Evolutionarily Stable Strategies

Game Theory · Module 7 — Evolutionary Game Theory

Node 7.2

Position in Knowledge Graph

System: Game Theory **Structural Domain:** Evolutionary Game Theory **Node:** 7.2

Node 7.1 established that the rest points of the replicator equation are the symmetric Nash equilibria. But not every rest point is stable — some are saddle points, some are neutral centers. The question becomes: which Nash equilibria can survive the pressure of evolutionary perturbation? This is the question Maynard Smith and Price (1973) answered by defining the **evolutionarily stable strategy** (ESS), perhaps the single most influential concept in evolutionary biology since the modern synthesis. An ESS is a strategy that, once established in a population, cannot be invaded by any rare mutant. The formal definition requires two conditions: the strategy is a best response to itself (Nash), *and* it does strictly better against any alternative best response than that alternative does against itself (the stability condition). The second condition is the new ingredient — it is what separates ESS from Nash and what guarantees dynamic stability under replicator dynamics. The payoff of this node is a clean comparison: ESS is strictly stronger than Nash, every ESS is asymptotically stable under the replicator, but not every asymptotically stable rest point is an ESS. The three concepts — Nash, ESS, replicator stability — form a strict hierarchy, and understanding where each one sits is essential for the rest of Module 7.

Formal Definition

Fix a finite symmetric two-player game (S, A) with strategy set $S = \{1, \dots, n\}$ and payoff matrix A. A mixed strategy is a vector $p \in \Delta^{n-1}$. The expected payoff when a p-player meets a q-player is

$$u(p, q) = p^\top A q.$$

Definition (Maynard Smith and Price, 1973). A strategy $p^* \in \Delta^{n-1}$ is an **evolutionarily stable strategy** (ESS) if for every strategy $q \neq p^*$:

1. **Nash condition:** $u(p^*, p^*) \geq u(q, p^*)$.
2. **Stability condition:** If $u(p^*, p^*) = u(q, p^*)$, then $u(p^*, q) > u(q, q)$.

Equivalently, p^* is an ESS iff for every $q \neq p^*$, there exists $\bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon})$:

$$u(p^*, (1-\epsilon)p^* + \epsilon q) > u(q, (1-\epsilon)p^* + \epsilon q).$$

This says: in a population playing p^*, a small fraction ϵ of mutants playing q cannot invade, because the incumbent p^* does strictly better against the post-invasion mixture than the mutant q does.

Theorem (ESS implies Nash). Every ESS is a symmetric Nash equilibrium. The converse is false in general.

Theorem (Strict Nash implies ESS). If p^* is a strict Nash equilibrium (unique best response to itself), then p^* is an ESS. The stability condition is vacuously satisfied because no $q \neq p^*$ achieves $u(q, p^*) = u(p^*, p^*)$.

Theorem (ESS and replicator stability). If p^* is an ESS, then it is an asymptotically stable rest point of the replicator equation. The converse is false: there exist asymptotically stable rest points of the replicator that are not ESS.

Intuition

The Nash condition says p^* is a best response to itself — no mutant can do *better* against the incumbent. But Nash allows ties: there may be alternative strategies q that do *equally well* against p^*. If such a neutral mutant appeared, Nash alone would not predict its fate. The stability condition resolves this: if q ties with p^* against the incumbent, then p^* must strictly beat q in the head-to-head matchup $u(p^*, q) > u(q, q)$. In biological terms: even if the mutant is equally fit in the old environment, the incumbent is fitter in a slightly mutant-contaminated environment.

Think of it as a two-round test. Round 1: can the mutant invade if the population is all p^*? If not (strict Nash), the test is over — p^* is ESS. If the mutant is "neutral" (ties in round 1), we proceed to round 2: can the mutant invade when the population is mostly p^* with a small fraction of the mutant itself? The stability condition says no.

The biological origin matters. Maynard Smith was thinking about animal conflicts: two animals meet and each chooses an aggressive or passive behavior. A population strategy is ESS if no rare mutation can spread. The concept was developed for biology but applies verbatim to any population where strategies spread by differential success — including poker.

An important negative result: **not every symmetric game has an ESS.** The classic example is Rock-Paper-Scissors (7.1). The unique symmetric Nash is $(1/3, 1/3, 1/3)$, but it fails the stability condition because any pure strategy q satisfies $u(p^*, p^*) = u(q, p^*) = 0$ (the Nash condition binds) and $u(p^*, q) = 0 = u(q, q)$ does not give the required strict inequality. Rock-Paper-Scissors has no ESS — which is consistent with the replicator dynamics being neutrally stable (closed orbits) rather than asymptotically stable at the interior.

