Axiomatic bargaining

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

When two states stand on the brink of war, both sides usually understand — before the first shot is fired — that fighting will be costly, that some side will likely win and the other lose, and that the territory or policy stake at issue could in principle be divided between them. Fearon (1995) pressed the question this observation invites: if both sides could foresee the costly war and its eventual settlement, why didn't they reach the settlement directly and skip the war? The puzzle is real because the structural setup — two parties with conflicting preferences over a divisible stake, with a costly outside option both prefer to avoid — is the canonical setup of *bargaining theory*, and the structural question of what bargained outcome rational parties should reach is the question this handout's apparatus is designed to answer.

Bargaining theory shows up across political-economy modeling because the same structural setup recurs everywhere. A legislator and an executive negotiating over a budget allocation. A union and a firm negotiating over wages. A trade negotiation between two states. A coalition formation in a parliamentary government, with each member having conflicting preferences over the policy bundle. A constitutional convention dividing power between branches. In each case, several parties with conflicting preferences face a set of feasible alternatives and must select one, with each party able to walk away to a known outside option that all of them prefer to avoid. The structural question — which feasible alternative gets selected, given that each party has bargaining power and conflicting preferences — is the bargaining-theoretic question, and the welfare-function apparatus of the previous handout is exactly what supplies the answer.

Three structural insights organize the handout. First, a bargaining problem is formally a pair $(S, \boldsymbol{d})$ consisting of a set S of attainable utility profiles (the Pareto frontier of which is what gets bargained over) and a disagreement point $\boldsymbol{d}$ Pareto-dominated by at least one alternative (§2). Second, a *bargaining solution* is a rule that picks a Pareto-optimal point from each bargaining problem; the most-cited solution is the *Nash bargaining solution*, characterized axiomatically by Nash (1950) as the unique solution satisfying Pareto efficiency, scale invariance, symmetry, and independence of irrelevant alternatives (§3–§4). Third, alternative axiomatizations relax some of Nash's axioms in ways with substantive content for political-economy applications — most prominently Kalai and Smorodinsky (1975), which weakens IIA in favor of a monotonicity property (§5) — and the various solutions correspond to different ways of selecting weights in the welfare-function characterization of #28 (§6). The Fearon war-bargaining setup runs through the handout as the canonical example.

This handout closes the welfare-and-bargaining cluster (#28–#29). The Pareto handout supplied the apparatus for characterizing the set of efficient alternatives; the present handout uses that apparatus to characterize how rational parties select among them. The strategic / Rubinstein-style alternating-offers bargaining theory — which derives bargaining outcomes from extensive-form game-theoretic primitives rather than axiomatic principles — lives in the forthcoming game-theory cluster.

2 Bargaining problems

A reform commission with two factions disagreeing over how much of a controversial provision to keep; two states disputing a contested border, with war as the outside option; a labor-management negotiation over wages, with strike as the outside option; a coalition formation negotiation in a parliamentary government, with the no-coalition default as the outside option. Each of these settings has a common structural shape: several parties with conflicting preferences over a set of feasible alternatives, each party able to walk away to a default outcome that all parties prefer to avoid. Bargaining theory formalizes this shape via the pair (feasible utility profiles, disagreement point), and develops solution concepts that select an outcome.

Throughout this handout we focus on the two-player case for clarity. The general n-player case is structurally identical and the standard references handle it; we flag the generalization where it matters.

Definition 1 (Bargaining problem)**.** A *(two-player) bargaining problem* is a pair $(S, \boldsymbol{d})$ where:

- $S \subseteq \mathbb{R}^2$ is a non-empty, compact, convex set of *feasible utility profiles* (u_A, u_B) achievable by some agreement;
- $\boldsymbol{d} = (d_A, d_B) \in S$ is the *disagreement point*, the utility profile resulting if the players fail to agree;
- there exists $\boldsymbol{u} \in S$ with $u_A > d_A$ and $u_B > d_B$ (i.e., the disagreement point is strictly Pareto-dominated by some feasible alternative).

The set of all bargaining problems on $\mathbb{R}^2$ is denoted $\mathfrak{B}$.

The strict-domination clause is what makes bargaining a non-trivial problem: if no feasible alternative strictly Pareto-dominates the disagreement point, the parties have no shared interest in agreement, and the bargaining problem is empty in the substantive sense. The compactness and convexity of S are standard regularity conditions; convexity in particular reflects the assumption that players can randomize between agreements, so any convex combination of attainable utility profiles is itself attainable.

