalitional Nash equilibrium under* $\pi$ is an action profile $a^* = \{a^*_S\}_{S\in\pi}$, with $a^*_S \in A_S$ for every $S \in \pi$, such that for each $S$,

$$\sum_{i\in S} f_i(a^*) \geqslant \sum_{i\in S} f_i(a_S, a^*_{-S}) \text{ for all } a_S \in A_S.$$ [21]

A coalitional Nash equilibrium is a precursor to Nash-in-Nash equilibrium. If it is unique for all $\pi$, it generates a scalar value $v(S,\pi)$ to every coalition $S$ in every structure $\pi$ (Lucas 1963, Thrall and Lucas 1963, Myerson 1977, Ray and Vohra 1999), which is then divided in accordance with the coalitional bargaining solution. This nice separability property holds in many applications, as we shall see in the next Section.

## 4. Applications

The following applications are designed to highlight the issues that arise when studying viable coalition structures. Each can be generalized to accommodate larger player sets, but

[21] Coalitional Nash equilibria can also be defined for general games which are unconstrained by any norms for payoff division. Every $S \in \pi$ chooses a *Pareto-optimal response*: there is no $a'_S \in A_S$ such that $f_S(a'_S, a^*_{-S}) \gg f_S(a)$. See Ichiishi (1981), Ray and Vohra (1997) and Haeringer (2004)).

the main emphasis here is on the direct implications of our recursive definition. We place particular focus on the Coasean question: is there some viable *and* efficient outcome?

The inefficiencies illustrated in some of these examples are a direct result of externalities, as will become additionally clear from Section 5.3. But it is important to appreciate that the standard intuition for inefficiency in the face of externalities does not apply. Binding agreements can always be written. For instance, consider any two-person strategic setting, such as the Prisoner's Dilemma. Then the grand coalition is indeed viable, and its Nash bargain is efficient. However, efficiency is not inevitable with three or more players.

Example 2, a three-country "environmental problem," shows how a single country breaks away from environmental cooperation when it (correctly) predicts that the others will be willing to remain together. The grand coalition is unable to prevent this breakaway.

Example 3, a three-firm "R&D race," shows how the departure of a single firm makes the remaining firms stay together (as in Example 2) but (*unlike* Example 2) they overpower the outsider. The threats here come from larger subsets of firms. To forestall them, the grand consortium divides its payoff unequally, despite its ex ante symmetry, but remains viable.

Example 4, "a Cournot cartel," has yet other features. With three symmetric firms, the departure of a single firm causes the remaining structure to disintegrate, unlike in Examples 2 in 3, but this nullifies the original threat posed by the departing firm. In contrast, with more firms or in non-symmetric settings, a single firm can pose a bigger threat because the remaining firms stick together. and only inefficient intermediate structures might be viable.

As we shall see from these examples, a plethora of outcomes is possible even in relatively simple three-agent situations. But that does not mean that the theory is devoid of power: any *particular* setting can be fully solved for, and there is a clear prediction regarding the viability of every coalitional structure in that setting.

**Example 2.** [An Environmental Problem] We build on Ray and Vohra (2001). Three countries engage in pollution control $(z_1, z_2, z_3)$, with aggregate benefit $z_1 + z_2 + z_3$ accruing equally to each country. The cost of $z_i$, borne by country $i$, is $(1/3a_i^2)z_i^3$, where $a_i^2$ is a proxy for technical knowhow or population size. The payoff to country $i$ is given by

$$(z_1 + z_2 + z_3) - \frac{z_i^3}{3a_i^2}. \tag{10}$$

Assume that a coalition once formed can make unlimited transfers among its members. Of course, if the grand coalition forms it can internalize the externalities and achieve Pareto efficiency. But there are six threats to any such outcome, one each coming from one of three deviating countries that could stand in their own, and three more coming from doubleton deviants. Indeed, we will see that: (a) these coalitional threats could make the grand coalition non-viable even when it is efficient; that this happens because (b) intermediate coalition structures of the form $\{jk, \ell\}$ are viable, even though (c) threats from individual players $j$ or $k$ or unequal bargaining weights may force an unequal division of the surplus in the doubleton coalition $jk$, so as to maintain viability of the intermediate structure.

The simplicity of this example lies in the fact that externalities are additive, so that production decisions are uniquely pinned down for any coalition of countries. For instance, under the singleton coalition structure, each country will produce $z_i = a_i$. Under this unique coalitional Nash equilibrium for the singleton structure $\pi^s$, the payoff to country $i$ is:

$$v(\{i\}, \pi^s) = (a_1 + a_2 + a_3) - \frac{a_i}{3}.$$

So any $i \in jk$ in any coalition structure of the form $\pi^{jk} = \{jk, \ell\}$ poses the threat

$$\Theta^*(\{i\}, \pi^{jk}) = \{v(\{i\}, \pi^s)\} = \left\{(a_1 + a_2 + a_3) - \frac{a_i}{3}\right\}. \tag{11}$$

to that structure. With that in mind, consider Nash-in-Nash for a "doubleton" structure, say $\pi^{23} = \{1, 23\}$. Country 1 will continue to choose $z_1 = a_1$. Since the coalition $\{23\}$ can make unlimited transfers among its members, it will seek to maximize $2(z_1 + z_2 + z_3) - (1/3a_2^2)z_2^3 - (1/3a_3^2)z_3^3$, taking $z_1$ as given. That leads to new outputs of $\sqrt{2}a_2$ and $\sqrt{2}a_3$ for countries 2 and 3 in the coalition $\{23\}$, with resulting coalitional payoff given by:

$$\begin{aligned} v(\{23\}, \pi^{23}) &= 2(a_1 + \sqrt{2}a_2 + \sqrt{2}a_3) - \frac{1}{3a_2^2}\sqrt{8}a_2^3 - \frac{1}{3a_3^2}\sqrt{8}a_3^3 \\ &= 2a_1 + \frac{\sqrt{32}}{3}a_2 + \frac{\sqrt{32}}{3}a_3. \end{aligned}$$

Is this enough aggregate surplus for $\{23\}$ to make the coalition structure $\pi^{23}$ viable? In fact it is, because the sum of the threats posed by countries 2 and 3 can be expressed as:

$$\Theta^*(\{2\}, \pi^{23}) + \Theta^*(\{3\}, \pi^{23}) = (2a_1 + 2a_2 + 2a_3) - \frac{a_2}{3} - \frac{a_3}{3} = 2a_1 + \frac{5}{3}a_2 + \frac{5}{3}a_3,$$

which is smaller than $v(\{23\}, \pi^{23})$. Coalition $\{23\}$ can therefore divide its surplus to guard against a threat from either of its members, and so $\{1, 23\}$ is viable. Their coalitional Nash solution may not result in equal division, however, even if $\gamma_2^{23} = \gamma_3^{23}$. For instance, if $a_3$

is higher than $a_2$, country 2 poses the bigger threat, and if the gap is large enough, country 2's threat could be a binding constraint, leading to unequal division within $\{23\}$.[22]

Now we turn to the grand coalition $\pi^g$ where all countries attempt to cooperate. Given the availability of unlimited transfers, this coalition will maximize its aggregate payoff, given by $3(z_1 + z_2 + z_3) - (1/3a_1^2)z_1^3 - (1/3a_2^2)z_3^3 - (1/3a_3^2)z_3^3$. It follows that country $i$ will produce $\sqrt{3}a_i$ at cost $\sqrt{3}a_i$, so that:

$$\begin{aligned} v(\{123\}, \pi^g) &= 3(\sqrt{3}a_1 + \sqrt{3}a_2 + \sqrt{3}a_3) - \sqrt{3}a_1 - \sqrt{3}a_2 - \sqrt{3}a_3 \\ &= 2\sqrt{3}(a_1 + a_2 + a_3). \end{aligned}$$

Under the structure $\pi^{23}$, country 1 enjoys the blessings of pollution control carried out by countries 2 and 3, and so, under $\pi^g$ and given the viability of $\{1, 23\}$, poses the threat

$$\Theta^*(\{1\}, \pi^g) = a_1 + \sqrt{2}a_2 + \sqrt{2}a_3 - \frac{1}{3}a_1 = \frac{2}{3}a_1 + \sqrt{2}a_2 + \sqrt{2}a_3 \tag{12}$$

to the viability of the grand coalition. Because the game is naturally superadditive, and transfers are available, the grand coalition can protect itself from this particular threat: it can retain both coalitions $\{1\}$ as well as $\{23\}$ by suitably (though unequally) dividing its aggregate worth. But alas, this is not the only threat that the grand coalition has to deal with. *Any* country $i$ can precipitate the (viable) intermediate coalition structure $\pi^{jk}$ to earn an amount analogous to the payoff in (12). The sum of these individual threats is:

$$\sum_{i=1}^{3} \Theta^*(\{i\}, \pi^g) = (\frac{2}{3} + 2\sqrt{2})(a_1 + a_2 + a_3),$$

which is greater than the worth of the grand coalition.[23] The grand coalition therefore cannot simultaneously guard against all such threats. So the coarsest viable structures are $\pi^{12}$, $\pi^{13}$ and $\pi^{23}$, which are all Pareto inefficient. In any such structure, payoffs are given by a Nash-in-Nash equilibrium, with possibly unequal division within the doubletons. These predictions are remarkably robust to inequalities in $\{a_i\}$ as also in bargaining weights.[24] $\diamond$

---

[22]Country 2 can precipitate the singletons and earn $a_1 + a_2 + a_3 - \frac{a_2}{3} = a_1 + \frac{2}{3}a_2 + a_3$. Equal division of the surplus within $\{23\}$ results in $a_1 + \frac{\sqrt{8}}{3}a_2 + \frac{\sqrt{8}}{3}a_3$. The latter is lower than the threat if $a_3$ is large enough. Note that the the country with the better technology (or larger population) incurs the greater cost.