Structural Role

7.2 occupies the central position in Module 7's hierarchy of stability concepts:

- **7.1 Replicator dynamics** provides the dynamics; 7.2 provides the stability criterion for rest points of those dynamics.
- **7.3 Hawk-Dove** is the premier example where the interior mixed equilibrium is an ESS, and where the ESS concept reveals the biological logic of mixed aggression.
- **7.4 Population games** extends the invasion analysis to large populations with general payoff structures.
- **7.5 Stochastic evolution** provides the probabilistic analogue: an ESS is the absorbing state that a stochastic dynamic converges to in the long run.
- **2.1 Nash equilibrium** — ESS strictly refines Nash. The relationship is: strict NE $\subset$ ESS $\subset$ NE, with each inclusion strict in general.

The concept hierarchy that 7.2 establishes is:

$$\text{Strict NE} \;\Rightarrow\; \text{ESS} \;\Rightarrow\; \text{Nash equilibrium}$$

$$\text{ESS} \;\Rightarrow\; \text{asymptotically stable under replicator} \;\Rightarrow\; \text{Lyapunov stable under replicator} \;\Rightarrow\; \text{rest point of replicator}$$

Each implication is strict. Understanding this chain is the main structural deliverable of the node.

Mathematical Development

The two-condition characterization

For $p^* \in \Delta^{n-1}$, let $C(p^*) = \{i \in S : (Ap^*)_i = u(p^*, p^*)\}$ be the **carrier** of p^* — the set of pure strategies that are best responses to p^*. The Nash condition says every strategy in the support of p^* is in $C(p^*)$, which is automatic for a symmetric Nash equilibrium.

Proposition. The stability condition is equivalent to: for all $q \neq p^*$ with $\operatorname{supp}(q) \subseteq C(p^*)$,

$$u(p^*, q) > u(q, q), \qquad \text{i.e.} \qquad (p^*)^\top Aq > q^\top Aq.$$

In matrix notation: $(p^* - q)^\top Aq > 0$ for all such $q \neq p^*$.

This is the condition that will connect to the Lyapunov function in 7.1. The relative-entropy derivative along the replicator was $\dot{H} = -(p^* - x)^\top Ax$; the ESS stability condition guarantees $(p^* - x)^\top Ax > 0$ near p^*, hence $\dot{H} < 0$, hence asymptotic stability.

Proof that ESS implies asymptotic stability

Theorem. If p^* is an interior ESS, then it is asymptotically stable under the replicator equation.

Proof. Use the relative entropy Lyapunov function $H(p^*, x) = \sum_i p_i^* \log(p_i^*/x_i)$. From 7.1:

$$\dot{H} = -(p^* - x)^\top Ax.$$

At $x = p^*$, $\dot{H} = 0$. We need $\dot{H} < 0$ for $x \neq p^*$ in a neighborhood. ESS condition: for $q \neq p^*$ with $u(q, p^*) = u(p^*, p^*)$, we have $(p^* - q)^\top Aq > 0$. By continuity, for x near p^*, we have $(p^* - x)^\top Ax > 0$. Hence $\dot{H} < 0$. □

Counterexample: asymptotically stable but not ESS

Consider the three-strategy game

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{pmatrix}.$$

The edge $x_3 = 0$, $x_1 + x_2 = 1$ consists entirely of Nash equilibria (strategies 1 and 2 are equivalent, strategy 3 is dominated). None of these is ESS — the stability condition fails between strategies 1 and 2 since $u(e_1, e_2) = 0 = u(e_2, e_2)$. Yet the face $x_3 = 0$ is globally attracting (strategy 3 always has below-average fitness). The set of Nash equilibria on this face is Lyapunov stable as a set, but no individual point in it is ESS. This separates the concepts.

Bishop–Cannings theorem

Theorem (Bishop and Cannings, 1978). If p^* is an interior ESS, then every pure strategy in the support of p^* earns the same payoff against p^*, and no pure strategy outside the support earns as much. Moreover, no proper mixture of strategies in the support of p^* can be an ESS.

This is the analogue of the indifference principle for mixed Nash, plus a uniqueness result: an interior ESS is the *unique* symmetric Nash with that support.