Example 2 (The Fearon war-bargaining problem)**.** Two states A and B contest a divisible territory of size 1. A bargained agreement assigns share $x \in [0, 1]$ to A and $1 - x$ to B, yielding utilities $u_A(x) = x$ and $u_B(x) = 1 - x$. The set of feasible utility profiles is

$$S = \{(x, 1 - x) : x \in [0, 1]\} \cup \{\text{convex combinations / lotteries}\} = \{(u_A, u_B) \in \mathbb{R}^2_+ : u_A + u_B \leq 1\}.$$

The disagreement point is the expected-utility profile if war breaks out: A wins the entire territory with probability $p \in (0, 1)$ and loses with probability $1 - p$, and both states pay war costs $c_A, c_B > 0$. Thus

$$\boldsymbol{d} = (p - c_A,\ 1 - p - c_B).$$

Note that $d_A + d_B = 1 - c_A - c_B < 1$, so $\boldsymbol{d}$ lies strictly below the Pareto frontier $\{u_A + u_B = 1\}$ and is strictly Pareto-dominated by every point on the frontier between $u_A = p - c_A$ and $u_A = p + c_B$. The structural fact that $\boldsymbol{d}$ sits strictly inside the Pareto frontier is exactly Fearon's puzzle: there is always a non-empty range of bargained outcomes both states prefer to war, and the substantive question is which one rational states should reach.

The set of *individually rational* alternatives in a bargaining problem is the set of $\boldsymbol{u} \in S$ with $\boldsymbol{u} \geq \boldsymbol{d}$ componentwise: the alternatives each player weakly prefers to the disagreement point. The set of *Pareto-optimal* alternatives in S is its Pareto frontier $\partial^P S$ (developed in #28). The intersection $\partial^P S \cap \{\boldsymbol{u} \geq \boldsymbol{d}\}$ — the Pareto-optimal individually-rational alternatives — is the *bargaining range*: the set of alternatives that pass both efficiency and individual-rationality screens.

Example 3 (Bargaining range in the Fearon problem)**.** Continuing Example 2. The Pareto frontier is $\{(x, 1-x) : x \in [0,1]\}$. Individual rationality requires $u_A \geq p - c_A$ (i.e., $x \geq p - c_A$) and $u_B \geq 1 - p - c_B$ (i.e., $1 - x \geq 1 - p - c_B$, equivalently $x \leq p + c_B$). The bargaining range is therefore

$$[p - c_A,\, p + c_B] \quad \text{(in } A\text{'s share)},$$

an interval of width $c_A + c_B$ — the sum of the war costs. The political-economy reading of Fearon's puzzle is that the bargaining range is non-empty whenever $c_A + c_B > 0$, which is generically the case; the bargaining-theoretic question is which point in this range the parties should reach.

3 The Nash bargaining solution

Different bargaining problems plausibly call for different bargained outcomes, and the question of which point in the bargaining range gets selected has a long methodological history. The first axiomatic answer — and still the most-cited one — is the *Nash bargaining solution* (Nash, 1950), which characterizes the bargained outcome as the maximizer of a particular product:

$$\mathfrak{N}(S, \boldsymbol{d}) = \operatorname*{arg\,max}_{\boldsymbol{u} \in S, \boldsymbol{u} \geq \boldsymbol{d}} (u_A - d_A)(u_B - d_B).$$

The maximand is the product of utility *gains over the disagreement point*, sometimes called the *Nash product*. The solution selects the alternative that maximizes this product, subject to feasibility and individual rationality.

The substantive content of the Nash bargaining solution is best seen on examples; we'll return to its axiomatic characterization in §4.

Example 4 (Nash bargaining in the Fearon problem)**.** For the Fearon bargaining problem, the Nash product is $(x - (p - c_A))(1 - x - (1 - p - c_B)) = (x - p + c_A)(p + c_B - x)$, maximized over $x \in [p - c_A, p + c_B]$. Differentiating,

$$\frac{d}{dx}\big[(x - p + c_A)(p + c_B - x)\big] = (p + c_B - x) - (x - p + c_A) = 2p + c_B - c_A - 2x,$$

which vanishes at

$$x_{\mathfrak{N}} = p + \frac{c_B - c_A}{2}.$$

The political-economy reading is direct: the Nash bargaining outcome equals the win probability p plus half the asymmetry in war costs, with the cost-asymmetry favoring whichever side has the smaller cost. When $c_A = c_B$ (symmetric costs), the bargained share equals the win probability p exactly — the bargained outcome "is" the expected war outcome, with the cost savings split evenly. When costs are asymmetric, the high-cost side gets a worse bargain because it has more to lose from war.

