BROWN
Orlando Bravo Center
for Economic Research

Nash Bargaining with Coalitional Threats

Bravo Working Paper # 2024-001

Debraj Ray and Rajiv Vohra*

March 2024

Abstract We axiomatically characterize bargaining outcomes in the presence of coalitional threats. As in Nash's solution, these involve the product of payoffs net of disagreement points, but coalitional threats appear as conventional constraints, and are not netted out from payoffs as disagreement points are. This asymmetry is implied by a new "expansion axiom" (along with standard axioms), one that is automatically satisfied in the standard bargaining problem. We then endogenize coalitional threats using internal consistency, requiring coalitions to be constrained by *their* subcoalitions just as the grand coalition is. For games with convex payoff sets, this consistent solution coincides with one in which the only threat from each coalition is their "standard" Nash solution, unconstrained by subcoalitions. For transferable-utility games, this observation uncovers a connection between the coalitional solution and the egalitarian solution of Dutta and Ray (1989, 1991).

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

Nash Bargaining with Coalitional Threats

Debraj Ray and Rajiv Vohra[†]

March 2024

Abstract We axiomatically characterize bargaining outcomes in the presence of coalitional threats. As in Nash's solution, these involve the product of payoffs net of disagreement points, but coalitional threats appear as conventional constraints, and are not netted out from payoffs as disagreement points are. This asymmetry is implied by a new "expansion axiom" (along with standard axioms), one that is automatically satisfied in the standard bargaining problem. We then endogenize coalitional threats using internal consistency, requiring coalitions to be constrained by *their* subcoalitions just as the grand coalition is. For games with convex payoff sets, this consistent solution coincides with one in which the only threat from each coalition is their "standard" Nash solution, unconstrained by subcoalitions. For transferable-utility games, this observation uncovers a connection between the coalitional solution and the egalitarian solution of Dutta and Ray (1989, 1991).

1. Introduction

This paper studies Nash bargaining with coalitional threats. We establish an axiomatic characterization of a bargaining solution in the presence of exogenously given coalitional threats. Our solution is closely related to Nash's solution for bargaining problems, but coalitional threats are incorporated in a particular way that we shall discuss in detail. We then go on to endogenize coalitional threats by imposing internal consistency, which asks that the threats posed by any coalition must form a solution to a parallel, recursively defined problem, constrained by threats from *their* subcoalitions.

The first exercise faithfully attempts to follow Nash's original axioms, making necessary changes to accommodate the presence of coalitions. A *game with coalitional threats* is a triple $G = (F, \mathbf{\Theta}, d)$, where $F \subset \mathbb{R}^n$ is a set of feasible outcomes for N, the grand coalition of all players, $\mathbf{\Theta}$ collects sets $\Theta(S) \subset \mathbb{R}^{|S|}$ of payoff threats that each coalition S has at its disposal, and d is a vector of disagreement payoffs for each player when no coalition has

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

formed at all. We prove that a solution $\sigma(G)$ satisfies our axioms if and only if

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

over the set of all allocations $x \in F$ that are unblocked by some coalitional threat (S, y), where $S \subset N$ and $y \in \Theta(S)$; see Theorem 1 for a precise statement.

While this result and our axiomatic formalization are new, it owes much to Kaneko (1980). Kaneko works with a rich domain of feasible sets without any coalitional threats. But "feasibility" is a broad concept. It could include non-blocking by coalitions. Had we treated no-blocking on par with feasibility *by assumption*, Theorem 1 would stand as a modest generalization of Kaneko's Theorem 2, with some interpretative interest. But our approach does not assume this equivalence; rather, it is derived from a substantially broader framework. Our derivation lays bare a central point of interest to us, which is the asymmetric treatment of individual disagreement payoffs and coalitional threats embodied in (1).

Specifically, disagreement payoffs are subtracted from overall payoffs, as they are in Nash's theorem, "before" maximization proceeds. But coalitional threats, including individual threats, appear as standard constraints that bind in the conventional way. Nash's original theorem in a two-player setting characterizes a solution as maximizing the product of payoff excesses over individual disagreement payoffs:

$$\sigma(F, d) = \arg\max_{\mathbf{x} \in F} \{(x_1 - d_1)(x_2 - d_2) | x \geqslant d\} \tag{2}$$

for every bargaining problem, where F is a compact, convex set of feasible payoffs and $d \in F$ is the disagreement payoff. That is, the bargain sets aside what the players can get on their own, and then divides any surplus. This familiar property emerges from the invariance axiom, which states that an affine transformation of payoffs should not affect the solution. It is often viewed as a "fairness" principle; see, e.g., the "disputed garment principle" in Aumann and Maschler (1985), or the pre-kernel of Davis and Maschler (1965).

Now consider instead the maximization of the product of *gross* payoffs from F, subject to the no-blocking constraints imposed by individual threats $\zeta_i \equiv \max \Theta(\{i\})$:

$$\tilde{\sigma}(F, \zeta) = \arg\max_{x \in F} \{x_1 x_2 | x \geqslant \zeta\}. \tag{3}$$

(3) is fully compatible with (1) if we allow for disagreement payoffs d_i and individual threats ζ_i to differ for each i, normalizing the former to zero using Nash's invariance axiom. However, if we refuse to make a distinction between these two objects — which may well

be reasonable — then (3) is killed off by the invariance axiom. It all depends on how we view d_i, which is the payoff accruing to any individual if *no* coalition forms, compared to the payoff ζ_i that accrues when the single player i goes off on her own.

In a dynamic context, Binmore, Shaked, and Sutton (1989) and Compte and Jehiel (2010) draw a distinction between d and ζ.[1] Binmore et al. (1989) note that ζ refers to outside options or threats that players can access only if they *abandon* negotiations.[2] They provide experimental evidence as well as strategic analysis (via a Rubinstein-style alternating offers game with outside options) in support of (3) over (2). In their espousal of (3), they write:

> "The attraction of split-the-difference lies in the fact that a larger outside option seems to confer greater bargaining power. But how can a bargainer use his outside option to gain leverage? By threatening to play the deal-me-out card. 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].

Unlike this approach, our coalitional solution (1) is entirely based on axiomatics. It allows for d to equal or differ from ζ, and therefore in effect for both (2) and (3). In this sense we take a weaker stance. At the same time, and from a completely different perspective, we agree with Binmore et al. (1989) that there is *no* parallel room for ambiguity when it comes to coalitional threats. No coalitional threat is subtracted from payoff in our solution, whatever "subtract" might exactly mean in a vector-valued context. Specifically, such threats are binding only in the standard sense of the term: they affect the solution only when it lies on the edge of the unblocked set. That asymmetry could of course be assumed directly by asserting that coalitional constraints are just one instance of feasibility; on par with technological constraints, for instance. But simply *assuming* that equivalence does not allow us to examine it. Rather, our result is derived from a simple "expansion axiom" that's automatically met in a world without coalitional threats. Specifically:

Suppose there are just two feasible allocations x and y, neither of which is blocked by any threat, and suppose that x belongs to the solution. Then there exists some $\lambda \in \mathbb{R}^N$ with $\lambda_i > 1$ for all $i \in N$ such that under an unchanged threat constellation, $\lambda \otimes x$ belongs

[1] See also Binmore, Rubinstein, and Wolinsky (1986), Chatterjee, Dutta, Ray, and Sengupta (1993) and Okada (1996).