[23]This follows from the fact that $2/3 + 2\sqrt{2} = 3.495$ while $2\sqrt{3} = 3.464$.

[24]They are also robust to costly within-coalition transfers. If the marginal cost of transfers is nondecreasing, the sum of payoffs is maximized at one point on the Pareto frontier, and the product of payoffs at another. For instance, when $\{23\}$ maximizes its Nash product, the sum of payoffs will fall short of $v(\{23\}, \pi^{23})$ when $a_2 \neq a_3$. However, as long $a_2 + a_3$ is not too far below $\sqrt{2}a_2 + \sqrt{2}a_3$, the intermediate structure $\{1, 23\}$ will remain viable. In this case, the conclusion that the intermediate coalition structures are the coarsest viable

**Example 3.** [An R&D Race] We extend Loury (1979) and Reinganum (1982) to incorporate coalition formation in R&D races. To emphasize a contrast with Example 2, we first consider three *symmetric* firms with equal bargaining weights, engaged in an R&D race. A formed coalition $S \subseteq \{123\}$ works towards a Poisson prize with parameter $\lambda_S$ given by

$$\lambda_S = \sqrt{2e_S}, \tag{13}$$

where $e_S$ is the overall research effort undertaken by $S$. This effort is derived from synergistic research budgets $\{x_j\}$ of firms in $S$, as follows:

$$e_S = [\sum_{j \in S} x_j^{1/\alpha}]^{\alpha}, \tag{14}$$

where $\alpha \geqslant 1$ is a measure of synergy. (There is no coordination and so no synergy across firms in different coalitions, and the Poisson arrival rates across such firms are additive.)

The prize has value 1, and so has present value $\exp(-r\tau)$ to whoever wins it at random arrival time $\tau$, where $r$ is the common discount rate. But coalition $S$ obtains the prize only if its Poisson arrival happens first. Define $\lambda_{-S}$ to be the aggregate arrival rate over coalitions other than $S$ (i.e., $\lambda_{-S} \equiv \sum_{W \in \pi - S} \lambda_W$). Then the instantaneous probability of first arrival at date $\tau$ in favor of $S$ is $\lambda_S \exp(-\lambda_S \tau) \exp(-\lambda_{-S}\tau) d\tau$, so that the expected value to coalition $S$ is $\int \exp(-r\tau)\lambda_S \exp(-\lambda_S \tau) \exp(-\lambda_{-S}\tau) d\tau = \lambda_S/(r + \lambda_S + \lambda_{-S})$.

If transfers are freely available within a coalition, then by symmetry and the convexity of costs, it is optimal to divide the overall effort equally among firms in $S$. Let $s$ be the cardinality of $S$. From (13) and (14), the budget needed to generate an arrival rate of $\lambda_S$ must be given by $(1/2s^{\alpha})\lambda_S^2$, and so the expected aggregate payoff to $S$ is given by:

$$\Sigma_S \equiv \frac{\lambda_S}{r + \lambda_S + \lambda_{-S}} - \frac{1}{2s^{\alpha}}\lambda_S^2. \tag{15}$$

This is a strictly concave problem in $\lambda_S$ with an interior solution, so first-order equality conditions are necessary and sufficient for characterizing best responses:

$$\frac{1}{r + \Lambda} - \frac{\lambda_S}{(r + \Lambda)^2} - \frac{1}{s^{\alpha}}\lambda_S = 0. \tag{16}$$

where $\Lambda \equiv \lambda_S + \lambda_{-S}$. The grand coalition of all firms has no rivals, and so we can set $\lambda_{-S} \equiv 0$ in (15). The resulting problem is a simple optimization problem, and (after some

ones remains valid. (The fact the grand coalition's feasible payoffs are also restricted doesn't affect the rest of the argument because it only makes the grand coalition less powerful.)

manipulation and noting that $s = 3$) the first order condition in (16) reduces to:

$$\frac{1}{3^\alpha}\lambda_N(r+\lambda_N)^2 = r, \tag{17}$$

which uniquely pins down $\lambda_N$. Using (15) and (17), the payoff of $N$ is then given by:

$$\Sigma(N) \equiv \frac{\lambda_N}{r+\lambda_N} - \frac{1}{2\cdot 3^\alpha}\lambda_N^2, \text{ where } \lambda_N \text{ uniquely solves (17).} \tag{18}$$

Now consider a structure $\pi^{jk} = \{jk, \ell\}$, where $jk$ is a coalition and $\ell$ a standalone firm. With $jk$ and $\ell$ maximizing payoff as in (15), the first-order conditions reduce to:

$$r+\lambda_\ell = \frac{1}{2^\alpha}\lambda_{jk}(r+\lambda_{jk}+\lambda_\ell)^2 \text{ (for } jk\text{), and } r+\lambda_{jk} = \lambda_\ell(r+\lambda_{jk}+\lambda_\ell)^2 \text{ (for } \{\ell\}\text{).} \tag{19}$$

The aggregate payoff to coalition $jk$ can therefore be written as

$$\Sigma(jk, \pi^{jk}) = \frac{\lambda_{jk}}{r+\lambda_{jk}+\lambda_\ell} - \frac{1}{2\cdot 2^\alpha}\lambda_{jk}^2. \tag{20}$$

The heart of this example lies in the following observation: for any $r > 0$, there is $\bar{\alpha} > 1$ such that if $\alpha > \bar{\alpha}$, then the payoff $\Sigma(jk, \pi^{jk})$ exceeds 2/3. Indeed, as synergy levels become large ($\alpha \to \infty$), the payoff $\Sigma(jk, \pi^{jk})$ converges to 1, which is the maximum possible surplus. If the degree of impatience is small ($r \simeq 0$), it is easy to get a sense of the threshold $\bar{\alpha}$ via some direct calculations. When $r \simeq 0$,

$$\Sigma(jk, \pi^{jk}) \simeq \frac{2^{\alpha/2}}{2^{\alpha/2}+1} - \frac{2^{\alpha/2}}{2(2^{\alpha/2}+1)^2} > 2/3 \text{ for all } \alpha \geqslant 3.3.$$

The details of these arguments are in the Supplementary Notes, but intuitively, high synergies enable the doubleton coalition to easily overpower its single opponent. Contrast this with the doubleton structure in Example 2, where the singleton free-rider gains relative to its doubleton opponent, the latter providing the bulk of pollution control.

We combine this observation with a short description of equilibrium under the singleton structure, in which firm $i$ (or more pedantically, the coalition $\{i\}$) replaces $jk$ in (15) and (16). This can easily be shown to yield a unique equilibrium that's symmetric (see Supplementary Notes). The exact details are unimportant, except to note that:

$$\Theta^*(i, \pi^{jk}) = (b, b, b), \text{ where } b \in (0, 1/3). \tag{21}$$

So for large enough $\alpha$, the threat of a singleton breakaway from $\pi^{jk}$ is empty, because the net payoff to the doubleton exceeds $2/3$ and (by symmetry) will be divided equally. So the main threat to the grand R&D consortium comes from a potential doubleton breakaway:

$$\Theta^*(jk, \{N\}) = (y, y), \text{ where } y > 1/3 \text{ for } \alpha \text{ large enough.} \tag{22}$$

Indeed, there are *three* such outcomes as we move over all $\{jk\}$ that can break away. The grand coalition has no more than a surplus of 1 to play with; indeed, for finite synergy and $r > 0$, it will have strictly less. Moreover, for large $\alpha$, $y \in (1/3, 1/2)$ for all $\{jk\}$, so these are serious threats. With equal bargaining weights, the grand coalition attempts to maximize the Nash product $x_1 x_2 x_3$, subject to the constraint that $\max\{x_j, x_k\} \geqslant y$ for every $j$ and $k$. Because $y > 1/3$ for large $\alpha$, the solution cannot be symmetric, *even if* all the firms are symmetric. The argument extends to any number of firms. The main threats will all arise from (minimal) majority coalitions. But the grand coalition survives.

Indeed, the viability of the grand coalition also survives asymmetric bargaining weights. The intermediate structures are still viable as they were before, because the unblocked set for the doubletons continues to be nonempty. However, we must now amend (22) to

(23) $$\Theta^*(jk, \{N\}) = (y_j^{jk}, y_k^{jk}) \geqslant (b, b), \text{ with } y_j^{jk} + y_k^{jk} > 2/3 \text{ for } \alpha \text{ large enough,}$$

where $y^{jk}$ is the (unique) Nash-in-Nash payoff for $\{jk\}$ within $\pi^{jk}$ and $b$ is given by (21). The Supplementary Notes show that an unblocked allocation can still be found for the grand coalition. Unequal treatment of member firms continues to hold, at least for all "consistent" bargaining weights that maintain their relative values across all coalitions.[25]

The example highlights how the viability of the grand coalition may be predicated on treating its symmetric members asymmetrically. Note the difference from Example 2, where the most powerful threats come from singleton deviants (or more generally, minority coalitions), and cannot be appeased by the grand coalition. $\diamond$

**Example 4.** [Cartels in Oligopoly] $n$ firms sell a homogeneous product. They can form coalitions. Coalitions in a structure interact non-cooperatively in the sense of Cournot. Within a coalition, transfers are used freely to implement coalitional Nash solutions. We study viable coalition structures; in particular, whether or not full cartelization occurs.