The Nash product can be re-expressed in logarithmic form: maximizing $(u_A - d_A)(u_B - d_B)$ is equivalent to maximizing $\log(u_A - d_A) + \log(u_B - d_B)$ on the bargaining range. The log form makes clear that the Nash solution is a (logarithmic) welfare-function maximizer over surplus utility, with equal weights on the two players' log-gains. This is the first connection back to the welfare-function characterization of #28 — a connection we develop in §6.

A useful generalization replaces the equal weights with arbitrary positive weights:

$$\mathfrak{N}_{\alpha,1-\alpha}(S, \boldsymbol{d}) = \underset{\boldsymbol{u} \in S, \boldsymbol{u} \geq \boldsymbol{d}}{\arg\max} (u_A - d_A)^{\alpha}(u_B - d_B)^{1-\alpha},$$

called the *generalized* or *asymmetric Nash bargaining solution* with bargaining weight $\alpha \in (0, 1)$ on player A. When $\alpha = 1/2$ this reduces to the standard Nash solution; for $\alpha \neq 1/2$, the high-weight player gets a more favorable bargain. The asymmetric weights are usually interpreted as reflecting differences in patience, outside options, or institutional bargaining power, and they are the workhorse for political-economy applications where the players are not symmetrically situated.

4 Axiomatic characterization

The Nash bargaining solution might look ad-hoc — why specifically the product, why specifically the gains over the disagreement point? — and Nash (1950)'s contribution was to show that it is in fact *uniquely* determined by four structural axioms on the bargaining-solution rule itself. The argument is the methodological foundation of the entire axiomatic-bargaining literature: the Nash solution is not derived from a particular cardinal interpretation of utility or a particular game-theoretic protocol; it is derived from minimal structural conditions on what counts as a reasonable bargaining solution.

Definition 5 (Bargaining solution)**.** A *bargaining solution* is a function $f : \mathfrak{B} \to \mathbb{R}^2$ that assigns to each bargaining problem $(S, \boldsymbol{d})$ a feasible utility profile $f(S, \boldsymbol{d}) \in S$.

The four Nash axioms.

- **(PE)** *Pareto efficiency*: for every $(S, \boldsymbol{d}) \in \mathfrak{B}$, $f(S, \boldsymbol{d}) \in \partial^P S$. (The selected outcome is Pareto-optimal.)
- **(SYM)** *Symmetry*: if S is symmetric (i.e., $(u_A, u_B) \in S$ implies $(u_B, u_A) \in S$) and $d_A = d_B$, then $f_A(S, \boldsymbol{d}) = f_B(S, \boldsymbol{d})$. (Symmetric problems have symmetric solutions.)
- **(SI)** *Scale invariance*: for any positive affine transformations $\phi_A(u) = a_A u + b_A$ and $\phi_B(u) = a_B u + b_B$ $(a_A, a_B > 0)$, define the rescaled problem $(\phi(S), \phi(\boldsymbol{d}))$ by applying ϕ_A to first coordinates and ϕ_B to second coordinates. Then $f(\phi(S), \phi(\boldsymbol{d})) = \phi(f(S, \boldsymbol{d}))$. (The solution is invariant under positive affine rescaling of utilities.)
- **(IIA)** *Independence of irrelevant alternatives*: if $(S, \boldsymbol{d}), (S', \boldsymbol{d})$ are bargaining problems with $S' \subseteq S$ and $f(S, \boldsymbol{d}) \in S'$, then $f(S', \boldsymbol{d}) = f(S, \boldsymbol{d})$. (Removing alternatives that aren't selected leaves the selection unchanged.)