[2] Binmore et al. (1989) refer to our disagreement payoff as an impasse point and an individual threat as an outside option. We prefer our terminology in this axiomatic setting with no explicit timeline.

to the solution from the two-point set $\{\lambda \otimes x, \lambda \otimes y\}$, where for any $z \in \mathbb{R}^N$, $\lambda \otimes z = (\lambda_1 z_1, \ldots, \lambda_n z_n)$.

In a world without coalitions, the expansion axiom above is automatically implied by the invariance axiom. With coalitional threats, it is an independent assumption. This seemingly innocuous restriction, combined with other standard restrictions, delivers the asymmetric nature of coalitional threats discussed above. See Example 1 for more discussion.

We then turn to an endogenous determination of the coalitional threat sets $\mathbf{\Theta}$. We focus exclusively on the internal consistency of those sets, leaving an exploration of "external consistency" to a subsequent contribution. Internal consistency requires that coalitions are to be constrained (by *their* subcoalitions) in just the same way as the grand coalition is. If the latter attempts a Nash-like bargain that is immune to blocking, so must the former, viewed as a "mini-society" which is subject to threats from *its* subcoalitions. But that naturally imposes a credibility constraint on each coalition's threat set: coalitions realize that they cannot block freely and will be subject to the same forces as the grand coalition.[3]

Following this idea, internally consistent coalitional Nash solutions are defined recursively using Theorem 1, building up from singleton coalitions. Each coalitional solution in turn becomes a threat set facing still larger coalitions. Despite its conceptual directness, this recursively constructed collection of solutions appears convoluted and difficult to apply. Examples 2 and 3 suggest that in general, no easy characterization of the recursive solution might exist, at least none amenable to easy applicability. However, under the assumption that all feasible payoff sets are *convex*, Theorem 2 achieves a significant simplification, arguing that the resulting solution is equivalent to the maximizers of (1), but taken over all payoff allocations that are unblocked using the *unconstrained* Nash outcome for each coalition. That is, for each coalition, we compute its Nash bargaining solution relative to just its individual disagreement points and no other threats. This is not a conceptually accurate computation by any means. After all, a block using the unconstrained Nash solution is not, in general, credible. Nevertheless, this easy though artificial problem delivers the "correct" recursive answer. Theorem 1, which shows that coalitional constraints appear in the conventional form discussed earlier, is crucial to achieving this characterization.

[3]In noncooperative game theory, that sort of consistency is, of course, implicit in the very notion of subgame perfection. Coalitional analogues appear there too, as in the work of Bernheim, Peleg, and Whinston (1987), Bernheim and Ray (1989) and Farrell and Maskin (1989). In cooperative game theory, the same idea appears in the credible core (Ray 1989) in solution concepts that rely on constrained welfare solutions (Dutta and Ray 1989, 1991), and less directly in the concept of reduced games (Aumann and Maschler 1985, Peleg 1986).

Our final results pertain to transferable utility games, in which (after the normalization of intercept and slope) payoffs can be transferred 1-1 across players in every coalition. This setting yields a close connection between the coalitional solution and the *egalitarian solution* introduced in Dutta and Ray (1989, 1991). Their notion of "constrained egalitarianism" combines a commitment to egalitarianism 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 (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. Dutta and Ray restrict their attention to internal blocking to develop their notion of *constrained egalitarianiasm*, and so employ the same recursive structure as we do here.

Theorem 3 asserts that the internally consistent Nash solution must be a subset of the set of egalitarian solutions. At one level, this is intuitive: with symmetric bargaining power, the Nash product favors equality of payoffs, though the solution is constrained by coalitional threats. It therefore stands to reason that there should be a close connection between the internally consistent Nash solution and constrained egalitarianism; namely, that the former should be housed within the latter. The nonintuitive part comes from the fact that subset inclusion lower down in the recursion expands the unblocked set, opening up the possibility for Nash solutions at higher levels to lose all connections with the constrained egalitarian set. That those connections are in fact *not* lost is a crucial implication of Theorem 2.

Theorem 3 also provides conditions under which the internally consistent Nash bargaining solution is nonempty. It turns out that this is true of all superadditive games: those for which the worth of any coalition exceeds the total worth of any partition of it. This is a significantly lighter condition than one that guarantees the non-emptiness of the core, which is what we would have contend with in the absence of any consistency restriction on the behavior of blocking coalitions. We end with some remarks on the uniqueness of the Nash solution. This is a deep and interesting question on which we do not make full progress — the unblocked sets are typically nonconvex, after all — but sufficient conditions can be provided for the internally consistent Nash solution set to be a singleton.

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 or strong. The reader must judge whether the transparent and simple nature of the expansion assumption is illuminating in this sense, while at the same time its implication (jointly with the other standard axioms) is striking in the asymmetry it generates across different constraints even as it retains the simplicity of Nash's original characterization.

2. A Characterization of Nash Bargaining Under Coalitional Threats

2.1. **Notation.** $N = \{1, \ldots, n\}$ is the set of players, or the grand coalition. A coalition S is any nonempty subset of N, and a subcoalition is any strict subset of N. For any coalition S, $\mathbb{R}^S$ denotes $|S|$-dimensional Euclidean space with coordinates indexed by the elements of S. For any $x \in \mathbb{R}^S$, $\|x\| \equiv \sum_{i \in S} |x_i|$, and for $T \subset S$, 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. For any x and $\lambda \in \mathbb{R}^m$, write $\lambda \otimes x \equiv (\lambda_1 x_1, \ldots, \lambda_m x_m)$. Given $\tau = (\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$, an *affine transform* of $z \in \mathbb{R}^m$ with respect to τ is $\alpha + \beta \otimes z$. An affine transform of a set $Z \subseteq \mathbb{R}^S$ with respect to τ is $\{z' \in \mathbb{R}^S | z' = \alpha + \beta \otimes z$ for some $z \in Z\}$.

2.2. **Coalitional Threats.** A *game with coalitional threats* is a triple $G = (F, \boldsymbol{\Theta}, d)$, where $F \subset \mathbb{R}^N$ is a set of feasible payoffs for the grand coalition, $\boldsymbol{\Theta} = \{\Theta(S)_{S \subset N}\}$ are sets of *threats* in $\mathbb{R}^S$ for each subcoalition S, and $d \in \mathbb{R}^N$ is a vector of disagreement payoffs. Later, we endogenize $\Theta(S)$, but for now we regard these threats as exogenously given.

The disagreement payoff d is to be interpreted as some status quo payoff when no arrangement is in place in *any* subcoalition. Compare it with individual threat payoffs in $\Theta(\{i\})$ when the coalition $\{i\}$ has formed, the best of which are presumably no smaller than d_i, though the door is left open for d_i to possibly equal the individual threat payoff.

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

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

This definition clarifies that a "threat" only refers to the blocking allocation of a coalition, and has no additional connotations for the payoffs of the complementary coalition.

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

[Dom 1] For each subcoalition S, $\Theta(S)$ is nonempty and compact with $z \geqslant d_S$ for every $z \in \Theta(S)$. In particular $\zeta_i \equiv \max_{x \in \Theta(\{i\})} u_i(x)$ is well defined, with $\zeta_i \geqslant d_i$ for every i.

[Dom 2] F is nonempty and compact, and contains some $x \gg d$.