There is a linear inverse demand curve, with units normalized so that $p = A - x$, where $p$ is price, $A > 0$ and $x$ is aggregate output. Firm $i$ has constant unit cost $c_i$. Define $C \equiv \sum_j c_j$. Under the baseline singleton structure, it is easy to show that each firm $i$ produces

(24) $$x_i = \frac{A + C}{n + 1} - c_i,$$

[25]If bargaining weights are entirely arbitrary across coalitions, then it is possible to have equal division in the grand coalition, though we do not think of this as a realistic possibility.

provided the following necessary and sufficient condition for interior output holds:

$$A + C > (n + 1)[\max_j c_j]. \tag{25}$$

Using (24) to calculate aggregate quantity and then price $p$, firm $i$'s equilibrium profit is:

$$(p - c_i)x_i = \left(\frac{A + C}{n + 1} - c_i\right)^2. \tag{26}$$

Now consider any coalition structure $\pi$. Given linear costs and costless transfers, all production within a coalition will be carried by its lowest-cost members — the other members are effectively paid to stay out of the way. The Supplementary Notes show that condition (25) also guarantees positive production by every coalition in coalitional equilibrium in any coalitional structure. That equilibrium is governed by the same equations that we had for the singletons, and in particular for every $S \in \pi$, its aggregate profit is

$$v(S, \pi) = \left[\frac{A + C(\pi)}{|\pi| + 1} - c(S)\right]^2, \tag{27}$$

where $c(S) \equiv \min_{i \in S} c_i$, $C(\pi) \equiv \sum_{T \in \pi} c(T)$ and $|\pi|$ is the cardinality of $\pi$. We can now use this formula to deduce viable structures in any given setting. To illustrate this, consider the case of just three firms. Then the singleton structure $\pi^s$ will yield a profit of

$$v(i, \pi^s) \equiv \left[\frac{A + C}{4} - c_i\right]^2 = \frac{(A - c)^2}{16}, \tag{28}$$

the last equality holding in the special case when all firms are symmetric with $c_1 = c_2 = c_3 = c$. Likewise, in an intermediate structure $\pi^{jk}$ of the form $\{jk, \ell\}$, we see that

$$v(\ell, \pi^{jk}) \equiv \left[\frac{A + \min\{c_j, c_k\} - 2c_\ell}{3}\right]^2 = \frac{(A - c)^2}{9} \text{ and}$$

$$v(jk, \pi^{jk}) \equiv \left[\frac{A + c_\ell - 2\min\{c_j, c_k\}}{3}\right]^2 = \frac{(A - c)^2}{9}, \tag{29}$$

while the aggregate payoff to the grand coalition is

$$v(N, \pi^g) \equiv \left[\frac{A - \min\{c_1, c_2, c_3\}}{2}\right]^2 = \frac{(A - c)^2}{4}, \tag{30}$$

where again, the last equalities are for equal-cost firms. In that special case, note that the intermediate coalition structure *cannot* be viable, because (28) and (29) jointly imply that

$$v(jk, \pi^{jk}) = \frac{(A - c)^2}{9} < 2\frac{(A - c)^2}{16} = v(j, \pi^s) + v(k, \pi^s),$$

from which it follows that any breakaway from the grand coalition must precipitate the singleton structure. But now the grand coalition is indeed viable, for by (28) and (30),

$$v(N,\pi^g) = \frac{(A-c)^2}{4} > 3\frac{(A-c)^2}{16} = v(1,\pi^s) + v(2,\pi^s) + v(3,\pi^s).$$

This non-viability of intermediate structures and the resulting viability of the grand coalition is robust to coalitional bargaining weights, as long as costs are identical. But intermediate structures may become viable with heterogeneous costs. Suppose that

$$A = 10, c_1 = 2, c_2 = 3, \text{ and } c_3 = 4.$$

Then (28) implies that $v(1,\pi^s) \simeq 7.56, v(2,\pi^s) \simeq 3.06,$ and $v(3,\pi^s) \simeq 0.56$. Likewise, (29) tells us that

$$[v(1,\pi^{23}) = 9 \text{ and } v(23,\pi^{23}) = 4], [v(2,\pi^{13}) = 4 \text{ and } v(13,\pi^{13}) = 9], [v(3,\pi^{12}) \simeq 1.78 \text{ and } v(12,\pi^{12}) \simeq 11.11],$$

and (30) tells us that $v(N,\pi^g) = 16$. These computations show that in contrast to the symmetric case, *every* intermediate structure may now be viable. For instance, $v(23,\pi^{23}) = 4 > 3.06 + 0.56 = v(2,\pi^s) + v(3,\pi^s)$, so $(23)$ will not split up under the structure $\pi^{23} = \{1, \{23\}\}$. The precise division of the surplus within $\{23\}$ will, of course, depend on the Nash bargaining weights, but firm 2 will get at least 3.06 and firm 3 at least 0.56. In general, the threat posed to the grand coalition from this particular structure can be written as

$$\Theta^*(1,\pi^g) = \{9\}, \Theta^*(23,\pi^g) = \{(x_2^{23}, x_3^{23})\}, \text{ where } x_2^{23} \geqslant 3.06, x_3^{23} \geqslant 0.56 \text{ and } x_2^{23} + x_3^{23} = 4.$$

Similar computations for the other intermediate structures show that

$$\Theta^*(2,\pi^g) = \{4\}, \Theta^*(13,\pi^g) = \{(x_1^{13}, x_3^{13})\}, \text{ where } x_1^{13} \geqslant 7.56, x_3^{13} \geqslant 0.56 \text{ and } x_1^{13} + x_3^{13} = 9$$

and

$$\Theta^*(3,\pi^g) = \{1.78\}, \Theta^*(12,\pi^g) = \{(x_1^{12}, x_2^{12})\}, \text{ where } x_1^{12} \geqslant 7.56, x_2^{12} \geqslant 3.06 \text{ and } x_1^{12} + x_2^{12} = 11.11.$$

As it so happens, the grand coalition can withstand all these threats, regardless of the bargaining weights. The unblocked set for $N$ consists of any split $x$ of the monopoly profits with $x_1 \geqslant 9$, $x_2 \geqslant 4$ and $x_3 \geqslant 1.78$. So the unique viable structure is one in which the grand coalition forms, and deploys its coalitional Nash solution, the exact division depending on barganing weights. For instance, with equal weights, the unique solution is:

$$\sigma^{\text{sym}}(N) = (9, 4, 3)$$

Had the intermediate structures not been viable (as in the equal cost case), firms 1 and 2 would have received only their Cournot payoffs. It is due to the viability of intermediate coalition structures that they can extract more. While the grand coalition might still withstand the threats (as in this example), that is not generally the case, pointing to another important potential difference when costs are heterogenous. Consider, for instance:

$$A = 23, c_1 = 1, c_2 = 3, c_3 = 5.$$

Again using (28), we have $v(1, \pi^s) = 49, v(2, \pi^s) = 25, v(3, \pi^s) = 9$, while by (29):

$[v(1, \pi^{23}) = 64$ and $v(23, \pi^{23}) = 36], [v(2, \pi^{13}) = 36$ and $v(13, \pi^{13}) = 64], [v(3, \pi^{12}) = 21.78$ and $v(12, \pi^{12}) = 75.11]$,

and by (30), $v(N, \pi^g) = 121$. Observe that $\{1, \{23\}\}$ is viable because firms 2 and 3 together earn 36, which can be divided to ensure that 1 gets at least 25 and 3 gets at least 9. It can similarly be shown that the other two intermediate coalition structures are also viable, the precise allocations depend on the Nash weights. So the viability of the grand coalition hinges on its ability to generate some $x = (x_1, x_2, x_3)$ such that $x_1 \geqslant 64$, $x_2 \geqslant 36$, and $x_3 \geqslant 21.76$. These add up to $121.78 > v(N, \pi^g)$, so the grand coalition is not viable. $\diamond$

These applications exemplify the full integration of Nash bargaining into coalition formation, as opposed to some allocation rule such as unconditional equal division (Farrell and Scotchmer 1988, Bloch 1996). Alternatives such as equal division often bypass the need to hold coalitions together, thereby generating widespread inefficiency. Our coalitional Nash solution is constrained by the pragmatics of coalitional integrity. Nevertheless, that does not entirely prevent inefficiency, as we saw on more than one occasion in the examples.

In our framework, inefficiency can happen when there isn't too much of it. With "excessive" inefficiency, there would be too much to lose following a deviation, thereby eradicating the gains for a renegade subcoalition and avoiding that inefficiency. For instance, in Example 4, a single firm refrains from deviating, knowing that it would lead the entire structure to collapse. In contrast, in Example 2, the inefficiency occurs because there isn't "too much" of it when a single country breaks off from an agreement — the others stay together. The viability of an intermediate coalition structure is a double-edged sword: it avoids large degrees of inefficiency, but also threatens the viability of an efficient outcome.

## 5. Coalitional Nash Bargaining without Externalities

5.1. **Characteristic Functions.** Nash-in-Nash equilibrium also applies, of course, to settings without externalities, though in this case it is a bit pedantic to use the term "Nash-in-Nash". To describe it better, note that for each coalition, we are back to the coalitional Nash solution of Section 2, but with threats constructed endogenously via the recursive procedure in Section 3.2. When there are no externalities, we shall therefore refer to each coalitional component of a Nash-in-Nash equilibrium for a coalition structure as its coalitional Nash solution (with the understanding that the coalition threats are now endogenous).