Each axiom has a substantive reading worth stating. PE is the efficiency baseline (no rational solution should leave a Pareto improvement on the table). SYM is the no-favoritism principle (a solution rule shouldn't depend on which player is labeled A vs. B when the situation itself is symmetric). SI is the cardinal-invariance principle (the solution shouldn't depend on the units in which utilities are measured, since utilities are only ordinally meaningful up to positive affine transformation; this rules out interpersonal utility comparisons of the kind utilitarianism implicitly invokes). IIA is the structural-restriction axiom (eliminating non-selected alternatives shouldn't change the selection); it is the most controversial of the four, and it is what Kalai and Smorodinsky (1975) weakens.

Theorem 6 (Nash 1950)**.** *A bargaining solution $f : \mathfrak{B} \to \mathbb{R}^2$ satisfies (PE), (SYM), (SI), and (IIA) if and only if $f = \mathfrak{N}$, the Nash bargaining solution.*

Proof sketch. The "only if" direction has two steps. First, by (SI), the solution is determined by its values on bargaining problems with $\boldsymbol{d} = \mathbf{0}$ and a particular normalization of utilities (e.g., the ideal point at $(1, 1)$). Second, on this normalized class, (PE) + (SYM) determines the solution on *symmetric* problems (where S is symmetric around the diagonal): the unique Pareto-optimal symmetric point is the unique candidate. Third, (IIA) extends from symmetric problems to all problems: given an arbitrary $(S, \boldsymbol{d})$ with $\boldsymbol{d} = \mathbf{0}$ and a Nash-product maximizer $\boldsymbol{u}^*$, embed S in a larger symmetric problem S^* for which $\boldsymbol{u}^*$ is the symmetric Pareto-optimum (by aligning the level set of the Nash product through $\boldsymbol{u}^*$ as the Pareto frontier of S^*); by (PE) + (SYM), $f(S^*, \mathbf{0}) = \boldsymbol{u}^*$, and by (IIA), $f(S, \mathbf{0}) = \boldsymbol{u}^*$ as well, since $S \subseteq S^*$. The "if" direction is a direct verification that $\mathfrak{N}$ satisfies all four axioms. Osborne and Rubinstein (1990, Ch. 2) works through the proof in detail. □

The substantive content of the theorem is that the Nash bargaining solution is not chosen because the product-of-gains formula is intrinsically appealing; it is chosen because (PE), (SYM), (SI), and (IIA) collectively pin it down as the unique solution rule satisfying these structural conditions. The methodological move — characterizing a solution by axioms rather than by primitives — is the same move at work in the welfare-economics framework of #28 (where the welfare-function representation is characterized by structural conditions on the imputation set) and in the axiomatic decision theory of #26 (where vNM utility is characterized by axioms on preferences over lotteries). The axiomatic foundation is not the only one available; the Nash solution has a second derivation from an entirely different methodological route, and the connection between the two is one of the substantive payoffs of bargaining theory.[1]

5 Kalai–Smorodinsky and the IIA debate

The Nash axiom that has drawn the most substantive scrutiny is (IIA): the requirement that removing non-selected alternatives leave the selection unchanged. The substantive critique is that

[1] The Nash bargaining solution was originally derived by Nash (1950) from a different methodological angle: Nash modeled the bargaining process as a game and showed that the Nash product is the unique outcome satisfying his axioms, but also that it is the limit of the unique subgame-perfect equilibrium of an alternating-offers bargaining game as the discount factor approaches one (this latter result, completed by Rubinstein (1982) and Binmore, Rubinstein, and Wolinsky (1986), is sometimes called the "Nash program" — the connection between axiomatic and strategic bargaining). The structural reading is that the axiomatic characterization and the strategic characterization deliver the same solution, providing two distinct foundations for the Nash product. The strategic side is the topic of the forthcoming game-theory cluster's bargaining handout, which will develop the alternating-offers framework, derive Rubinstein's solution, and connect it to the Nash bargaining solution via the Nash program. For present purposes, the axiomatic characterization is the operative one — it gives a foundation for the Nash solution that does not depend on a particular game-theoretic protocol or on cardinal interpretations of utility.

(IIA) treats the structure of the bargaining set as irrelevant beyond which point gets selected, when in fact the structure of S — in particular, the maximum each player could attain if she could dictate the outcome — arguably should affect the bargained share. Two players with very different "ideal" utility levels should arguably split the surplus differently from two players with similar ones, and (IIA) is the axiom that rules this out.

Kalai and Smorodinsky (1975) replace (IIA) with a *monotonicity* axiom and derive a different bargaining solution.

