Evolutionary Stable States (ESS)

Maynard Smith and Price (1973) defined the notion of an *evolutionary stable strategy* as immune to invasion by mutants:

- Their focus was on monomorphic populations: every member plays the same strategy, which can be a mixed strategy.

- We are concerned with a polymorphic population of agents each programmed with a pure strategy.

- We have seen the equivalence of these two problems.

Hence we can adapt the concept of an evolutionary stable strategy to a population setting:

- The term we shall use is *evolutionary stable state* (ESS).
Invasion

Let the state be \( x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \).

Consider a game \( F \), where \( F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \\ \vdots \\ F_n(x) \end{pmatrix} \).

Consider an invasion of mutants who make up a fraction \( \epsilon \) of the post-entry population.

The shares of each strategy in the mutant population are represented by \( y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \).
Invasion

Therefore, the post-entry population state is:

\[ x_\epsilon = (1 - \epsilon)x + \epsilon y = \begin{pmatrix} (1 - \epsilon)x_1 + \epsilon y_1 \\ (1 - \epsilon)x_2 + \epsilon y_2 \\ \vdots \\ (1 - \epsilon)x_n + \epsilon y_n \end{pmatrix}. \]

The average payoff in the incumbent population in the post-entry state is \( x'F((1 - \epsilon)x + \epsilon y) \).

The average payoff in the mutant population in the post-entry state is \( y'F((1 - \epsilon)x + \epsilon y) \).
Uniform Invasion Barrier

The average payoff in the incumbent population is higher if:

\[(y - x)'F((1 - \varepsilon)x + \varepsilon y) < 0.\]  

(1)

State \(x\) is said to admit a **uniform invasion barrier** if there exists an \(\bar{\varepsilon} > 0\) such that (1) holds for all \(y \in X - \{x\}\) and \(\varepsilon \in (0, \bar{\varepsilon})\).

That is, for all possible mutations \(y\), as long as the mutant population is less than fraction \(\bar{\varepsilon}\) of the postentry population, the incumbent population receives a higher average payoff.
Definition. State \( x \in X \) is an **evolutionary stable state (ESS)** of \( F \) if there exists a neighborhood \( O \) of \( x \) such that:

\[
(y - x)'F(y) < 0 \quad \text{for all } y \in O - \{x\}.
\]  

(2)

In other words, if \( x \) is an ESS, then for any state \( y \) sufficiently close to \( x \), a population playing \( x \) will receive a larger average payoff in state \( y \) than a population playing \( y \) (i.e. \( x \) is a better reply to \( y \) than \( y \) is to itself).

Note that this considers invasions of other states \( y \) by \( x \) rather than invasions of \( x \) by other states. Hence it is not clear, at present, why this should be a stability condition.
ESS and Invasion Barriers

**Theorem 7.1.** State $x \in X$ is an *evolutionary stable state* (ESS) if and only if it admits a uniform invasion barrier.

Thus if $x$ is stable in the face of an arbitrarily large population of entrants who mutate to a nearby state, then it is stable in the face of a sufficiently small population of entrants who mutate to an arbitrary state.
What is the relationship between ESS and NE?

**Definition.** Suppose that $x \in X$ is a NE. Then $(y - x)'F(x) \leq 0$ for all $y \in X$.

In addition, suppose there exists a neighborhood of $x$ that does not contain any other NE.

Then $x$ is an **isolated NE**.

**Proposition 7.2.** Every ESS is an isolated NE.
Proof

Let $x$ be an ESS of $F$, $O$ be the nhd posited in (2) and $y \in X - \{x\}$ (not necessarily in $O$).

Then for all $\varepsilon > 0$ sufficiently small, the postentry state $x_\varepsilon = \varepsilon y + (1 - \varepsilon)x$ is in $O$.

Given $x$ is an ESS, this implies that:

$$(x_\varepsilon - x)'F(x_\varepsilon) < 0$$
$$(\varepsilon y + (1 - \varepsilon)x - x)'F(x_\varepsilon) < 0$$
$$\varepsilon(y - x)'F(x_\varepsilon) < 0$$
$$(y - x)'F(x_\varepsilon) < 0.$$  

(3)
Proof

Taking $\varepsilon \to 0$ yields:

$$(y - x)'F(x) \leq 0,$$

by the continuity of $F$. That is, $x$ is a NE.

To establish that $x$ is isolated, note that if $w \in O - \{x\}$ were a NE then $(w - x)'F(w) \geq 0$, contradicting the supposition that $x$ is an ESS [by (2)]. □

The converse of Proposition 7.2 is not true.

- The mixed equilibrium of a two-strategy coordination game is a counterexample.
More on ESS and Nash

Therefore, *ESS is stronger than NE*.

In particular, an ESS satisfies the additional property:

Suppose there exists a state $y$ which is an alternative best reply to $x$, i.e. $(y - x)'F(x) = 0$.
—Then $(y - x)'F(y) < 0$, i.e. $x$ is a better reply to $y$ than $y$ is to itself.