Our aim in this Section is three-fold. First, we show that our recursive definition admits a simple non-recursive characterization. Second, we show that in TU games that satisfy grand coalition superadditivity, there is always an efficient coalitional Nash solution, *even when* the core is empty. Thirdly, we uncover a philosophical connection between the symmetric coalitional Nash solution and an ethic based on "pragmatic egalitarianism".

Our primitives are now summarized by a characteristic function: a collection of nonempty compact sets of feasible payoffs $F(S)$ for every $S$, along with disagreement payoffs $d$, with $x \geqslant d_S$ for all $x \in F(S)$. To rule out uninteresting cases, we presume that $x \gg d_S$ for some $x \in F(S)$ whenever $F(S) \neq \{d_S\}$. Normalize $d = 0$. A TU characteristic function is one in which each $S$ has a worth $v(S) \geqslant 0$, and $F(S) = \{x \in \mathbb{R}^S_+ | \sum_{i \in S} x_i \leqslant v(S)\}$.

We quickly redo our recursive definition for this simpler setting. For singleton sets $\{i\}$, set $\Theta^*(\{i\}) = F(\{i\})$. Now suppose that for some $S$ we already have $\Theta^*(T)$ for every $T \subset S$, and that each such $\Theta^*(T)$ is nonempty and compact. Define:

(31) $\quad U^*(S) \equiv \{x \in F(S) | x \text{ is unblocked by any } (T, y) \text{ with } T \subset S \text{ and } y \in \Theta^*(T)\}.$

If $U^*(S) = \emptyset$, declare $S$ to be ineffective, write $\Theta^*(S) = \{d_S\} = \{0_S\}$ and set $\sigma^*(S) = \emptyset$. Otherwise, $U^*(S)$ is nonempty and compact. In this case, set:

$$\Theta^*(S) \equiv \sigma^*(S) \equiv \underset{x \in U^*(S)}{\arg\max} \prod_{j \in S} x_j^{\gamma_j(S)}, \tag{32}$$

where $\gamma(S)$ are the bargaining weights for $S$. Now the recursion can continue to larger sets. Once completed, we have a full collection $\{\sigma^*(S)\}$ from (31) and (32). Each nonempty $\sigma^*(S)$ has a dual character: it is a solution for $S$, *and* a threat set that constrains larger coalitions, so $\Theta^*(S) = \sigma^*(S)$. As before, our overall solution is any minimal viable refinement $\pi$ of $N$, along with the coalitional Nash solution for each coalition in $\pi$.

Since subcoalitions are constrained to block with their own coalitional Nash solutions, the core of a characteristic function game could be a strict subset of $U^*(N)$. Our solution may therefore not lie in the core, making it diverge from that of Compte and Jehiel (2010). Indeed, the core could be empty while $U^*(N)$ is not. See Examples 5 and 6 below.

5.2. **A Simple Characterization.** Untangling a recursive definition can be hard work, and it is generally necessary to navigate every coalition structure, as we did in the examples of Section 4. However, when feasible payoff sets $F(S)$ for each coalition are *convex*, a simple characterization is available. To this end, define the *unconstrained Nash bargaining solution* $\Psi(S)$ for coalition $S$ (with $d$ normalized to zero) by ignoring all coalitional constraints:

$$\Psi(S) = \arg\max_{x \in F(S)} \prod_{j \in S} x_j^{\gamma_j(S)}.$$

If $F(S)$ is convex, $\Psi(S)$ has a single element $\psi(S)$. Let

(33) $$U^{\text{naive}}(S) \equiv \{x \in F(S) | x \text{ is unblocked by any } (T, \psi(T)) \text{ with } T \subset S\}$$

be the "naive" unblocked set for $S$ using the (generally non-credible) solutions $\psi(T)$ for each $T \subset S$ as threats. No recursive definition is needed for $U^{\text{naive}}(S)$. And yet:

**Theorem 2.** *Assume that $F(S)$ is compact and convex for every coalition $S$, and that $x \gg d_S = 0$ for some $x \in F(S)$ whenever $F(S) \neq \{0\}$. Then $U^*(S) = U^{\text{naive}}(S)$, and so*

$$\sigma^*(S) = \arg\max_{x \in U^{\text{naive}}(S)} \prod_{j \in S} x_j^{\gamma_j(S)};$$

*that is, the coalitional Nash solution for any coalition need only guard against the threats posed by the unconstrained Nash solutions of its subcoalitions.*

The theorem achieves a significant simplification. Instead of working through the entire recursion, which is demanding computationally, Theorem 2 asserts that the solution can be found by using an artificial threat for each subcoalition, consisting of the *unconstrained* Nash solution for that subcoalition.[26] The Supplementary Notes establish a substantially more general version of this result by weakening convexity, but also provide a counterexample to show that the convexity of feasible sets cannot be dispensed with entirely.

5.3. **Viability and Efficiency.** We return to the Coasian question asked of the examples in Section 4: is there a viable coalition structure that is efficient? As we saw in Section 4, this is a subtle issue when there are externalities, and it remains so even when externalities are absent. If, as in Section 2, we take coalitional threats to be exogenously given, the unblocked set for the grand coalition is the core of the characteristic function, which could well be empty. Indeed, in a TU game, the core is nonempty if and only if that game is

[26] The careful reader will observe that the proof of Theorem 2 applies to a wider class of solutions: those that emerge from the maximization of any system of coalitional welfare functions that are quasiconcave and strictly increasing in individual payoffs, under the assumption that all feasible sets are convex.

balanced (Bondareva 1963, Shapley 1967). When balancedness fails and the core is empty, any unblocked solution is inefficient, *even when there are no externalities*.

The coalitional Nash solution leaves more room for the unblocked set of the grand coalition to be nonempty. Subcoalitions are no longer free to block with any feasible payoff profile; they must block with their own coalitional Nash solutions. This fact, as well as the structure of the Nash solutions, will show that an efficient solution is guaranteed for all TU games satisfying two mild conditions. First, Nash bargaining weights are *consistent* in the following sense: there is $\gamma = (\gamma_1, \ldots, \gamma_n) \gg 0$ such that within any coalition, the bargaining weights are given by $\gamma_S = \{\gamma_i\}_{i \in S}$. Second, the game is *grand-coalition superadditive* (GCS): $v(N) \geqslant v(S) + v(T)$ for any pair of disjoint subcoalitions $S$ and $T$. This is, of course, significantly weaker than the Bondareva-Shapley requirement of balancedness.

**Theorem 3.** *Consider any GCS TU game with consistent bargaining weights. Then the grand coalition is viable, and its coalitional Nash solution is Pareto-efficient.*[27]

Theorem 3 stands in remarkable contrast to the possibility of an empty core in superadditive characteristic function games. As our proof shows, the contrast depends intimately on the use of the coalitional Nash solution at every level, both in the solution to be achieved and in the blocks or threats that potentially impede the achievement of that solution. Our argument is constructive, and makes crucial use of Theorem 2. We first build a particular structure and allocation, and then show that that allocation cannot be blocked by subcoalitions using *their* unconstrained Nash bargaining solution.

The following example illustrates how Theorem 3 relates to the core.

**Example 5.** Consider a TU game in which $N = \{123\}$, $v(i) = 0$ for $i$, $v(ij) = 0.8$ for all two-player coalitions and $v(N) = 1$. This is a GCS game but it is not balanced, and its core is empty. Now consider the symmetric solution with $\gamma_i = 1$ for all $i$. The coalitional Nash solution for every two-player coalition involves equal division, so each player gets $0.4$. While the core is empty, the unblocked set for $N$ relative to these threats is *not*. It includes, for example the allocation $(0.4, 0.4, 0.2)$ as well as all its permutations. Indeed, the coalitional Nash solution consists of precisely these allocations. ◇

Whether or not Theorem 3 generalizes to all TU games, and not just GCS TU games, is an open question. What we do know is that Theorem 3 does not generalize free of charge

[27] That is, there is no allocation $y$ for any structure $\pi$ with $\sum_{i \in S} y_i \leqslant v(S)$ for each $S \in \pi$ such that $y > x$.

to NTU games, even when they are superadditive. In such games, the feasible set can be arbitrary in ways that make it impossible to exploit the deeper structure of the coalitional Nash solution, as we do in Theorem 3.[28]

The GCS property is automatic when the grand coalition can do everything that subcoalitions can, as would occur with the embedding of $G$ into a standard game (this is true of all the applications in Section 4, for instance). But the grand coalition need not be viable when there are externalities, as we have already seen in Section 4.

5.4. **Constrained Egalitarianism.** We conclude by exploring an ethical foundation of the coalitional Nash solution. In what follows, we will focus on the solution with symmetric weights, so that $\gamma_i(S) = 1$ for all $S$ and $i \in S$. We also work with TU games, so that payoffs can be compared and transferred across individuals.

As already noted, the "split-the-difference" property of the Nash bargaining solution isn't an "ethical axiom," but a technical consequence of the invariance axiom.[29] And yet, given its undertones of "I set aside what's mine, you set aside yours, and we split the rest," this property has assumed a moral significance for what might be regarded as "fair"; see, e.g., the disputed garment principle in Aumann and Maschler (1985), or the foundations of the pre-kernel in Davis and Maschler (1965).