[Dom 1] asks that individual threats be no smaller than disagreement payoffs, which is presumably true almost by definition: every agent should be able to enjoy *at least* her disagreement payoff when she "forms her own coalition." [Dom 2] asks that feasible payoffs for the grand coalition include some allocation that strictly dominates individual disagreement payoffs. Both conditions also add on the technical restriction of compactness.[4]

Coalition S is *ineffective* if $\Theta(S) = \{d_S\}$, and *effective* otherwise. Let $\mathbf{\Theta}^0$ denote threats when all subcoalitions are ineffective and let $\mathcal{G}^0 \subset \mathcal{G}$ contain all games $G = (F, \mathbf{\Theta}^0, d)$. Each such game is a standard bargaining problem without coalitional threats. The allowance for nontrivial threat constellations $\mathbf{\Theta} \neq \mathbf{\Theta}^0$ is our central point of departure.

2.3. **Coalitional Nash Solution.** A *solution* σ assigns to every $G \in \mathcal{G}$ a nonempty subset $\sigma(G)$ of $U(G)$ whenever $U(G) \neq \emptyset$, and the empty set otherwise. Unlike Nash, our 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 therefore to the possibility of multiple outcomes. From this perspective, our approach benefits from the literature on Nash bargaining over nonconvex sets; see especially Kaneko (1980).[5] We now impose axioms on solutions, beginning with:

[**Par**] $\sigma(G) \subseteq U(G)$ is Pareto optimal in the unblocked set $U(G)$ (and therefore in F).[6]

Next, we adopt the affine invariance axiom. In Nash's original conception, payoffs are generated from outcomes, including 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.[7] The Nash invariance axiom asserts that:

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

Our next assumption is placed on bargaining problems in which no coalition has power:

[**Sym**] Suppose that every subcoalition is ineffective and $d_i = d_j$ for all $i, j \in N$. If, for some permutation π 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 $x \in \sigma(G)$.

[4]Comprehensiveness or free disposal is a standard property of payoff sets, but its compatibility with compactness is easily restored by noting that no payoff below the disagreement points will ever be relevant.

[5]Zhou (1997) and Serrano and Shimomura (1998) also study Nash bargaining with nonconvexities.

[6]An allocation $x \in A$ is Pareto-optimal in A if there does not exist $x' \in A$ such that $x' \gg x$.

[7]Formally, $G' = (F', \mathbf{\Theta}', d')$ is an affine transform of $G = (F, \mathbf{\Theta}, d)$, with respect to τ, if F' is an affine transform of F with respect to τ, $\Theta'(S)$ is an affine transform of $\Theta(S)$ with respect to τ_S for every S, and d' is an affine transform of d with respect to τ.

For two-player games, this is just the standard symmetry assumption of Nash.

Next, we suitably modify Nash's independence of irrelevant alternatives to allow for a multi-valued solution. We adapt the version of this axiom by Kaneko (1980) to our framework in a minimal way. Our adaption allows for changes in the feasible set F but leaves coalitional threats $\mathbf{\Theta}$ unchanged.

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

We follow Kaneko (1980) again in assuming upper hemicontinuity of the solution.

[**UHC**] Let G^k be a sequence such that $(F^k, \mathbf{\Theta}^k, d^k)$ converges in the (product) Hausdorff metric to $G = (F, \mathbf{\Theta}, d)$.[8] 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 in Shapley and Shubik (1974).

The above axioms are all standard conditions, mildly adapted to the coalitional setting. The next condition is new, though it is automatically implied in games without coalitional threats. Say that $\lambda \in \mathbb{R}^N_+$ is an *expansion* if $\lambda_i > 1$ for all $i \in N$.

[**Exp**] Suppose that $G = (F, \mathbf{\Theta}, d)$ 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 expansion λ such that $\lambda \otimes x \in \sigma(\{\lambda \otimes x, \lambda \otimes y\}, \mathbf{\Theta}, d)$

[Exp] states that if x is chosen from two unblocked alternatives x and y, then there is *some* expansion λ for which $\lambda \otimes x$ will is chosen from the (still unblocked) set $\{\lambda \otimes x, \lambda \otimes y\}$. Notice that this assumption is *not* implied by scale invariance because the threats are not scaled up or down. In Section 2.4, we discuss [Exp] and the role played by it.

Theorem 1. *A solution* $\sigma(G)$ *satisfies axioms* [Par], [Inv], [Sym], [IIA], [UHC] *and* [Exp] *for every game* $G \in \mathcal{G}$ *if and only if it maximizes the product of the payoffs in excess of the individual disagreement payoffs within the set of unblocked allocations:*[9]

$$\sigma(G) = \arg\max_{x \in U(G)} \prod_{j \in N} [x_j - d_j], \tag{4}$$

---

[8]More precisely, the sequence in either of the first two components *either* contains only empty sets after some finite index, and the presumed limit is also an empty set, or the sequence contains only nonempty sets after a finite index, the presumed limit is also nonempty, and convergence occurs in the Hausdorff metric.

[9]Recall that $U(G) = \{x \in F | x \text{ is not blocked by any threat } (S, y)\}$.

and we shall refer to it as the coalitional Nash solution.[10]

2.4. **Discussion.** Within a sub-domain with all coalitions ineffective, our proof of Theorem 1 follows (but generalizes) the argument in Kaneko (1980) by defining a preference ordering over payoff allocations in bargaining problems, and then using IIA, affine invariance and symmetry to establish that such an ordering must be represented by the symmetric Nash product defined on payoffs net of disagreement. Other axioms in Kaneko (1980) can be dropped for free; see proof for details. Axiom [Exp] then constructs a bridge from this ordering to a corresponding ordering for games with coalitional threats.

Of course, we could have *assumed* that unblocked sets in our sense are equivalent to feasible sets in the sense of Nash. That would be tantamount to admitting from the outset that coalitional threats are no different from any other constraint on feasibility, and our generalization of Kaneko's arguments would apply with no need for Axiom [Exp]. But our analysis does not assume this, allowing (at least in principle) for coalitional threats to be distinct from a mere delineator of feasibility. The fact that [Exp] then forces the same result uncovers a more basic property of the coalitional Nash solution, to which we now turn.

Recall that Nash's original theorem characterizes a solution as maximizing the product of payoff *excesses* over disagreement points:

$$\sigma(F,d) = \arg\max_{\mathbf{x}\in F} \{(x_1-d_1)(x_2-d_2) | x \geqslant d\} \tag{5}$$

for every bargaining problem (F,d), where F is compact and convex and d is the disagreement payoff. The subtraction of disagreement points emerges from the symmetry and invariance axioms, and our coalitional solution, suitably specialized to two-person bargaining, yields the same formula. Alternatives such as

$$\hat{\sigma}(F,d) = \arg\max_{\mathbf{x}\in F} \{x_1x_2 | x \geqslant d\}$$

are killed off by the invariance axiom. However, a parallel solution such as

$$\tilde{\sigma}(F,\zeta) = \arg\max_{x\in F} \{x_1x_2 | x \geqslant \zeta\}\,. \tag{6}$$

where ζ is now the vector of individual threat points, is also implied by Theorem 1 if we allow for disagreement and threats to differ, and normalize the former to zero using

[10]Compte and Jehiel (2010) define a similar notion by taking the unblocked set to be the core of a characteristic function game. This will become a meaningful distinction only when, in Section 3, we make coalitional threats endogenous. Our approaches, however, are complementary in that we are interested in an axiomatic characterization of (4) while they derive it as an equilibrium of a dynamic game.

invariance. *Threats* are not "subtracted" from payoffs, and are binding in the standard sense of the term — indeed, the extension of (6) to all coalitions yields precisely (4).