Definition 7 (Ideal point)**.** For a bargaining problem $(S, \boldsymbol{d})$, the *ideal point* is

$$\boldsymbol{u}^*(S, \boldsymbol{d}) = (u_A^*, u_B^*), \quad u_i^* = \max\{u_i : \boldsymbol{u} \in S, \boldsymbol{u} \geq \boldsymbol{d}\}.$$

The ideal point records each player's individually-rational maximum — what each player would get if she were the dictator (subject to the other player's individual rationality being respected).

(MON) *Monotonicity*: if $(S, \boldsymbol{d})$ and $(S', \boldsymbol{d})$ have the same ideal point and $S \subseteq S'$, then $f(S', \boldsymbol{d}) \geq f(S, \boldsymbol{d})$ componentwise. (Expanding the bargaining set without changing the ideal point makes both players weakly better off.)

Theorem 8 (Kalai–Smorodinsky 1975)**.** *A bargaining solution $f : \mathfrak{B} \to \mathbb{R}^2$ satisfies (PE), (SYM), (SI), and (MON) if and only if $f = \mathfrak{K}$, where $\mathfrak{K}(S, \boldsymbol{d})$ is the unique Pareto-optimal point on the line segment from $\boldsymbol{d}$ to $\boldsymbol{u}^*(S, \boldsymbol{d})$.*

The Kalai–Smorodinsky solution selects the unique alternative on the Pareto frontier that "equalizes the proportional gains" — each player gets the same fraction of her maximum-possible gain over the disagreement point. Geometrically, draw the line segment from $\boldsymbol{d}$ to $\boldsymbol{u}^*$; intersect with $\partial^P S$; the intersection is $\mathfrak{K}$.

Example 9 (Kalai–Smorodinsky in the Fearon problem)**.** For the Fearon problem of Example 2, the ideal point is $\boldsymbol{u}^* = (1, 1)$ in the original utility profile (each player's individually-rational max is the entire stake), *or* more carefully, the ideal point with respect to the bargaining range is $\boldsymbol{u}^* = (p + c_B, 1 - p + c_A)$ (each player's max conditional on the other being individually rational). Working with the latter: the line from $\boldsymbol{d} = (p - c_A, 1 - p - c_B)$ to $\boldsymbol{u}^* = (p + c_B, 1 - p + c_A)$ has slope $(1 - p + c_A - (1 - p - c_B))/(p + c_B - (p - c_A)) = (c_A + c_B)/(c_A + c_B) = 1$. Setting $u_A = p - c_A + t(c_A + c_B)$ and $u_B = 1 - p - c_B + t(c_A + c_B)$ for $t \in [0, 1]$, and intersecting with $u_A + u_B = 1$ gives $t = 1/2$, hence

$$x_{\mathfrak{K}} = u_A = p - c_A + \tfrac{1}{2}(c_A + c_B) = p + \frac{c_B - c_A}{2}.$$

For the Fearon bargaining problem, the Kalai–Smorodinsky solution coincides with the Nash bargaining solution. This is a feature of the linear-utility / linear-frontier setting, not a general fact: when the Pareto frontier is curved (e.g., concave), the two solutions diverge. Thomson (2010, Ch. 3) works through the diverging-solutions cases in detail.

The substantive contrast between Nash and Kalai–Smorodinsky bites in settings where one player's ideal-point utility is much larger than the other's. Nash's (IIA) makes the bargained share insensitive to non-selected parts of the Pareto frontier; KS's (MON) makes it sensitive. For political-economy applications with asymmetric outside options or asymmetric stakes (e.g., a bargain between a unitary state and a federal subunit, where the subunit has much less to gain from any agreement), the choice between Nash and Kalai–Smorodinsky reflects a substantive judgment about whether the structural maximum-stake matters or only the gain from the bargained outcome.

6 Bargaining solutions as welfare-function maximizers

The Nash and Kalai–Smorodinsky solutions, characterized axiomatically in §4–§5, can equivalently be characterized as welfare-function maximizers in the sense of the previous handout (#28). The connection is structurally clean: each bargaining solution corresponds to a particular choice of welfare function, and the bargaining-theoretic axioms translate into structural restrictions on the welfare function. This connection is the bridge between the bargaining-theoretic and welfare-economics frameworks, and it is the structural answer to the question raised in #28's §5: where do the welfare weights come from when individuals' preferences are the primitive? Bargaining theory is one possible answer.