But our analysis has shown how fragile this interpretation is. Apart from the disagreement point, *all* coalitional threats enter as conventional constraints in our characterization. They affect the outcome only when it lies on the edge of the unblocked set. To us, that reopens the question of whether split-the-surplus can be viewed as an easy moral imperative. Our solution suggests a connection to a different ethic, one that we explore in this Section.

The constrained egalitarian solution (Dutta and Ray 1989, 1991) combines a commitment to equality with the pragmatics of coalitional participation. A social ethic (egalitarianism, in the case at hand) is applied to the greatest degree possible, while remaining limited by

[28]The familiar roommate problem can be used to establish this claim. Suppose $N = \{1, 2, 3\}$, $F(\{i\}) = \{0\}$ for all $i \in N$, $F(12) = \{(3, 2)\}, F(2, 3) = \{(3, 2)\}, F(1, 3) = \{(2, 3)\}$ and $F(N) = \{3, 2, 1), (1, 3, 2), (2, 1, 3)\}$. Clearly, every coalition structure of the form $\{jk, \ell\}$ is viable because no player in a two-player coalition will break away to precipitate the coalition structure of singletons. But this makes the grand coalition non-viable; every efficient allocation is blocked some two-player coalition. Thus, efficient viability may not be possible even when there are no externalities.

[29]If payoffs are expected utilities, as Nash takes them to be, the subtraction of disagreement points is viewable as an additive shift, and any solution must be invariant to such shifts.

the need to secure individual and coalitional buy-ins. The following example illustrates the notion of "pragmatic egalitarianism" used in Dutta and Ray:

**Example 6.** Consider a three-player TU game, with $d$ normalized to zero, $v(N) = 1$, $v(\{12\}) = 0.8$ and $v(S) = 0$ for all other $S$. Coalitions want distributional equality to the greatest degree possible, but don't want to break up (all this to be formalized below). Players 1 or 2 on their own obtain zero payoff, so $\{12\}$ can fully enjoy its egalitarian goals, implementing $\{0.4, 0.4\}$ if $\{12\}$ were to form. But the grand coalition $N$ is constrained by the possibility that players 1 and 2 could exit the grand coalition (with their credible threat $\{0.4, 0.4\}$) if pushed too far. A "constrained egalitarian solution" is therefore given by the set of two allocations $\{0.4, 0.3, 0.3\}$ and $\{0.3, 0.4, 0.3\}$, which are the Lorenz-maximal elements of the unblocked set for the grand coalition. The coalitional Nash solution also consists of these two allocations. In contrast, the core of this game is the set $\{x \in \mathbb{R}^3_+ \mid x_1 + x_2 \geqslant 0.8\}$ and the Compte-Jehiel (2010) solution is $(0.4, 0.4, 0.2)$. $\diamond$

Dutta and Ray (1989, 1991) generalize this idea to societies of all sizes and any specification of coalitional worths $\{v(S)\}_{S \subseteq N}$. Consider two payoff allocations $x$ and $y$ in $\mathbb{R}^k$ that add to the same total, arranged such that $x_i \leqslant x_{i+1}$ and $y_i \leqslant y_{i+1}$ for all $i = 1, \ldots, k-1$. Say that $x$ *majorizes* (or *Lorenz-dominates*) $y$ if $\sum_{i=1}^{j} x_i \geqslant \sum_{i=1}^{j} y_i$ for every $j = 1, \ldots, k$, with strict inequality for some $j$. This partial ordering is known to agree with the ethics of egalitarianism (see, e.g., Kolm 1969, Dasgupta, Sen and Starrett 1973 and Fields and Fei 1978). For any set of allocations $A$ adding to the same total, let $L(A)$ be its set of *Lorenz-maximal* elements: those allocation in $A$ not majorized by another allocation in $A$.

Now we define the constrained egalitarian solution $E(S)$ for every coalition $S$. For any singleton coalition $\{i\}$, define $E(\{i\}) = \{v(i)\}$. Recursively, consider any coalition $S$ and suppose that we've defined $E(T)$ for every strict subset $T$ of $S$. Then define

$$U^e(S) \equiv \{x \in F(S) | x \text{ is unblocked by any } (T, y) \text{ with } T \subset S \text{ and } y \in E(T)\}.$$

and, if $U^e(S) \neq \emptyset$, define the *constrained egalitarian solution* for coalition $S$ by

$$E(S) = L(U^e(S)).$$

Proceed in this manner until all coalitions are covered.

We can now connect constrained egalitarianism to the coalitional Nash solution:

**Theorem 4.** *In a TU game, the coalitional Nash solution $\sigma^{\text{sym}}$ with equal bargaining weights is a subset of the constrained egalitarian solution for every coalition $S$:*

$$\sigma^{\text{sym}}(S) \subseteq E(S).$$

*Suppose additionally that a TU game is superadditive, in that for every coalition $S$:*

$$v(S) \geqslant \sum_{j=1}^{m} v(T_j) \text{ for all partitions } (T_1, \ldots, T_m) \text{ of } S. \tag{34}$$

*Then for all $S$, $\sigma^{\text{sym}}(S)$ is nonempty, and is found by maximizing the Nash product over the set of allocations that are unblocked by any subcoalition using equal division.*

Theorem 4 states that the coalitional Nash solution with equal bargaining weights can be interpreted as a problem of constrained egalitarianism. The Nash product seeks to generate equality of outcomes, but could be prevented from doing so by coalitional threats. By Theorem 2, we can restrict ourselves to the naive threats imposed by unconstrained Nash bargaining. In a TU game, that's just equal division; hence the result. But Theorem 2 is also crucial in a different sense. It allows us to sidestep a common problem with the recursive definition; namely, that inductive arguments based on set-inclusion typically fail. Suppose inductively $\sigma^*(T) \subseteq E(T)$ for all coalitions of size $k$ or smaller. That *widens* the unblocked set $U^*(S)$ for a larger coalition $S$ relative to the Dutta-Ray unblocked set $U^e(S)$, and the inductive argument fails at this point. Theorem 2 avoids this line of reasoning altogether.

As a final remark of separate interest, one can use Theorem 4 to ask when the coalitional Nash solution is unique, at least up to a renaming of players. This is a deep question to which we do not have a complete answer. But Dutta and Ray (1991) provide some leads. They invoke the following partial ordering from Maschler and Peleg (1966): for two players $i$ and $j$, say that $i \succsim j$ if for all $S \subseteq N - \{ij\}$ (possibly empty), $v(S \cup \{i\}) \geqslant v(S \cup \{j\})$. This ordering is transitive (cf. Maschler and Peleg 1966), but additionally, there is a wide class of games for which $\succsim$ is complete. In such cases, we have

**Corollary 2** (to Theorem 4 and Dutta and Ray 1991, Theorem 3)**.** *If $\succsim$ is complete, then every allocation in $\sigma(N)$ is identical up to a permutation of the players.*

We omit the proof, which is an immediate consequence of combining Theorem 4 and Dutta and Ray (1991), Theorem 3. It remains to be seen whether we can do better. Our completeness condition is sufficient but far from necessary. For instance, it can be shown that the

Nash bargaining solution must contain a single allocation (again up to possible permutations) for all three player games, though $\succsim$ is not always complete for such games.[30]

## 6. Summary and Research Directions

We began by studying coalitional bargaining, in the axiomatic spirit of Nash's seminal work on two-person bargaining. As in his solution, our solution maximizes the product of payoffs net of individual disagreements, but all coalitional threats appear as conventional constraints. They are not netted out from payoffs. This gives rise to a new interpretation of Nash bargaining. In particular, we relate our coalitional solution to a notion of pragmatic egalitarianism, in which the social norm combines a preference for equal division with the practical need to secure individual and coalitional buy-ins.

In the baseline model, we presumed that every coalition has a well-defined set of threats for its members. We then went on to show how this seeming relic from cooperative game theory can be nicely embedded into an ambient setting with cross-player externalities. This embedding can be achieved by starting with a game in strategic form (actually a slight generalization of it) and allowing players to form coalitions. Then the threat set of each coalition $S$ depends additionally on the coalition *structure* that the coalition is embedded in. That set, in turn, is formed by noncooperative interaction across coalitions, coupled with the use of our coalitional Nash solution *within* coalitions. That is, threats are both end-solutions for a given coalition structure and constraints on play in coarser structures — all the way "up" to the grand coalition. We illustrate our solution concept with three different applications: the production of public goods with global externalities, competition in R&D, and Cournot interaction across firms.

Our coalitional Nash solution significantly expands the set of situations which can be studied by the Nash-in-Nash idea of combining cooperative solutions within coalitions with non-cooperative strategic interaction across them.

It should be reiterated that we have worked throughout with the case of "internal blocking," in which every coalition anticipates further deviations by its *sub*coalitions, but does not entertain other deviations engineered by coalitions that intersect our coalition but are not necessarily subsets of it. As already discussed, this extension will require that we replace our recursive definition of internal consistency by a consistency notion based on a fixed

[30]The coalitional Nash solution also exhibits a single allocation when the grand coalition has the highest average worth; that is, $v(N)/|N| \geqslant v(S)/|S|$ for every coalition $S$, but $\succsim$ might be incomplete.

point argument, or by a set of rules that prevents the same coalition structure from forming twice in any allowable blocking chain This is the subject of research to come, and the analysis here will serve as a base for such research. The Supplementary Notes contain further remarks on this topic.