In a dynamic bargaining model, Binmore, Shaked, and Sutton (1989) and Compte and Jehiel (2010) draw a distinction between d and ζ.[11] As already discussed, Binmore, Shaked, and Sutton (1989) view individual threats as outside options that players can access only when they abandon negotiations altogether. In similar vein, Compte and Jehiel (2010) effectively separate d from ζ, viewing d as a per-period return that accrues when bargaining is still ongoing.[12] With d and ζ thus separated and the former normalized to zero, Compte and Jehiel (2010) connect to the equilibria of an N-player model of coalitional bargaining. They show that an efficient stationary equilibrium must maximize the product of the players' payoffs across all the payoffs unblocked by any coalition:[13]

$$x^* = \arg\max_{x \in \text{Core}} \prod_{j \in N} x_j, \tag{7}$$

which is the solution we obtain axiomatically when disagreement payoffs are separated from individual threats, reducing to (6) in the two-person case. In short, our approach allows for both (5) and (6), depending on whether or not we insist on $d = \zeta$.

Indeed, in a narrower domain in which ζ is restricted to always equal d, we would still be able to prove Theorem 1, despite the loss of one degree of freedom across d and ζ. (That degree of freedom is never used in the proof.) In that domain ζ would be subtracted from payoffs. But there is no parallel latitude for *nonsingleton* threats, which are binding only in the conventional sense. That stark difference arises from Axiom [Exp], which states that a "revealed preference" for x over y is maintained for some expansion of both x and y.

As already noted, [Exp] automatically follows from affine invariance for all "pure bargaining" problems, in which coalitions are ineffective. Given our continuity axiom [UHC], [Exp] is also automatically satisfied for all problems with or without coalitional threats

[11] A similar approach to ζ also underlies Chatterjee, Dutta, Ray, and Sengupta (1993) and Okada (1996).

[12] Binmore, Rubinstein, and Wolinsky (1986) adopt a different perspective, also in the context of a noncooperative bargaining game. They allow for some exogenous low-probability event that causes bargaining to break down altogether, leaving players stranded with d. In the equilibrium of that game, it turns out those payoffs must be subtracted from bargaining payoffs and the Nash product applied to the net amounts, whereas in the absence of breakdown, it is the gross bargaining payoff that enters into the Nash product. While this is an interpretation that we choose not to pursue here, it has potentially interesting implications in a coalitional setting. Presumably, the probability distribution over coalitional payoff allocations following an exogenous "breakdown event" would pin down the particular form assumed by the Nash product.

[13] Chatterjee, Dutta, Ray, and Sengupta (1993) connect bargaining in coalitional games to the egalitiarian solution of Dutta and Ray (1989), which maximizes the product in (7). We return to this connection below.

when just one of the two outcomes in F is in the solution, but not the other. So the real bite of [Exp] pertains to the knife-edge case in which *both* outcomes are chosen in a two-outcome problem, asserting that that indifference is preserved for some expansion of the outcomes. Because such situations of indifference are relatively “rare” or “non-generic”, that makes [Exp] an even weaker restriction, and yet its implications are substantial.

To illustrate how [Exp] works, we drop it, and construct a new solution which satisfies all our other axioms. (In so doing, we establish the independence of [Exp]. The Supplementary Notes show that the other axioms are similarly independent.) Formally, we construct a solution $\phi(G)$ for every game $G \in \mathcal{G}$ that satisfies [Par], [Inv], [Sym], [IIA] and [UHC].

By [Inv], there is no loss of generality in assuming $d = 0$. For every $S \subset N$, let $\tilde{\Theta}(S)$ denote the Pareto frontier of $\Theta(S)$ and let $a_i(S)$ be the mean payoff to i in S given the uniform distribution over $\tilde{\Theta}(S)$. For every $i \in N$, let $S(i)$ be the collection of all subcoalitions of N that contain i. Write the average of the average payoff to i as $a_i \equiv [\sum_{S \in S(i)} a_i(S)]/m$, where $m = 2^{n-1} - 1$ is the number of coalitions in $S(i)$. Now define a solution by

$$\phi(G) = \underset{x \in U(G)}{\arg\max} \prod_{j \in N} [x_j - a_j].$$

It is easy to see that ϕ satisfies [Par], [Inv], [Sym], [IIA] and [UHC]. However, there exist games for which ϕ fails to satisfy [Exp]. The following example makes this clear.

Example 1. Consider a three-person game with $d = (0, 0, 0)$, in which all coalitions except N and $\{1, 2\}$ are ineffective. The threat set of coalition $\{1, 2\}$ is given by:

$$\Theta(\{1, 2\}) = \{(x_1, x_2) \in \mathbb{R}^2_+ | x_1 + x_2 \leqslant 2\},$$

and the grand coalition is equipped with the feasible set

$$F = \{y, z\}, \text{ where } y = (3, 3, 1) \text{ and } z = (2, 2, 4).$$

Note that both y and z are unblocked. Observe that $(a_1, a_2, a_3) = (1, 1, 0)$, so that:

$$\phi(G) = \underset{x \in \{y,z\}}{\arg\max} (x_1 - 1)(x_2 - 1)x_3 = \{y, z\}, \tag{8}$$

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 ϕ 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 $G' = (\{\lambda \otimes y, \lambda \otimes z\}, \Theta, 0)$,

$$\phi(G') = \underset{x \in \{\lambda \otimes y, \lambda \otimes z\}}{\arg\max} (x_1 - 1)(x_2 - 1)x_3 = \{\lambda \otimes z\}, \tag{9}$$

which, given (8), contradicts [Exp]. To prove (9), let $d_i \equiv \lambda_i - 1$ for each i. Note that

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

Letting the maximand in ϕ be denoted by $f(x) = (x_1 - 1)(x_2 - 1)x_3$, we have:

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

and

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

Clearly, $f(\lambda \otimes z) > f(\lambda \otimes y)$, which implies (9). ◇

2.5. **Proof of Theorem 1.** 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 through the proof. Using [Inv], [UHC] and [Exp], we will first reduce any two-allocation problem with threats to a two-allocation bargaining problem, and then match the analysis to a generalized version of the argument in the proof of Theorem 2, Kaneko (1980).

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 $\mathbf{\Theta}' = \alpha \otimes \mathbf{\Theta}$ 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 $\lambda \otimes x \in \sigma(\{\lambda \otimes x, \lambda \otimes y\}, \bar{\mathbf{\Theta}}, 0)$. Let $\alpha = (1/\lambda_1, \ldots, 1/\lambda_n)$. Then $1 \gg \alpha \geqslant 0$ (in fact $\alpha \gg 0$), and defining $\mathbf{\Theta} = \alpha \otimes \bar{\mathbf{\Theta}}$, 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 $\precsim$.[14] For any totally ordered subset $\mathcal{T}^c$ of $\mathcal{T}$, define

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

If there is $\mathbf{\Theta}^* \in \mathcal{T}^c$ of the form $\mathbf{\Theta}^* = \alpha^* \otimes \bar{\mathbf{\Theta}}$ 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 is a sequence $\{\mathbf{\Theta}^k\}$ in $\mathcal{T}^c$ with $\mathbf{\Theta}^k = \alpha^k \otimes \bar{\mathbf{\Theta}}$, with $1 \gg \alpha^k \geqslant 0$ for every k, and $\|\alpha^k\| \to a$. Let $(\mathbf{\Theta}^*, \alpha^*)$ be any limit point of $\{\mathbf{\Theta}^k, \alpha^k\}$, the first component under the product

[14] Every element $\mathbf{\Theta}$ of $\mathcal{T}$ has the property that $\mathbf{\Theta} = \alpha \otimes \bar{\mathbf{\Theta}}$ for some α 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.