Proposition 10 (Nash bargaining as logarithmic welfare maximization)**.** *The Nash bargaining solution $\mathfrak{N}(S, \boldsymbol{d})$ maximizes the welfare function*

$$W_{\mathfrak{N}}(u_A, u_B) = \log(u_A - d_A) + \log(u_B - d_B)$$

over the bargaining range $\{\boldsymbol{u} \in S : \boldsymbol{u} \geq \boldsymbol{d}\}$.

The proof is direct: maximizing the Nash product is equivalent to maximizing its logarithm by monotonicity, and the logarithm of the product is the sum of logarithms. The welfare function is a (logarithmic) Bergson–Samuelson welfare function with equal weights on the two players' log-gains over the disagreement point. The substantive reading is that Nash bargaining is the special case of welfare-function maximization in which the welfare weights are equal but the units are log-gains-over-disagreement. The asymmetric Nash solution generalizes to $W_{\mathfrak{N}_\alpha}(u_A, u_B) = \alpha \log(u_A - d_A) + (1-\alpha)\log(u_B - d_B)$.

The Kalai–Smorodinsky solution can be characterized similarly, though with a more complicated welfare function involving the ideal point.

Example 11 (Welfare-function reading of Fearon)**.** In the Fearon problem of Example 2, the Nash bargaining solution maximizes $\log(u_A - (p - c_A)) + \log(u_B - (1 - p - c_B))$ over the bargaining range. The political-economy reading is that the bargained outcome is the maximizer of an equal-weighted Bergson–Samuelson welfare function with the pre-war expected utilities serving as the reference points. The disagreement-point-as-reference structure is what distinguishes the Nash bargaining solution from generic welfare maximization: the welfare function is anchored to what the parties would get if they failed to agree, not to some abstract "zero point" or to the parties' absolute utility levels.

The welfare-function reading also opens up a substantive line of analysis: comparative statics in bargaining problems can be derived by tracking how the relevant welfare function shifts as primitives change. In the Fearon setting, an increase in c_A (player A's war cost) shifts A's reference point downward, and the Nash-bargaining outcome's comparative-statics reading is that A gets a worse bargain. The general principle: the player whose disagreement-point utility falls gets a worse bargain; the player whose ideal-point utility rises (under MON) also gets a better bargain (under KS). The IFT-based comparative-statics tools from #20 deliver explicit formulas in particular settings.

Why bargained outcomes can fail to be reached. Even when the bargaining range is non-empty (every party would prefer some bargain to disagreement), bargaining can fail in practice. Fearon (1995) identifies three structural answers to the puzzle as it applies to war-bargaining: private information about types or capabilities together with strategic incentives to misrepresent them, commitment problems where one party would re-negotiate after the other complies, and issue indivisibility where the bargained outcome cannot be allocated continuously.[2]

7 What's next

This handout closes the welfare-and-bargaining cluster (#28–#29). Three strands extend it.

Strategic bargaining. The forthcoming game-theory cluster's bargaining handout will develop the strategic / extensive-form-game-theoretic side: Rubinstein (1982) alternating-offers bargaining as the canonical model, with the unique subgame-perfect equilibrium recovering the Nash bargaining solution as the discount factor approaches one (the Nash program). The strategic approach derives the bargained outcome from primitives (preferences, strategies, equilibrium concept) rather than from axioms, and it is the methodological complement to the axiomatic approach developed here.

Multilateral bargaining and legislative bargaining. The two-player axiomatic theory generalizes to n-player bargaining (the Nash and KS solutions both extend), but multilateral bargaining is most powerfully treated game-theoretically. Baron and Ferejohn (1989) legislative bargaining is the canonical model: a finite legislature bargains over a divisible policy with sequential proposers and recognition probabilities, with the equilibrium policy depending on patience and recognition structure. The Markov-perfect-equilibrium machinery from #24 is the structural input. The forthcoming game-theory cluster will develop legislative bargaining as an MPE application.

Mechanism design. When the bargaining problem is complicated by private information about preferences, the bargaining-theoretic question becomes a mechanism-design question: what bargaining protocols implement which Pareto outcomes when each party's type is private? The Myerson and Satterthwaite (1983) theorem is the central result — under generic conditions, no bargaining protocol simultaneously achieves Pareto efficiency, individual rationality, incentive compatibility, and budget balance. Mechanism design lives in the same forthcoming game-theory cluster.