## APPENDIX A. PROOFS AND SOME TECHNICAL REMARKS

By [Inv], there is no loss of generality in assuming that for every $G \in \mathcal{G}$, $d = 0$. This normalization will be in force throughout the proof. of Theorem 1. We shall make use of the following two Lemmas. The first one combines [Exp], [Inv] and [UHC] to show that a solution to any two-allocation problem with threats is also a solution to the same allocation problem when there are no threats.

**Lemma 1.** *Suppose a solution $\sigma(G)$ satisfies axioms* [Inv], [UHC] *and* [Exp] *for every game $G \in \mathcal{G}^0$. Then*

(35) *If $x$ and $y$ are unblocked by $\mathbf{\Theta}$ and $x \in \sigma(\{x, y\}, \mathbf{\Theta}, 0)$, then $x \in \sigma(\{x, y\}, \mathbf{\Theta}^0, 0)$;*

*that is, $x$ is also a solution to the pure bargaining problem with feasible set $F = \{x, y\}$.*

**Proof.** Consider any distinct $x, y \in \mathbb{R}^n_+$ and any effective collection of threats $\mathbf{\Theta}$ (that is, at least one coalition is effective), with both $x$ and $y$ unblocked, and with $x \in \sigma(\{x, y\}, \mathbf{\Theta}, 0)$. Say that $\mathbf{\Theta}' \precsim \mathbf{\Theta}$ if there exists $\alpha \in \mathbb{R}^n$ with $1 \gg \alpha \geqslant 0$ such that $\mathbf{\Theta}' = \alpha \circ \mathbf{\Theta}$ and $x \in \sigma(\{x, y\}, \mathbf{\Theta}', 0)$. Now fix $\bar{\mathbf{\Theta}}$ with $x$ and $y$ both unblocked under $\bar{\mathbf{\Theta}}$, and with $x \in \sigma(\{x, y\}, \bar{\mathbf{\Theta}}, 0)$. Let $\mathcal{T}$ collect all threat constellations $\mathbf{\Theta}$ such that $\mathbf{\Theta} \precsim \bar{\mathbf{\Theta}}$.

*We claim that $\mathbf{\Theta}^0 \in \mathcal{T}$, and in particular that $x \in \sigma(\{x, y\}, \mathbf{\Theta}^0, 0)$.*

First observe that $\mathcal{T}$ is nonempty. For by [Exp], there is $\lambda \gg 1$ such that $\lambda \circ x \in \sigma(\{\lambda \circ x, \lambda \circ y\}, \bar{\mathbf{\Theta}}, 0)$. Let $\alpha = (1/\lambda_1, \ldots, 1/\lambda_n)$. Then $1 \gg \alpha \geqslant 0$ (in fact $\alpha \gg 0$), and defining $\mathbf{\Theta} = \alpha \circ \bar{\mathbf{\Theta}}$, we deduce from [Inv] that $x \in \sigma(\{x, y\}, \mathbf{\Theta}, 0)$. So $\mathbf{\Theta} \in \mathcal{T}$.

Note that $\mathcal{T}$ is partially ordered by $\precsim$.[31] For any totally ordered subset $\mathcal{T}^c$ of $\mathcal{T}$, define

$$a \equiv \inf\{\|\alpha\| | \mathbf{\Theta} = \alpha \circ \bar{\mathbf{\Theta}} \text{ for some } \mathbf{\Theta} \in \mathcal{T}^c\}.$$

If there is $\mathbf{\Theta}^* \in \mathcal{T}^c$ of the form $\mathbf{\Theta}^* = \alpha^* \circ \bar{\mathbf{\Theta}}$ where $\|\alpha^*\| = a$, then clearly $\mathbf{\Theta}^* \precsim \mathbf{\Theta}$ for every $\mathbf{\Theta} \in \mathcal{T}^c$, and so $\mathbf{\Theta}^*$ is a lower bound for $\mathcal{T}^c$. Otherwise, by the definition of $a$, there

[31] Every element $\mathbf{\Theta}$ of $\mathcal{T}$ has the property that $\mathbf{\Theta} = \alpha \circ \bar{\mathbf{\Theta}}$ for some $\alpha$ with $1 \gg \alpha \geqslant 0$. Therefore $x$ and $y$ are also unblocked under $\mathbf{\Theta}$. Now the transitivity of $\precsim$ on $\mathcal{T}$ is immediate.

is a sequence $\{\boldsymbol{\Theta}^k\}$ in $\mathcal{T}^c$ with $\boldsymbol{\Theta}^k = \alpha^k \odot \bar{\boldsymbol{\Theta}}$, with $1 \gg \alpha^k \geqslant 0$ for every $k$, and $\|\alpha^k\| \to a$. Let $(\boldsymbol{\Theta}^*, \alpha^*)$ be any limit point of $\{\boldsymbol{\Theta}^k, \alpha^k\}$, the first component under the product Hausdorff metric and the second in the standard sense.[32] Then $\boldsymbol{\Theta}^* = \alpha^* \odot \bar{\boldsymbol{\Theta}}$. Moreover, because $x \in \sigma(\{x,y\}, \boldsymbol{\Theta}^k, 0)$ for every $k$ along the sequence of threat constellations $\{\boldsymbol{\Theta}^k\}$, it follows from [UHC] that $x \in \sigma(\{x,y\}, \boldsymbol{\Theta}^*, 0)$. Therefore $\boldsymbol{\Theta}^* \in \mathcal{T}$ and just as in the previous case, $\boldsymbol{\Theta}^* \precsim \boldsymbol{\Theta}$ for every $\boldsymbol{\Theta} \in \mathcal{T}^c$, and so serves as a lower bound for $\mathcal{T}^c$.

We may therefore apply Zorn's Lemma to assert that $\mathcal{T}$ admits a minimal element $\underline{\boldsymbol{\Theta}}$. We claim that $\underline{\boldsymbol{\Theta}} = \boldsymbol{\Theta}^0$; that is, all coalitions are ineffective under $\underline{\boldsymbol{\Theta}}$. Otherwise, if $\underline{\boldsymbol{\Theta}} \in \mathcal{T}$ is effective, it could be contracted further, which violates minimality. We have therefore proved the claim. And as an implication, we have established (35). ■

**Lemma 2.** *A solution $\sigma(G)$ satisfies axioms* [Inv], [IIA], [UHC] and [PO] *for every game $G = (F, \Theta^0, 0) \in \mathcal{G}^0$, with $F = \{x, y\}$, if and only if*

$$\sigma(G) = \arg\max_{x \in F \cap \mathbb{R}^N_+} \prod_{j \in N} x_j^{\gamma_j}$$

*for some $\gamma = (\gamma_1, \ldots, \gamma_n) \gg 0$.*

**Remarks.** It is easy to see that the weighted Nash solution satisfies the axioms of Lemma 2. To prove the converse, we shall follow closely the proof of Theorem 2 of Kaneko (1980), keeping in mind that we wish to dispense with two additional assumptions that he makes: a slightly stronger version of [Sym] as well as strong individual rationality:

[SIR] If $x \in \sigma(G)$, then $x \gg d$.

It is particularly important to drop [SIR], because it assumes precisely what we want to derive: the fact that disagreement points are generally not binding in the conventional sense, while, in contrast, coalitional constraints do bind in that sense.

**Proof.** Define a binary relation $\succsim$ on $\mathbb{R}^N_{++}$ by $x \succsim y$ if and only if $x \in \phi(\{x, y\})$. Because Kaneko's (and our) domain includes sets containing 2 points, $\phi(\{x, y\})$ is well defined.[33] We follow Kaneko's argument to assert that $\succsim$ is transitive, complete, continuous and monotonic in the sense that $x \succ y$ whenever $x > y$. The last property is a consequence of [PO]. It follows that $\succsim$ can be represented by a continuous, monotonic utility function $H : \mathbb{R}^N_{++} \to \mathbb{R}$; that is, $H(x) \geqslant H(y)$ if and only if $x \succsim y$. Next, Kaneko uses [Inv] to

32These are well-defined given that $0 \leqslant \alpha^k \ll 1$ for every $k$.

33We will later extend this binary relation to $\mathbb{R}^N_+$, a step that Kaneko can skip because he assumes [SIR].

apply a Lemma of Osborne (1976) to show that $H$ can be taken to be of the form

(36) $$H(x) = \prod_{j \in N} x_j^{\gamma_j},$$

where $\gamma \gg 0$. Osborne's Lemma only guarantees that $\gamma \geqslant 0$, but [PO] implies $\gamma \gg 0$.[34]

We now verify that $\succsim$ can be extended to $\mathbb{R}_+^N$ while maintaining the representation (36). By (36) applied to strictly positive vectors, we can see that for any $x \gg 0$, $i \in N$ and positive number $a$, there is $y \gg 0$ such that $y_i < x_i$, $y_j > a$ for all $j \neq i$ and $\prod_{j \in N} y_j^{\gamma_j} < \prod_{j \in N} x_j^{\gamma_i}$. By the representation property, this means that $x \succ y$, in the sense that $\phi(\{x, y\}) = \{x\}$.

Now consider any $z \in \mathbb{R}_+^N$. with $z_i = 0$ for some $i \in N$. By [PO], $\phi(\{y, z\}) = \{y\}$ for any $y \gg z$, and in particular, defining $a = \max_{j \in N} z_j$, this is true of the $y$ identified in the previous paragraph; i.e., $y \succ z$. Thus, we have

$$x \succ y \text{ or } \phi(\{x, y\}) = \{x\} \text{ and } y \succ z \text{ or } \phi(\{y, z)\} = \{y\}.$$

Observe that if $z \in \phi(\{x, y, z\})$, then by [IIA], $z \in \phi(\{y, z\})$, which contradicts $\phi(\{y, z\}) = \{y\})$. Similarly, by [IIA], $y \in \phi(\{x, y, z\})$ would contradict $\phi(\{x, y\}) = \{x\}$. Thus $\phi(\{x, y, z\}) = \{x\}$. Another application of [IIA] implies that $\phi(\{x, z\}) = \{x\}$, i.e., $x \succ z$.