Hausdorff metric and the second in the standard sense.[15] Then $\boldsymbol{\Theta}^* = \alpha^* \otimes \bar{\boldsymbol{\Theta}}$. 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 the following:

(10) If x and y are unblocked by some effective $\boldsymbol{\Theta}$ and $x \in \sigma(\{x, y\}, \Theta, 0)$, then $x \in \sigma(\{x, y\}, \boldsymbol{\Theta}^0, 0)$.

The remaining argument lines up (10) with Theorem 2 of Kaneko (1980) for pure bargaining problems. It is easy to see that [Inv], [Sym], [IIA] and [UHC] corresponds to axioms N.2-N.5 of Kaneko (1980) in the smaller domain $\mathcal{G}^0$ of bargaining problems. The Supplementary Notes show that [Par], while weaker than Kaneko's axiom N.1, suffices for the proof of his Theorem 2. The Notes also show that Kaneko's "strict individual rationality" can be dropped at no cost. Additionally, Kaneko's Theorem 2 also assumes a dimensionality condition which implies in particular that the domain includes all sets consisting of no more than three points. It can be checked that his proof only relies on this implication and not the dimensionality condition itself. Moreover, the implication follows from [Dom 1] and [Dom 2], which implies that all bargaining problems with compact feasible sets are in $\mathcal{G}^0$. We can now appeal to Theorem 2 of Kaneko to assert that for any $x, y \in \mathbb{R}^N_+$,

$$\text{If } x \in \sigma(\{x, y\}, \boldsymbol{\Theta}^0, 0), \text{ then } \prod_{j \in N} x_j \geqslant \prod_{j=1}^{n} y_j.$$

Recalling (10), this implies that

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

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 (11) that

$$\sigma(G) = \underset{x \in U(G)}{\arg\max} \prod_{j \in N} x_j.$$

When d is not necessarily 0, [Inv] can be invoked to yield (4), and the proof is complete. ■

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

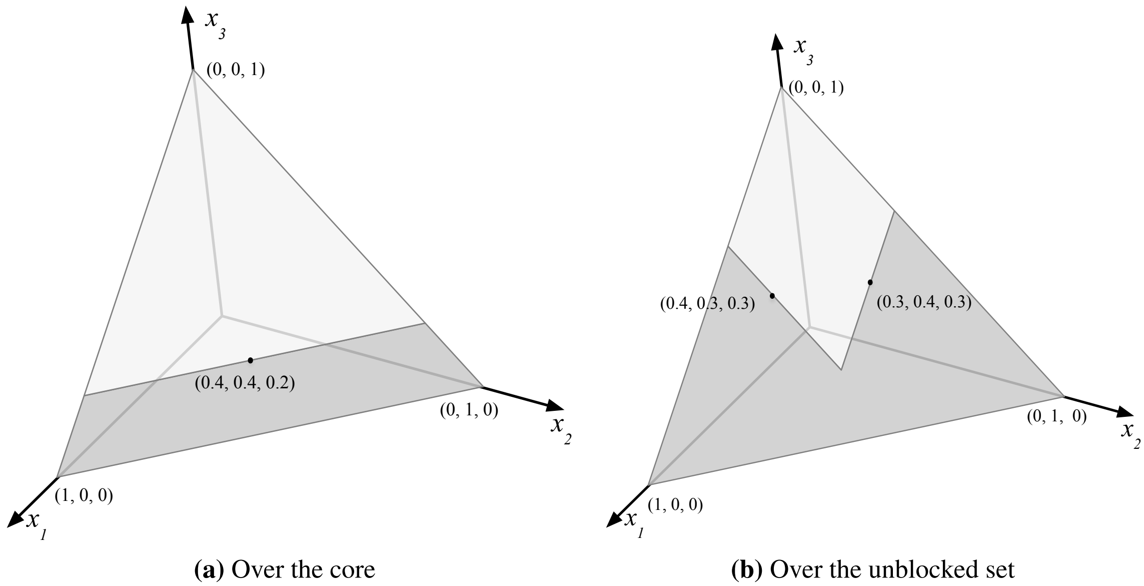

Figure 1. Maximizing payoff product with d normalized to 0. The darker area in each panel are the allocations that survive threats.

3. Internally Consistent Coalitional Threats

3.1. **Internal Consistency.** We now study threat sets in more detail, under the restriction that all threats to a coalition come from its own subcoalitions. We begin with a simple example to illustrate our recursive approach to endogenously constructing threat sets. The example is an instance of a *transferable utility* (TU) game: there is an affine transform of payoffs such that $d = 0$ and for each S, there is a scalar $v(S)$ so that its feasible set of allocations is given by $F(S) = \{x \in \mathbb{R}^S | \sum_{i \in S} v(\{i\}) \leqslant \sum_{i \in S} x_i \leqslant v(S)\}$.[16]

Example 2. Consider a three-player TU game with player set $N = \{1, 2, 3\}$. Suppose $v(N) = 1$, $v(\{1, 2\}) = 0.8$ and $v(S) = 0$ for all other S. In particular, we normalize both d and ζ to 0. Now, if $\{1, 2\}$ can use any feasible allocation as a threat, the unblocked set for the grand coalition — or the core of the game — is the set:

$$C = \{x \in \mathbb{R}^3_+ | x_1 + x_2 + x_3 = 1; x_1 + x_2 \geqslant 0.8\}$$

[16]This isn't standard, in that the usual definition allows for unlimited free disposal. We choke that off by requiring every allocation in S to be individually rational relative to the outside options $v(\{i\})$. It makes no difference as allocations failing that condition are completely irrelevant.

This is the shaded portion of the unit simplex (the efficient frontier) in Figure 1(a). If we maximize the product of payoffs over the core, the solution is $(0.4, 0.4, 0.2)$. This is the Compte and Jehiel (2010) solution for this game.

But if the grand coalition is to be subject to our axioms, consistency demands that the same restrictions be imposed on coalition $\{1, 2\}$ as well. That raises questions about what the "same" restrictions mean. In this paper, we take the position that $\{1, 2\}$ is only threatened by *its* subcoalitions. That leads to the notion of *internal consistency*. In our example, invoking (4), the unique solution for $\{1, 2\}$ is given by $(0.4, 0.4)$. This is its only credible threat, so $\Theta(\{1, 2\}) = \{(0.4, 0.4)\}$, while of course $\Theta(S)$ is the zero vector for all other subcoalitions. So if we set $F = F(N)$, the unblocked set for the grand coalition N is given — with a slight abuse of earlier notation — by:

$$U(N) = \{x \in F | \max\{x_1, x_2\} \geqslant 0.4\},$$

which is the shaded area in Figure 1(b). Restricting the threat from $\{1, 2\}$ to be the Nash solution for $\{1, 2\}$ results in an unblocked set for the grand coalition that is larger than its core. After all, to negate the putative threat, it is enough to offer 0.4 to just *one* of the players in $\{1, 2\}$. Note that the unblocked set is not convex. Maximizing the product of payoffs on it yields, again with a little abuse of notation:

$$\sigma(N) = \{(0.4, 0.3, 0.3), (0.3, 0.4, 0.3)\}.$$

If $N = \{1, 2, 3\}$ were embedded in turn into a larger game, there would be two threats from $\{1, 2, 3\}$, namely $(0.4, 0.3, 0.3)$ and $(0.3, 0.4, 0.3)$, and $\sigma(N)$ would be renamed to $\Theta(N)$. That suggests a recursive construction of threat sets from smaller to larger coalitions. $\diamond$