For graduate-level treatments at this handout's level: Osborne and Rubinstein (1990) *Bargaining and Markets* is the canonical reference covering both axiomatic and strategic bargaining; Thomson (2010) is the canonical reference on the axiomatic side specifically, with extensive treatment of alternative axiomatizations beyond Nash and Kalai–Smorodinsky; Powell (2002) on the political-economy of war literature.

[2] The Fearon framework is the canonical political-economy application of bargaining theory and has organized a generation of formal-theory work on conflict. Powell (2002, 2006) extend the framework to dynamic settings with shifting power, in which commitment problems become endogenous (a state with rising relative capability cannot credibly commit not to revise the bargain in its favor later, so the declining state may rationally fight today). Fearon (1998) develops the private-information branch with an explicit signaling model. Leventoğlu and Tarar (2008) and Wagner (2000) extend the framework to multi-round bargaining and to settings with audience costs. The conceptual contribution of the Fearon framework is to reframe war as a bargaining failure rather than a bargaining strategy, with the substantive question being which structural conditions undermine the bargaining range that bargaining theory says should exist. The welfare-and-bargaining apparatus of this handout is the structural underpinning; the political-economy of war literature has populated it with substantive content. The forthcoming game-theory cluster's strategic-bargaining handout will develop the private-information and commitment-problem branches with explicit extensive-form game-theoretic models.

8 Exercises

Exercise 12. *Fearon bargaining range.* Two states A and B contest a divisible territory of size 1. State A's probability of victory is $p = 0.6$; war costs are $c_A = 0.1$ and $c_B = 0.2$. (a) Identify the disagreement point $\boldsymbol{d}$. (b) Compute the bargaining range. (c) Find the Nash bargaining solution. (d) Find the Kalai–Smorodinsky solution and verify that it equals the Nash solution in this setting (by Example 9).

Exercise 13. *Asymmetric Nash solution.* Continuing the setup of Exercise 12, suppose state A has bargaining weight $\alpha = 0.7$ (representing institutional power, patience, etc.). (a) Solve the asymmetric Nash bargaining problem $\max_x (x - p + c_A)^\alpha (p + c_B - x)^{1-\alpha}$. (b) Verify your answer reduces to the symmetric case when $\alpha = 1/2$. (c) Discuss in two sentences how bargaining-weight asymmetry affects the Nash-bargained share.

Exercise 14. *Where Nash and Kalai–Smorodinsky diverge.* Let $S = \{(u_A, u_B) \in \mathbb{R}^2_+ : u_A^2 + u_B^2 \leq 1\}$ (a quarter-disc in the non-negative orthant) and $\boldsymbol{d} = (0, 0)$. (a) Find the Nash bargaining solution. (b) Find the Kalai–Smorodinsky solution. (c) Verify they differ, and identify the reason geometrically (the curvature of the Pareto frontier).

Exercise 15. *Verifying the Nash axioms.* Show that the Nash bargaining solution $\mathfrak{N}$ satisfies each of (PE), (SYM), (SI), (IIA). (a) Verify (PE) directly from the definition of the Nash product. (b) Verify (SYM) using the symmetry of the maximand. (c) Verify (SI) using the property of the product under positive affine rescaling. (d) Verify (IIA) using the definition of arg max on a smaller feasible set.

Exercise 16. *Comparative statics: cost asymmetry.* In the Fearon setting, fix $p = 0.5$ and $c_B = 0.1$. (a) Find $x_{\mathfrak{N}}$ as a function of c_A. (b) Show $\partial x_{\mathfrak{N}} / \partial c_A < 0$ and interpret: an increase in A's war cost results in a worse Nash-bargained share for A. (c) Discuss in two sentences the political-economy reading of the result: how does the structural property that the higher-cost side gets a worse bargain show up in international-relations bargaining theory?

Exercise 17. *Comparative statics: shifting power.* Fix $c_A = c_B = 0.1$ and let p be the parameter. (a) Find $x_{\mathfrak{N}}$ as a function of p. (b) Show $\partial x_{\mathfrak{N}} / \partial p > 0$: as A's war probability of victory rises, A's Nash-bargained share rises. (c) Connect this to Powell (2002)'s shifting-power model: discuss in two or three sentences how a state with rising p over time has incentives that may undermine the bargaining range.