We have shown that for any $x \in \mathbb{R}_{++}^N$ and any $z$ that is not in $\mathbb{R}_{++}^N$, $\phi(\{x, z\}) = \{x\}$. By [Dom 1], there exists $x \in F$ with $x \gg 0$. By [IIA] it follows that if $z_i = 0$ for any $i$, then $z \notin \phi(F)$. In other words $\phi$ satisfies [SIR], and the characterization $\sigma(G) = \arg\max_{x \in F} \prod_{j \in N} x_j^{\gamma_j}$ holds over $\mathbb{R}_+^N$ for all pure bargaining problems with two-point feasible sets. ■

**Proof of Theorem 1.** It is straightforward that the coalitional Nash solution satisfies [Inv], [IIA], [PO] and [Exp]. It follows from the argument used in the proof of Theorem 2 in Kaneko (1980) that it also satisfies [UHC]. We now prove the converse.

Lemma 2 implies that there exists $\gamma \gg 0$ such that for any $x, y \in \mathbb{R}_+^N$,

$$\text{If } x \in \sigma(\{x, y\}, \boldsymbol{\Theta}^0, 0), \text{ then } \prod_{j \in N} x_j^{\gamma_j} \geqslant \prod_{j \in N} y_j^{\gamma_j}.$$

Recalling (35) from Lemma 1, this implies that even if $\boldsymbol{\Theta} \neq \boldsymbol{\Theta}^0$:

(37) If $x$ and $y$ are unblocked by $\boldsymbol{\Theta}$ and $x \in \sigma(\{x, y\}, \boldsymbol{\Theta}, 0)$, then $\prod_{j \in N} x_j^{\gamma_j} \geqslant \prod_{j \in N} y_j^{\gamma_j}$.

---

[34] At this point in his proof, Kaneko focuses on the symmetric Nash solution by appealing to a form of symmetry and setting $\gamma = (1, \ldots, 1)$.

Now consider a game $G = (F, \boldsymbol{\Theta}, 0) \in \mathcal{G}$ and $x \in \sigma(G)$. Take any $y$ that is unblocked by $\boldsymbol{\Theta}$. Since $x, y \in F$, and $x \in \sigma(G)$, [IIA] implies that $x \in \sigma(\{x, y\}, \boldsymbol{\Theta}, 0)$. As this holds for any $y$ unblocked by $\boldsymbol{\Theta}$, it follows from (37) that

$$\sigma(G) = \arg\max_{x \in U(G)} \prod_{j \in N} x_j^{\gamma_j}.$$

■

**Remarks:** Although we have relied on Kaneko (1980) in proving Lemma 2, the literature on Nash bargaining with nonconvex feasible sets[35] allows for other options provided some form of symmetry is assumed. Particularly relevant in our context is Mariotti (1998), which shows that Kaneko's characterization can be sharpened to drop [UHC] as well as strong individual rationality. Mariotti (1998) does, however, assume *anonymity* which is stronger than [Sym]. Indeed, our proof of Corollary 1 (but not of Theorem 1) could have proceeded by applying Theorem 4.3 of Mariotti (1998) in place of Lemma 2. The main advantage of doing so would be that [UHC] wouldn't be needed in this part of the proof. However, we do need [UHC] when using Zorn's Lemma to prove Lemma 1. In fact, [UHC] cannot be dropped from Theorem 1, as we show in the Supplementary Notes.

**Proof of Theorem 2.** Our argument combines Claims 1–3 below.

**Claim 1.** For any coalition $S$, $U^*(S) \subseteq U^{\text{naive}}(S)$.

*Proof.* Suppose $x \in U^*(S)$ but $x \notin U^{\text{naive}}(S)$. Then there is some coalition $T \subset S$ such that $\psi(T) \gg x_T$. If $\psi(T) \in \Theta(T)$, that would contradict $x \in U^*(S)$. If $\psi(T) \notin \Theta(T)$, then there is $W \subset T$ and $y \in \Theta(W)$ with $y \gg \psi(T)_W$. But then $(W, y)$ also blocks $x$, and we again have a contradiction to the supposition that $x \in U^*(S)$. ■

**Claim 2.** Suppose there exists $x \in \Theta(T)$ such that $x \neq \psi(T)$. Then there exists $W \subset T$ such that $\psi(W) \geqslant x_W$.

*Proof.* We proceed by induction on coalition size. Clearly the assertion is trivially true for all coalitions of size 2 or less: for such coalitions $\Theta(T) = \Psi(T) = \{\psi(T)\}$. Now suppose that the lemma is true for any coalition of size $k$ or less, where $k \geqslant 2$. Let $T$ have cardinality $k + 1$. Suppose that $x \in \Theta(T)$ and $x \neq \psi(T)$. Observe that the Nash product value of $\psi(T)$ strictly exceeds that of $x$. By the quasi-concavity of the Nash product, this must be also true of any allocation of the form $tx + (1 - t)\psi(T)$ for $t \in (0, 1)$, which all lie in $F(T)$ because $F(T)$ is convex. Therefore each such allocation must be blocked by

[35] See Xu and Yoshihara (2020) and Thomson (2022).

some subcoalition of $T$ using some element of its coalitional Nash solution. In particular, by taking $t$ to 1, we see that there exist $H \subset T$ and $y \in \Theta(H)$ such that

$$y \geqslant x_H \tag{38}$$

If $y = \psi(H)$ we are done. If $y \neq \psi(H)$, then noting that $|H| \leqslant k$, the induction hypothesis implies there is $W \subset H$ with $\psi(W) \geqslant y_W$. Combining this information with (38), we must conclude that $\psi(W) \geqslant x_W$. ■

**Claim 3.** For any coalition $S$, $U^{\text{naive}}(S) \subseteq U^*(S)$.

*Proof.* Suppose there is $x \in U^{\text{naive}}(S)$ but with $x \notin U^*(S)$. Then there is some coalition $T$ and $y \in \Theta(T)$ such that $y \gg x_T$. Because $x \in U^{\text{naive}}(S)$, we have $y \neq \psi(T)$. But then by Claim 2, there exists $W \subset T$ such that $\psi(W) \gg y_W \gg x_W$, which contradicts the hypothesis that $x \in U^{\text{naive}}(S)$. ■

The Theorem is an immediate consequence of Claims 1 and 3. ■

**Proof of Theorem 3.** For any coalition $S$, define $\gamma_S \equiv \sum_{j \in S} \gamma_j$ and $\alpha(S) \equiv v(S)/\gamma_S$. Let $S_1 \in \arg\max_{S \subseteq N} \alpha(S)$. Recursively, having defined $S_1, \ldots S_{k-1}$, let $S_k \in \arg\max \alpha(S)$ over $S \subseteq N - \cup_{j=1}^{k-1} S_j$. This yields a structure $\pi = \{S_1, \ldots S_m\}$, where $\alpha(S_j) \geqslant \alpha(S_{j+1})$ for all $j = 1, \ldots m-1$. Consider the allocation $y = (y_1, \ldots, y_n)$ in $\pi$ obtained by horizontally stacking the unconstrained Nash payoffs for each coalition in $\pi$. That is, if $i \in S_j$, then (remembering that $d = 0$) $y_i = \gamma_i \alpha(S_j)$. Finally, recalling grand-coalition superadditivity, we note that there is $x \geqslant y$ that is feasible for the grand coalition, and Pareto-efficient.

We claim that there is no $T \subseteq N$ and $z \in \mathbb{R}^T$ such that $z$ is the unconstrained Nash bargaining solution for $T$ and $z \gg x$. For if not, then there is $T$ and $z$ feasible for $T$ such that $z_i = \gamma_i \alpha(T) > y_i$ for all $i \in T$. Let $S_k$ be the lowest indexed coalition in $\pi$ that has a nonempty intersection with $T$. Then for any $i \in T \cap S_k$,

$$\gamma_i \alpha(T) > \gamma_i \alpha(S_k).$$

But this means that $\alpha(T) > \alpha(S_k)$, which contradicts our construction of the partition $\pi$.

So the allocation $y$ — and *a fortiori* the allocation $x$ — is unblocked by any coalition $T$ using its unconstrained Nash bargaining solution. Thus, $U^{\text{naive}}(N) \neq \emptyset$. Because $F(S)$ is compact and convex for all $S$, Theorem 2 applies, and $U^*(N) \neq \emptyset$. We can now maximize $\prod_{j \in N} x_j$ over this non-empty, compact set to obtain the coalitional Nash solution. Because the grand coalition is viable and GCS holds, any such solution is efficient. ■

**Proof of Theorem 4.** In a TU game with symmetric weights, the unconstrained Nash solution for every $T$ is the same as the *equal division solution*:

(39) $$\psi(T) = e(T) = \left(\frac{v(T)}{|T|}, \ldots, \frac{v(T)}{|T|}\right).$$

Because $F(S)$ is compact and convex for all $S$, Theorem 2 applies. Given that $\psi(T) = e(T)$ as noted in (39), we must conclude that

(40) $$\Theta^*(S) = \sigma^*(S) = \underset{x \in U^{\text{naive}}(S)}{\arg\max} \prod_{j \in S} x_j.$$

where

(41) $$U^{\text{naive}}(S) = \{x \in F(S) | x \text{ is unblocked by any } (T, y) \text{ with } T \subset S \text{ and } y = e(T)\}$$

is nonempty. But by Theorem 1 of Dutta and Ray (1991),

(42) $$E(S) = L(U^{\text{naive}}(S))$$

when $U^{\text{naive}}(S) \neq \emptyset$. Given (40) and (42), it suffices to prove that for every nonempty compact set $A \subset \mathbb{R}^k$ in which all allocations have the same total, $\arg\max_{x \in A} \prod_{i=1}^k x_i \subseteq L(A)$. But the product maximand is increasing and strictly quasi-concave, and the agreement of such functions (in value) with majorization is well known; see, e.g., Kolm (1969), Atkinson (1970) and Dasgupta, Sen and Starrett (1973).