The analytical advantage of internal consistency is precisely that it is amenable to recursive treatment. A more general consistency notion, which allows for subcoalitions of a coalition to combine with other players who are not in the coalition, is beyond the scope of the current exercise, but will form the subject of a forthcoming paper. The difficulty is that there is no starting point for a recursion, and a fixed point argument along with other conceptual considerations intrude into the story; see Ray and Vohra (2015, 2019) for related exercises. One could, of course, close the construction by attaching arbitrary division methods at certain nodes, but it could be argued that this violates the spirit of consistency.[17] In Example

[17]The solution of Serrano and Shimomura (1998), which attempts to incorporate "external threats," illustrates the difficulty. In their exercise, an efficient payoff profile x across all players in N precipitates a particular two-player game on any pair $\{i, j\}$. The induced feasible set contains all allocations in $F(N)$ 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

2 that we develop into a fuller theory below, there is no discrepancy: coalitional threats are determined by the same considerations that are applied to the grand coalition. It is for the sake of that precise concordance that we are less ambitious in this paper.

3.2. **Recursive Definition.** There is, in the background, a disagreement payoff vector d. Consider a characteristic function $\mathbf{F} = \{F(S)_{S \subseteq N}\}$ where N is the grand coalition and for every coalition S, $F(S)$ is a non-empty, compact set of feasible payoffs, with $x \geqslant d_S$ for all $x \in F(S)$. To rule out uninteresting cases, we also presume that $x \gg d_S$ for some $x \in F(S)$ whenever $F(S) \neq \{d_S\}$. Observe that $\zeta_i = \max F(\{i\})$ is player i's individual threat. Whether or not ζ is automatically linked to d is a matter to be settled by the application at hand. It will make no difference to the analysis, as long as we remember the two interpretations discussed in Section 2.4. Henceforth we normalize $d = 0$.

We will recursively pin down endogenous threat sets $\mathbf{\Theta}^*(S)$ (or equivalently, solutions $\sigma^*(S)$) for every subcoalition S. Note the slight abuse of notation as these are only indexed by S and not the entire game G as we did before.) For singleton sets $\{i\}$, set $\Theta^*(\{i\}) = \Theta^*(\{i\}$. Now suppose that for some coalition S we have already determined the sets $\Theta^*(T)$ for every $T \subset S$, and that each such $\Theta^*(T)$ is nonempty and compact. We can now use these threats to construct the unblocked set for coalition S:

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

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:

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

which is the internally consistent solution for the game restricted to S and its subcoalitions.

Now the recursion, indexed by cardinality, can continue to still larger sets. When concluded at the level of the grand coalition, we can define the *internally consistent coalitional Nash*

---

joining some coalition T that doesn't include j, once again paying off each $k \in T$ to the tune of precisely x_k. 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 $\{i, j\}$. (There are other conditions in the Serrano-Shimomura axiomatization, including *converse consistency*, which asks that any x with the above consistency property must be a solution to the original game.) In TU games, this approach characterizes the pre-kernel of Davis and Maschler (1965). Note that when i approaches some coalition (without j in it), the other members of such a coalition are not similarly able to make use of their outside options to negotiate; they are assumed to simply acquiesce to be paid according to x. In Example 2, the pre-kernel consists of the unique allocation $(0.45, 0.45, 0.1)$, different from both the coalitional Nash solution and the internally consistent Nash solution.

solution, or *internally consistent Nash solution* for short, as

$$\sigma^*(N) = \underset{x \in U^*(N)}{\arg\max} \prod_{j \in N} x_j.$$

3.3. **A Characterization.** From an applied perspective, this is an unwieldy concept which appears to necessitate recursion not just to define but to fully describe. At every stage, we would need to know the internally consistent Nash solution for every subcoalition of a given coalition before unblocked sets for that coalition — or its internally consistent Nash solution — can be calculated. The problem is not unlike the computation of subgame perfect equilibria along game trees. Two alternatives present themselves that have at least the advantage of easier computability. One is the solution that Compte and Jehiel (2010) obtain in a noncooperative bargaining model, which entails the maximization of the Nash product over the core of the game. This is obviously simpler but asymmetrically ascribes to subcoalitions the total freedom to choose any allocation it pleases, while restricting the grand coalition. It is obvious that the "inconsistent solution" thus obtained will generally be different from our recursive solution; see, for instance, Example 2.

A second, equally naive, and even simpler alternative is to presume that each coalition blocks with its *standard (unconstrained)* Nash solution, one that ignores all coalitional constraints, except the disagreement points. We cannot condone such naïveté either. It stretches credibility in another direction, in that coalitions anticipate that they, too, will engage in bargaining but fail at the same time to acknowledge that they will similarly face constraints from their subcoalitions. Once again, the solution thus obtained will generally be different from our recursive solution. Our next example illustrates this,[18] but the reader interested in the broader argument can skip it for now and return later for details.

Example 3. Let $N = \{1, 2, 3, 4\}$ with $d_i = \zeta_i = 0$ for all i, and:

(i) $F(N) = \{x \in \mathbb{R}^4_+ | x_1 + x_2 + x_3 + x_4 \leqslant v\}$.

(ii) $F(\{1, 2, 3\})$ has just two payoff allocations $(1, 1, 1)$ and (a, a, b), where $(a, b) \gg 0$.

(iii) $F(\{1, 2\})$ consists of the single payoff allocation (c, c), with $c > 0$.

(iv) $F(S) = \{0_S\}$ for every other coalition S.

[18]We could have used a three-player game to make the same point, but then ζ would need to be distinct from d. We prefer to not rely on the possibility that $\zeta \neq d$.

(v) $a^2b < 1 < c < a < v$, and $b > \frac{v-c}{3} > 0$.[19]

The "naive" coalitional bargaining solution for N presumes that coalitions will block with their unconstrained Nash allocations. For coalition $\{1, 2\}$ this is just (c, c), and for coalition $\{1, 2, 3\}$ it is $(1, 1, 1)$, using the assumption that $a^2b < 1$. So, keeping in mind that $c > 1$, the unblocked set $U'(N)$ for N is the set of all allocations in $F(N)$ that are unblocked by the threat (c, c) from coalition $\{1, 2\}$. The naive solution for N is then easily seen to consist of the two allocations $(c, \frac{v-c}{3}, \frac{v-c}{3}, \frac{v-c}{3})$ and $(\frac{v-c}{3}, c, \frac{v-c}{3}, \frac{v-c}{3})$.[20] The *internally consistent* Nash solution for coalition $\{1, 2\}$ is still (c, c). But the solution for coalition $\{1, 2, 3\}$ is (a, a, b), now that the allocation $(1, 1, 1)$ is blocked by $\{1, 2\}$ with (c, c), and also because $a > c$ by (v). Returning, then, to the grand coalition, its internally consistent unblocked set $U^*(N)$ must consist of allocations unblocked by (a, a, b), and this set is clearly nonempty because $b < 1$. The naive solution $(c, \frac{v-c}{3}, \frac{v-c}{3}, \frac{v-c}{3})$ is blocked by the recursively defined threat from $\{1, 2, 3\}$; after all, $a > c$ and $b > \frac{v-c}{3}$ by assumption. The internally consistent solution is therefore different from the naive solution.[21] $\diamond$

Example 3 suggests that a simple, non-recursive description of the internally consistent Nash solution may not be possible. But if the feasible set $F(S)$ for every coalition is convex, as it indeed would be in the world of Nash's "anticipations" with lotteries, a straightforward characterization is available. As in Example 3, define the *standard Nash solution* $\Psi(S)$ for coalition S by ignoring all coalitional constraints:

$$\Psi(S) = \underset{x \in F(S)}{\arg\max} \prod_{j \in S} x_j,$$

where we recall once more that d has been normalized to zero. If $F(S)$ is convex, $\Psi(S)$ consists of a single allocation; call it $\psi(S)$. Let

$$U'(S) \equiv \{x \in F(S) | x \text{ is unblocked by any } (T, \psi(T)) \text{ with } T \subset S\},$$

be the unblocked set for S using the artificial disagreement points from the unconstrained Nash bargaining solutions for each subcoalition. Note that no recursive argument is required to define $U'(S)$.