ESS in two-strategy games

For $S = \{1, 2\}$ with payoff matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the analysis is exhaustive:

- If $a > c$ and $b > d$: strategy 1 strictly dominates. $p^* = e_1$ is the unique ESS.

- If $a < c$ and $b < d$: strategy 2 strictly dominates. $p^* = e_2$ is the unique ESS.
- If $a > c$ and $b < d$: coordination game. Both e_1 and e_2 are ESS. The interior Nash exists but is *not* ESS (it is unstable under the replicator — a saddle point on the simplex).
- If $a < c$ and $b > d$: anti-coordination game (Hawk-Dove type). Neither pure strategy is ESS. The unique interior Nash $p^* = (d-b)/(a-c+d-b)$ is the unique ESS.

This exhaustive classification is the template for 7.3.

Worked Example

Hawk-Dove with explicit ESS computation

Payoff matrix (Maynard Smith, 1982):

$$A = \begin{pmatrix} (V-C)/2 & V \\ 0 & V/2 \end{pmatrix}$$

where $V > 0$ is the resource value and $C > 0$ is the cost of fighting. Strategy 1 is Hawk (escalate), strategy 2 is Dove (display only).

Case 1: $V > C$. Then $(V-C)/2 > 0$ and $V > V/2$, so Hawk strictly dominates. $p^* = e_H = (1, 0)$ is the unique ESS.

Case 2: $V < C$. Hawk does not dominate. Check anti-coordination conditions: $(V-C)/2 < 0 = u(\text{Dove}, \text{Hawk})$ and $V > V/2$. This is the anti-coordination case. The unique interior ESS is:

$$p^* = \frac{V}{C}, \qquad \text{i.e. } x_H^* = V/C.$$

Verification of ESS conditions. At the interior equilibrium $p^* = (V/C, 1 - V/C)$:

Nash condition: both strategies earn the same payoff against p^*. Compute $u(H, p^*) = (V-C)/2 \cdot V/C + V \cdot (1 - V/C) = V^2/(2C) - V/2 + V - V^2/C = V(1 - V/(2C))$. And $u(D, p^*) = 0 \cdot V/C + (V/2)(1 - V/C) = V/2 - V^2/(2C) = V(1 - V/(2C))/1$. Equal: Nash condition holds.

Stability condition: need $u(p^*, q) > u(q, q)$ for all $q \neq p^*$. For two strategies, this reduces to checking the sign of $d^2/d\epsilon^2$ of the invasion fitness, which is $(a - c + d - b) = (V-C)/2 - 0 + V/2 - V = -C/2 < 0$. The negative sign confirms ESS. Alternatively: the interior equilibrium of an anti-coordination game is always ESS (from the classification above).

Biological interpretation. When fighting is costly ($V < C$), the population settles at a fraction V/C of Hawks. If the resource value increases, the population becomes more aggressive. If the cost of fighting increases, it becomes more peaceful. The ratio V/C is the single parameter that governs the evolutionary outcome.

Poker Anchor

Concept mapping

Formal object	Poker instantiation
Incumbent p^*	Solver-approved GTO strategy
Mutant q	A novel exploitative style (e.g., massive overbetting, unusual sizing)

Formal object	Poker instantiation
Nash condition	GTO is a best response to itself — no exploit gains against it
Stability condition	GTO beats the exploit in head-to-head more than the exploit beats itself
ESS	GTO is the strategy that cannot be invaded by any deviation
Non-ESS Nash	A population-wide "gentlemen's agreement" that collapses when tested

Concrete situation: why GTO is the ESS of poker

Consider a population of online poker regulars. Over years, the population converges toward solver-like play. Why? Because GTO satisfies both ESS conditions. First, no alternative style beats GTO (Nash condition): by definition, GTO is a Nash equilibrium, so no deviation is profitable against it. Second, when a new style appears that ties against GTO (because GTO is indifferent over certain mixed actions), GTO outperforms that new style in the mixed population (stability condition): the new style's weakness is exposed by the general population while GTO remains robust.

The styles that *fail* the ESS test are the metagame agreements: "everyone plays tight preflop and nobody squeezes." This can be a Nash equilibrium (if everyone believes everyone else is tight, no one gains by squeezing). But it fails the stability condition: a single aggressive player (mutant) who squeezes profitably against the tight population will see their style spread (by imitation or by new players entering with that style), and the agreement collapses. Any metagame norm that fails the ESS stability condition is evolutionarily fragile.