The second part follows immediately from Theorem 3 specialized to the case of equal weights. ■

## References


**Atkinson, Anthony**, "On the Measurement of Inequality," *Journal of Economic Theory*, 1970, *2* (3), 244–263.

**Aumann, Robert J and Michael Maschler**, "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud," *Journal of Economic Theory*, 1985, *36* (2), 195–213.

**Bagwell, Kyle, Robert W. Staiger, and Ali Yurukoglu**, ""Nash-in-Nash" tariff bargaining," *Journal of International Economics*, 2020, *122*, 103263.

**Bernheim, B. Douglas and Debraj Ray**, "Collective Dynamic Consistency in Repeated Games," *Games and Economic Behavior*, 1989, *1* (4), 295–326.

**Bernheim, B.Douglas, Bezalel Peleg, and Michael D Whinston**, "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 1987, *42* (1), 1–12.

**Bhattacharya, Ananyo**, *The Man from the Future: The Visionary Life of John von Neumann*, W. W. Norton & Company, 2022.

**Binmore, K, A Shaked, and J Sutton**, "An Outside Option Experiment," *Quarterly Journal of Economics*, 1989, *104* (4), 753–770.

**Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky**, "The Nash Bargaining Solution in Economic Modelling," *The RAND Journal of Economics*, 1986, pp. 176–188.

**Bloch, Francis**, "Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division," *Games and Economic Behavior*, 1996, *14*, 90–123.

**Bondareva, Olga N.**, "Some Applications of Linear Programming Methods to the Theory of Cooperative Games," *Problemy Kybernetiki*, 1963, *10*, 119–139.

**Burguet, Roberto and Ramon Caminal**, "Coalitional Bargaining with Consistent Counterfactuals," 2016. mimeo., Instituto de An'alisis Económico (CSIC), Barcelona.

**Chatterjee, Kalyan, Bhaskar Dutta, Debraj Ray, and Kunal Sengupta**, "A Noncooperative Theory of Coalitional Bargaining," *The Review of Economic Studies*, 1993, *60* (2), 463.

**Collard-Wexler, Allan, Gautam Gowrisankaran, and Robin S. Lee**, ""Nash-in-Nash" Bargaining: A Microfoundation for Applied Work," *Journal of Political Economy*, 2019, *127* (1), 163–195.

**Compte, Olivier and Philippe Jehiel**, "The Coalitional Nash Bargaining Solution," *Econometrica*, 2010, *78* (5), 1593–1623.

**Dasgupta, Partha, Amartya Sen, and David Starrett**, "Notes on the Measurement of Inequality," *Journal of Economic Theory*, 1973, *6* (2), 180–187.

**Davis, Morton and Michael Maschler**, "The Kernel of a Cooperative Game," *Naval Research Logistics Quarterly*, 1965, *12* (3), 223–259.

**Dutta, Bhaskar and Debraj Ray**, "A Concept of Egalitarianism Under Participation Constraints," *Econometrica*, 1989, *57* (3), 615–635.

**______ and ______**, "Constrained Egalitarian Allocations," *Games and Economic Behavior*, 1991, *3*, 403–422.

**Farrell, Joseph and Eric Maskin**, "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1989, *1* (4), 327–360.

**______ and Suzanne Scotchmer**, "Partnerships," *The Quarterly Journal of Economics*, 1988, *103* (2), 279–297.

**Fields, Gary S and John CH Fei**, "On Inequality Comparisons," *Econometrica*, 1978, pp. 303–316.

**Haeringer, Guillaume**, "Equilibrium Binding Agreements: A Comment," *Journal of Economic Theory*, 2004, *117* (1), 140–143.

**Harsanyi, John C.**, "A Bargaining Model for the Cooperative $n$-Person Game," in R. D. Luce and A. W. Tucker, eds., *Contributions to the Theory of Games IV*, Princeton: Princeton University Press, 1959, pp. 325–355.

**Ho, Kate and Robin S. Lee**, "Insurer Competition in Health Care Markets," *Econometrica*, 2017, *85* (2), 379–417.

______ **and** ______, "Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets," *American Economic Review*, February 2019, *109* (2), 473–522.

**Horn, Henrick and Asher Wolinsky**, "Bilateral Monopolies and Incentives for Merger," *The RAND Journal of Economics*, 1988, *19* (3), 408–419.

**Ichiishi, Tatsuro**, "A Social Coalitional Equilibrium Existence Lemma," *Econometrica*, 1981, *49* (2), 369–377.

**Isbell, J. R.**, "A modification of Harsanyi's Bargaining Model," *Bulletin of the American Mathematical Society*, 1960, *66* (2), 70–73.

**Kalai, Ehud and Meir Smorodinsky**, "Other Solutions to Nash's Bargaining Problem," *Econometrica*, 1975, *43* (3), 513–518.

**Kaneko, Mamoru**, "An Extension of the Nash Bargaining Problem and the Nash Social Welfare Function," *Theory and Decision*, 1980, *12*, 135–148.

**Kolm, S. Ch.**, "The Optimal Production of Social Justice," in Margolis J. and H. Guitton, eds., *Public Economics*, Macmillan UK, 1969, pp. 145–200.

**Lee, Robin S. and Kyna Fong**, "Markov-Perfect Network Formation: An Applied Framework for Bilateral Oligopoly and Bargaining in Buyer-Seller Networks," 2013. unpublished.

**Loury, Glenn C.**, "Market Structure and Innovation," *The Quarterly Journal of Economics*, 1979, *93* (3), 395–410.

**Lucas, William F.**, "On Solutions to $n$-Person Games in Partition Function Form." Ph.d. dissertation, University of Michigan, Ann Arbor, MI 1963.

**Mariotti, Marco**, "Nash bargaining theory when the number of alternatives can be finite," *Social Choice and Welfare*, 1998, *15* (3), 413–421.

**Maschler, Michael and Bezalel Peleg**, "A Characterization, Existence Proof and Dimension Bounds for the Kernel of a Game," *Pacific Journal of Mathematics*, 1966, *18* (2), 289–328.

**Myerson, Roger B.**, "Values of Games in Partition Function Form," *International Journal of Game Theory*, 1977, *6*, 23–31.

**Nash, John F.**, "The Bargaining Problem," *Econometrica*, 1950, *18* (2), 155–162.

**Osborne, D.K.**, "Irrelevant Alternatives and Social Welfare," *Econometrica*, 1976, *44* (5), 1001–1015.

**Ray, Debraj and Rajiv Vohra**, "Equilibrium Binding Agreements,," *Journal of Economic Theory*, mar 1997, *73* (1), 30–78.

______ **and** ______, "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior*, 1999, *26* (2), 286–336.

______ **and** ______, "Coalitional Power and Public Goods," *Journal of Political Economy*, 2001, *109* (6), 1355–1385.

**Reinganum, Jennifer F.**, "A Dynamic Game of R and D: Patent Protection and Competitive Behavior," *Econometrica*, 1982, *50* (3), 671–688.

**Serrano, Roberto and Ken Ichi Shimomura**, "Beyond Nash Bargaining Theory: The Nash Set," *Journal of Economic Theory*, 1998, *83* (2), 286–307.

**Shapley, Lloyd and Martin Shubik**, *Game Theory in Economics - Chapter 4: Preferences and Utility*, Rand Corporation Report R-904/4-NSF, 1974.

**Shapley, Lloyd S.**, "On Balanced Sets and Cores," *Naval Research Logistics Quarterly*, 1967, *14*, 453–460.

**Thomson, William**, "Chapter 35, Cooperative Models of Bargaining," in Robert J Aumann and Sergiu Hart, eds., *Handbook of Game Theory with Economic Applications*, Vol. 2, Elsevier, 1994, pp. 1237–1284.

______, "On the axiomatic theory of bargaining: a survey of recent results," *Review of Economic Design*, 2022, *26* (4), 491–542.

**Thrall, Robert M. and William F. Lucas**, "$N$-Person Games in Partition Function Form," *Naval Research Logistics Quarterly*, 1963, *10* (1), 281–298.

**von Neumann, John**, *Private Communication to Professor David Gale*, Container 24, John von Neumann and Klára Dán von Neumann Papers, Manuscript Division, Library of Congress, Washington, D.C., 1949.

**Xu, Yongsheng and Naoki Yoshihara**, "Nonconvex Bargaining Problems: Some Recent Developments," *Homo Oeconomicus*, 2020, *37* (1-2), 7–41.

**Zhou, Lin**, "The Nash Bargaining Theory with Non-Convex Problems," *Econometrica*, 1997, *65* (3), 681–685.