[19] For instance, $v = 3.05$, $a = 1.18$, $b = 0.66$ and $c = 1.1$, satisfy all the restrictions of the example, and additionally the game can be seen to be superadditive under these specifications.

[20] Because $a > 1$ and $a^2b < 1$, we have $b < 1$. Therefore $c > 1 > b > \frac{v-c}{3}$, and this assures us that the naive solution is indeed as stated in the main text.

[21] The exact form of $\sigma^*(N)$ will depend on the values of the parameters. If $v = 3.05$, $a = 1.18$, $b = 0.66$ and $c = 1.1$, it can be checked that $\sigma^*(N)$ consists of the two allocations $(1.1, 0.645, 0.66, 0.645)$ and $(0.645, 1.1, 0.66, 0.645)$. The naive solution consists of $(1.1, 0.65, 0.65, 0.65)$ and $(0.65, 1.1, 0.65, 0.65)$.

Theorem 2. *Assume that $F(S)$ is compact and convex for every coalition S. Then $U^*(S) = U'(S)$, and so when this common set is nonempty, the internally consistent Nash solution for any coalition need only guard against the threats posed by the standard Nash solutions of its subcoalitions. That is,*

$$\sigma^*(S) = \arg\max_{x \in U'(S)} \prod_{j \in S} x_j.$$

Specifically, under the assumptions of the Theorem, Example 3 does not apply.

The theorem achieves a significant simplification. Instead of finding the coalitional bargaining solution immune to recursively defined threat sets, which could be quite demanding computationally, Theorem 2 asserts that the same solution can be found by using artificial threat sets for each subcoalition, which consist only of a single point — the unconstrained Nash bargaining solution for that subcoalition.

The careful reader will have noticed that the results in this Section apply 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.

An alternative line of generalization comes from dropping the convexity assumption on feasible sets. A set $A \subseteq \mathbb{R}^m$ is *subconvex* if for every $x, y \in A$ and $t \in [0,1]$, there is $z \in A$ such that $z \geqslant tx + (1-t)y$, and a set $A \subseteq \mathbb{R}^m_+$ is *log subconvex* if $\ln A \equiv \{z \in \mathbb{R}^m | (z_1, \ldots, z_m) = (\ln x_1, \ldots, \ln x_m) \text{ for some } x \in A\}$ is subconvex.[22] Log subconvexity is a weak property that could apply to connected sets as well as to sets with isolated elements,[23] and moreover, *a convex set in $\mathbb{R}^m_+$ must be log subconvex.*[24] The Supplementary Notes show that Theorem 2 holds under the weaker assumption that all feasible sets are log subconvex, though the unconstrained Nash solution could now be multi-valued.

3.4. **Proof of Theorem 2.** The proof combines Claims 1–3 below.

Claim 1. For any coalition S, $U^*(S) \subseteq U'(S)$.

Proof. Suppose $x \in U^*(S)$ but $x \notin U'(S)$. 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

[22] Note that log subconvexity is only defined for subsets A of $\mathbb{R}^m_+$, so that $\ln x$ is well-defined in the extended reals for all $x \in A$. The vector ordering "$\geqslant$" is then applied to the extended reals in the obvious way.

[23] However, in Example 3, $F(\{1,2,3\})$ is not log subconvex.

[24] Let A be convex. Then for every $x, y \in A$ and $t \in (0,1)$, $z \equiv tx + (1-t)y \in A$. But we know that $\ln(z) \geqslant t \ln x + (1-t) \ln y$, which proves that A is log subconvex.

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

$$y \geqslant x_H \tag{12}$$

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 (12), we must conclude that $\psi(W) \geqslant x_W$. ■

Claim 3. For any coalition S, $U'(S) \subseteq U^*(S)$.

Proof. Suppose there exists $x \in U'(S)$ but $x \notin U^*(S)$. Since $x \in F(S)$, this means that there is some coalition T and $y \in \Theta(T)$ such that $y \gg x_T$. Because $x \in U'(S)$, 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 $x \in U'(S)$. ■

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

4. Transferable Utility Games

4.1. **Constrained Egalitarianism.** The goal of this section is to draw a connection between the internally consistent Nash solution and the constrained egalitarian solution (Dutta and Ray 1989, 1991), in the context of transferable utility (TU) games. Recall that a game with coalitional threats is TU if there is an affine transform of payoffs such that $d = 0$ and for each S, there is a scalar $v(S)$ so that its feasible set of allocations is given by $F(S) = \{x \in \mathbb{R}^S | \sum_{i \in S} v(\{i\}) \leqslant \sum_{i \in S} x_i \leqslant v(S)\}$.

Dutta and Ray (1989, 1991) proposed a solution concept for such games that combines a commitment to egalitarianism with the pragmatics of coalitional participation. Their idea was to design "constrained norms" under which the a social ethic (egalitarianism, in their case) could be applied to the greatest degree possible, while remaining limited by individual and coalitional buy-ins. The following example illustrates their concept:

Example 4. Consider the three-player TU game from Example 2, with player set $N = \{1, 2, 3\}$, d normalized to zero, $v(N) = 1$, $v(\{12\}) = 0.8$ and $v(S) = 0$ for all other S.

A commitment to egalitarianism on the part of the coalition $\{12\}$ is not constrained by players 1 or 2 acting on their own accord. The latter would get either player 0, so coalition $\{12\}$ can credibly carry out their social preference for egalitarianism. If formed, that coalition would implement the allocation $\{0.4, 0.4\}$. The same commitment exists for the grand coalition N, but it 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 natural candidate for 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. As we saw in Example 2, these are also the two allocations in the internally consistent Nash solution. $\diamond$

Dutta and Ray (1989, 1991) generalized this idea to societies of arbitrary size and any specification of worths $\{v(S)\}$ across coalitions. Specifically, they argued that the grand coalition would need to choose the Lorenz-maximal elements from its set of unblocked allocations. In turn, because every coalition is also presumed to subscribe to the social ethic of egalitarianism, any credible block would also need to be egalitarian (for the coalition doing the blocking) in the same way asked of the grand coalition. Throughout, Dutta and Ray restrict their attention to internal blocking to develop their notion of *constrained egalitarianiasm*, and so employ the same recursive structure as we do here.

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 ordering is well known to agree with the ethics of egalitarianism (see, e.g., Kolm 1969, Dasgupta, Sen, and Starrett 1973 and Fields and Fei 1978), though it is partial. For any set of allocations A adding to the same total, let $L(A)$ be its set of *Lorenz-maximal* elements: those allocations in A that are not majorized by any other allocation in A.