Exercise 18. *Static legislative bargaining as a Nash bargain.* A legislator and an executive bargain over a budget allocation $b \in [0, 1]$, with the legislator's utility $u_L(b) = b$ and the executive's $u_E(b) = 1 - b$. The disagreement (no-agreement) outcome is $\boldsymbol{d} = (0.2, 0.3)$ (each gets a small reservation utility from the unfilled office). (a) Identify the bargaining range. (b) Find the symmetric Nash bargaining solution. (c) If the legislator has greater institutional patience and an asymmetric bargaining weight $\alpha = 0.6$, find the asymmetric Nash bargaining solution.

Exercise 19. *Welfare-function reading of Nash.* Continuing Exercise 12: write the Nash bargaining problem as a welfare-function maximization in the form of Proposition 10. (a) Identify the welfare function explicitly. (b) Identify the welfare weights. (c) Discuss in two sentences how the structure of the welfare function (logarithmic in surplus, anchored at the disagreement point) differs from the standard utilitarian welfare function (linear in raw utility, anchored at zero).

Exercise 20. *Union–firm bargaining.* A union and a firm bargain over a wage $w \in [w_{\min}, w_{\max}]$. The union's utility is $u_U(w) = w$ (wage = welfare for the union), and the firm's utility is $u_F(w) = R - w$ (revenue minus wage). The disagreement point is the strike outcome: $\boldsymbol{d} = (w_{\min} - c_U, R - w_{\max} - c_F)$, where c_U and c_F are the strike costs. (a) Identify the bargaining range. (b) Find the symmetric Nash bargaining solution. (c) Discuss in two sentences how the structure of the wage solution depends on the asymmetry between strike costs c_U and c_F.

Exercise 21. *Bargaining failure.* Even when the bargaining range is non-empty (as it is generically in the Fearon setting), bargaining can fail in practice. In two or three sentences each, explain why the following structural conditions can produce bargaining failure: (a) private information about player types or capabilities, with strategic incentives to misrepresent; (b) commitment problems, where one party would re-negotiate after the other complies; (c) issue indivisibility, where the bargained outcome cannot be allocated continuously. (Cite Fearon (1995) as the canonical reference.)

References

Baron, David P. and John A. Ferejohn (1989). "Bargaining in Legislatures". In: *American Political Science Review* 83.4, pp. 1181–1206.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky (1986). "The Nash Bargaining Solution in Economic Modelling". In: *RAND Journal of Economics* 17.2, pp. 176–188.

Fearon, James D. (1995). "Rationalist Explanations for War". In: *International Organization* 49.3, pp. 379–414.

– (1998). "Bargaining, Enforcement, and International Cooperation". In: *International Organization* 52.2, pp. 269–305.

Kalai, Ehud and Meir Smorodinsky (1975). "Other Solutions to Nash's Bargaining Problem". In: *Econometrica* 43.3, pp. 513–518.

Leventoğlu, Bahar and Ahmer Tarar (2008). "Does Private Information Lead to Delay or War in Crisis Bargaining?" In: *International Studies Quarterly* 52.3, pp. 533–553.

Myerson, Roger B. and Mark A. Satterthwaite (1983). "Efficient Mechanisms for Bilateral Trading". In: *Journal of Economic Theory* 29.2, pp. 265–281.

Nash, John F. (1950). "The Bargaining Problem". In: *Econometrica* 18.2, pp. 155–162.

Osborne, Martin J. and Ariel Rubinstein (1990). *Bargaining and Markets*. San Diego: Academic Press.

Powell, Robert (2002). "Bargaining Theory and International Conflict". In: *Annual Review of Political Science* 5, pp. 1–30.

– (2006). "War as a Commitment Problem". In: *International Organization* 60.1, pp. 169–203.

Rubinstein, Ariel (1982). "Perfect Equilibrium in a Bargaining Model". In: *Econometrica* 50.1, pp. 97–109.

Thomson, William (2010). *Bargaining and the Theory of Cooperative Games: John Nash and Beyond*. Cheltenham, UK: Edward Elgar.

Wagner, R. Harrison (2000). "Bargaining and War". In: *American Journal of Political Science* 44.3, pp. 469–484.