What this understanding buys you

Not every equilibrium is robust. The distinction between "Nash but not ESS" and "ESS" maps directly to the distinction between "metagame norms that hold until someone tests them" and "strategies that are permanently stable." If you are evaluating whether a current table dynamic will persist, the ESS criterion is the right test: can a rare deviator invade?

GTO is not merely optimal, it is invasion-proof. The ESS framing gives a structural reason why GTO matters beyond the one-table analysis. In a population of players, GTO is the attractor that resists all mutant invasions. If you deviate from GTO, you may gain in the short run (against the current population), but the replicator will punish you: the strategies that exploit your deviation will grow, and your fitness will decline.

The meta always breaks toward the ESS. When you observe a metagame shift, you are watching a non-ESS equilibrium collapse under mutant invasion. The meta moves from the old equilibrium toward the ESS. Understanding this lets you predict the direction of meta shifts: they move toward the invasion-proof strategy.

Cross-Links

- **7.1 Replicator dynamics** — the dynamics for which ESS provides the stability concept.
- **7.3 Hawk-Dove** — the premier worked example of ESS in a biological context.
- **7.5 Stochastic evolution** — ESS as the stochastically stable state.
- **2.1 Nash equilibrium** — ESS strictly refines Nash.
- **2.2 Mixed-strategy Nash** — the indifference principle reappears in the Bishop–Cannings theorem.
- **Module 6 Mechanism design** — designing games so that the desired outcome is an ESS, not just a Nash.
- **9.5 Convergence to equilibrium** — learning dynamics that select ESS over non-ESS Nash.

Structural Compression

Invariant: An ESS is a Nash equilibrium satisfying an additional stability condition: among all strategies that tie against the incumbent, the incumbent wins the head-to-head. This is strictly stronger than Nash, strictly weaker than strict Nash (for mixed strategies), and equivalent to asymptotic stability of the replicator equation via the relative-entropy Lyapunov function. The hierarchy is: strict NE $\Rightarrow$ ESS $\Rightarrow$ Nash $\Rightarrow$ rest point of replicator.

Mechanism: Two-round invasion test: (1) can the mutant beat the incumbent in the incumbent's own environment? (Nash.) (2) If the mutant ties, can it survive in a slightly contaminated environment? (Stability.) The Lyapunov function $H(p^*, x)$ has $\dot{H} = -(p^* - x)^\top Ax$, which the ESS condition makes strictly negative, guaranteeing asymptotic stability.

Cross-System Echo: ESS is the biological analogue of **structural stability** in dynamical systems theory: a system is structurally stable if small perturbations of the vector field do not change the qualitative dynamics. An ESS is "structurally stable" in the strategy space: small perturbations (mutant invasions) of the population do not change the long-run outcome. The same logic appears in robust control theory (a controller is robust if small model perturbations do not destabilize the closed loop) and in machine learning (a classifier is robust if small adversarial perturbations do not flip predictions).

Exercises

1. Verify that every strict Nash equilibrium is an ESS by checking that the stability condition is vacuously satisfied.
2. Show that Rock-Paper-Scissors has no ESS. (Hint: check the stability condition at the unique interior Nash.)
3. For the two-strategy game $A = \begin{pmatrix} 3 & 0 \\ 5 & 1 \end{pmatrix}$ (Prisoner's Dilemma), find all Nash equilibria and determine which are ESS. Interpret biologically.
4. Prove the Bishop–Cannings theorem for two-strategy games: if the interior equilibrium $p^* = (p, 1-p)$ is an ESS, then both pure strategies earn the same payoff against p^*, and no other mixture is ESS.
5. Construct a symmetric three-strategy game that has exactly two ESS (both pure) and one interior Nash that is not ESS. Verify all three claims.
6. Show that in any two-strategy symmetric game, the number of ESS is at most 2 (the two pure strategies). In particular, the interior Nash in a coordination game is never ESS.
7. **Argue, structurally,** why the ESS concept is necessary beyond Nash equilibrium. What evolutionary failure mode does Nash alone miss, and why does the stability condition close this gap? Give a concrete biological or economic example of a Nash equilibrium that is not ESS and explain what goes wrong when a mutant invades.

Game Theory · Module 7 — Evolutionary Game Theory · Node 7.2
Concepts: evolutionary-stability, nash-equilibrium, stability
Prerequisites: 7.1, 2.1, 2.2
Enables: 7.3, 7.4, 7.5

hbar.university — own.your.cognition