Now we define solutions $E(S)$ 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, whenever this set is nonempty, set

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

This is the *constrained egalitarian solution* for coalition S.[25] Proceed in this manner until all coalitions are covered.

4.2. **Constrained Egalitarianism and the Internally Consistent Nash Solution.**

We now connect constrained egalitarianism to the internally consistent Nash solution. As always, we normalize individual disagreement points to zero.

Theorem 3. *In a TU game, the internally consistent Nash solution is a subset of the constrained egalitarian solution for every coalition S:*

$$\sigma^*(S) \subseteq E(S).$$

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

$$v(S) \geqslant \sum_{j=1}^{m} v(T_j) \text{ for all partitions } (T_1, \ldots, T_m) \text{ of } S. \tag{13}$$

Then for all S, $\sigma^*(S)$ 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 2 is crucial to the proof of this result, allowing us to sidestep a recursive argument. Such arguments generally fail for set-inclusion results. Under the recursive approach, the inductive hypothesis is that $\sigma^*(S) \subseteq E(S)$ for all coalitions of some size k or smaller. That *widens* the unblocked set for Nash bargaining for a larger coalition relative to the Dutta-Ray unblocked set, and prevents the inductive step from being completed. Theorem 2 avoids this line of reasoning altogether.

The existence of an internally consistent Nash solution is assured in all superadditive games (recall (13)), which is a very general class. That is because the ability of subcoalitions to

[25] Dutta and Ray actually use two different notions of blocking. Dutta and Ray (1989) define (T, y) to block $x \in F(S)$ if $y_T > x$, whereas (in line with the current paper) Dutta and Ray (1991) use the stronger blocking criterion $y_T \gg x$. The two induced solutions are different; see Dutta and Ray (1991). We use the latter here.

block an allocation is compromised by their anticipation that they, too, will seek a Nash bargain — and not be able to follow through with any arbitrary block. In contrast, a non-cooperative solution such as the one in Compte and Jehiel (2010) does not exhibit internal consistency; effectively, coalitional blocks are *unrestricted*. That limits the existence of a Nash bargain, even in TU games, to situations in which the game has a balancedness property, which is a far narrower class than mere superadditivity.[26] We do not view existence as a reason for choosing one concept over another, but the comparison is still of interest.

4.3. **Proof of Theorem 3.** In a TU game, the standard Nash solution for every S is the same as the *equal division solution*:[27]

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

To establish the first part of the theorem, begin by observing that for each S, the feasible set $F(S) = \{x \in \mathbb{R}^S | \sum_{i \in S} v(\{i\}) \leqslant \sum_{i \in S} x_i \leqslant v(S)\}$ is compact and convex. So Theorem 2 applies. Noting that the unconstrained Nash solution for any such T is equal to $e(T)$, we must conclude that

$$\Theta^*(S) = \sigma^*(S) = \arg\max_{x \in U'(S)} \prod_{j \in S} x_j. \tag{14}$$

where

$$U'(S) = \{x \in F(S) | x \text{ is unblocked by any } (T, y) \text{ with } T \subset S \text{ and } y = e(T)\} \tag{15}$$

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

$$E(S) = L(U'(S)). \tag{16}$$

when $U'(S) \neq \emptyset$. In the light of (14) and (16), 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 from Lemma 3 of Dutta and Ray (1991), who prove that under assumption (13), the set $U'(N)$ — see equation (15) with $S = N$ — is nonempty. It is obviously compact, so $\sigma^*(N) = \arg\max_{x \in U'(N)} \prod_{i=1}^k x_i$ must be nonempty. ■

[26] If a game is not TU, superadditivity does not guarantee that $\sigma^*(S)$ is nonempty, but balancedness does.

[27] Recall that the disagreement points are all normalized to zero.

Finally, one might ask when the internally consistent Nash solution is unique. 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 $S \subseteq N - \{i, j\}$ (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 1 (to Theorem 3 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 3 and Dutta and Ray (1991), Theorem 3. But 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.[28]

References

ATKINSON, A. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.

AUMANN, R. J. AND M. MASCHLER (1985): "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud," *Journal of Economic Theory*, 36, 195–213.

BERNHEIM, B., B. PELEG, AND M. D. WHINSTON (1987): "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory*, 42, 1–12.

BERNHEIM, B. D. AND D. RAY (1989): "Collective Dynamic Consistency in Repeated Games," *Games and Economic Behavior*, 1, 295–326.

BINMORE, K., A. RUBINSTEIN, AND A. WOLINSKY (1986): "The Nash Bargaining Solution in Economic Modelling," *The RAND Journal of Economics*, 176–188.

BINMORE, K., A. SHAKED, AND J. SUTTON (1989): "An Outside Option Experiment," *Quarterly Journal of Economics*, 104, 753–770.

CHATTERJEE, K., B. DUTTA, D. RAY, AND K. SENGUPTA (1993): "A Noncooperative Theory of Coalitional Bargaining," *The Review of Economic Studies*, 60, 463.

COMPTE, O. AND P. JEHIEL (2010): "The Coalitional Nash Bargaining Solution," *Econometrica*, 78, 1593–1623.

[28]The internally consistent 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.

DASGUPTA, P., A. SEN, AND D. STARRETT (1973): "Notes on the Measurement of Inequality," *Journal of Economic Theory*, 6, 180–187.

DAVIS, M. AND M. MASCHLER (1965): "The Kernel of a Cooperative Game," *Naval Research Logistics Quarterly*, 12, 223–259.

DUTTA, B. AND D. RAY (1989): "A Concept of Egalitarianism Under Participation Constraints," *Econometrica*, 57, 615–635.

——— (1991): "Constrained Egalitarian Allocations," *Games and Economic Behavior*, 3, 403–422.

FARRELL, J. AND E. MASKIN (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1, 327–360.

FIELDS, G. S. AND J. C. FEI (1978): "On Inequality Comparisons," *Econometrica*, 303–316.

KANEKO, M. (1980): "An Extension of the Nash Bargaining Problem and the Nash Social Welfare Function," *Theory and Decision*, 12, 135–148.

KOLM, S. C. (1969): "The Optimal Production of Social Justice," in *Public Economics*, ed. by M. J. and H. Guitton, Macmillan UK, 145–200.

MASCHLER, M. AND B. PELEG (1966): "A Characterization, Existence Proof and Dimension Bounds for the Kernel of a Game," *Pacific Journal of Mathematics*, 18, 289–328.

OKADA, A. (1996): "A Noncooperative Coalitional Bargaining Game with Random Proposers," *Games and Economic Behavior*, 16, 97–108.

PELEG, B. (1986): "On the Reduced Game Property and its Converse," *International Journal of Game Theory*, 15, 187–200.

RAY, D. (1989): "Credible Coalitions and the Core," *International Journal of Game Theory*, 18, 185–187.

RAY, D. AND R. VOHRA (2015): "The Farsighted Stable Set," *Econometrica*, 83, 977–1011.

——— (2019): "Maximality in the Farsighted Stable Set," *Econometrica*, 87, 1763–1779.

SERRANO, R. AND K. I. SHIMOMURA (1998): "Beyond Nash Bargaining Theory: The Nash Set," *Journal of Economic Theory*, 83, 286–307.

SHAPLEY, L. AND M. SHUBIK (1974): *Game Theory in Economics - Chapter 4: Preferences and Utility*, Rand Corporation Report R-904/4-NSF.

ZHOU, L. (1997): "The Nash Bargaining Theory with Non-Convex Problems," *Econometrica*, 65, 681–685.