Therefore:

- A strict NE is an ESS.
- A polymorphic population state (equivalent to a mixed NE) cannot be strict and hence must satisfy the additional property.
More on ESS and Nash

In the case in which agents are matched uniformly at random to play a normal form game (the case we have been focussing on), then it is easy to see why the additional property is required.

Suppose \((y - x)'F(x) = 0\), i.e. \(y\) is an alternative best reply to \(x\).

Then:

\[
(y - x)'F(\varepsilon y + (1 - \varepsilon)x) = \varepsilon(y - x)'F(y) + (1 - \varepsilon)(y - x)'F(x)
\]

\[
= \varepsilon(y - x)'F(y).
\]

(4)

Therefore, \((y - x)'F(y)\) must be negative for (1) to hold and hence, by Theorem 7.1, for \(x\) to be an ESS.
Example: Hawk Dove

<table>
<thead>
<tr>
<th></th>
<th>Hawk</th>
<th>Dove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawk</td>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>Dove</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

ESS: $x = \left( \frac{2}{3}, \frac{1}{3} \right)$. ESS payoff = 0.
Example: Hawk Dove

Consider a mutation $y$ such that $y_1 > x_1 = \frac{2}{3}$.

Check that $(y - x)'F((1 - \varepsilon)x + \varepsilon y) < 0$ for all such $y$:

$$(y_1 - x_1)[-2((1 - \varepsilon)x_1 + \varepsilon y_1) + 4((1 - \varepsilon)(1 - x_1) + \varepsilon(1 - y_1))].$$

This equals:

$$(y_1 - x_1)\varepsilon[-2y_1 + 4(1 - y_1)]$$

because $-2 \times \frac{2}{3} + 4 \times (1 - \frac{2}{3}) = 0$. This in turn equals:

$$(y_1 - x_1)\varepsilon[4 - 6y_1]$$

which is negative because $y_1 > \frac{2}{3}$ by hypothesis.

A similar argument can be applied to the case $y_1 < x_1$. Hence $x$ is an ESS.
The Prisoners' Dilemma

\[
\begin{array}{c|cc}
&C & D \\
\hline
C & 3 & 5 \\
3 & 0 & 1 \\
D & 5 & 1 \\
\end{array}
\]

\textbf{NE/ESS: } x = (0, 1).

Therefore, an ESS is not necessarily efficient.
Not Every Game has an ESS

\[
x = \left( \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)
\] is the unique NE and therefore the only possible ESS.
Not Every Game has an ESS

Note that $x$ is a polymorphic population state (equivalent to a mixed strategy), so any basis vector (pure strategy) is an alternative best reply to $x$.

Check that the additional property holds: $(e_1 - x)'F(e_1) < 0$, where $e_1 = (1, 0, 0)$, i.e. the pure-strategy $A$.

This is not the case: $x'F(e_1) = e_1'F(e_1) = 1$. 
The Iterated Prisoners’ Dilemma

- Two players engage in a series of PD games.

- The engagement ends after the current round with probability $\delta < \frac{1}{2}$. We call this the stopping probability.

- Consider a population in which three strategies are present:
  - $C$—always cooperate,
  - $D$—always defect,
  - $T$—tit-for-tat, i.e. start by cooperating, thenceforth cooperate in period $t$ if partner cooperated in $t - 1$. 
Expected Payoffs

Within each pairing:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$\frac{3}{\delta}$</td>
<td>0</td>
<td>$\frac{3}{\delta}$</td>
</tr>
<tr>
<td>D</td>
<td>$\frac{5}{\delta}$</td>
<td>$\frac{1}{\delta}$</td>
<td>$4 + \frac{1}{\delta}$</td>
</tr>
<tr>
<td>T</td>
<td>$\frac{3}{\delta}$</td>
<td>$\frac{1}{\delta} - 1$</td>
<td>$\frac{3}{\delta}$</td>
</tr>
</tbody>
</table>

Over all pairings:

\[
F_C(x) = (x_C + x_T) \frac{3}{\delta}
\]
\[
F_D(x) = x_C \frac{5}{\delta} + x_D \frac{1}{\delta} + x_T \left(4 + \frac{1}{\delta}\right)
\]
\[
F_T(x) = (x_C + x_T) \frac{3}{\delta} + x_D \left(\frac{1}{\delta} - 1\right)
\]
All-\(T\) is not an ESS

Let \(x = (x_D, x_C, x_T) = (0, 0, 1)\).

Consider any alternative state \(y\) such that \(y_D = 0\).

\[
(y - x)'F(y) = (0 ~ y_C ~ y_T - 1) \begin{pmatrix} F_D(y) \\ F_C(y) \\ F_T(y) \end{pmatrix}
\]

\[
= y_C F_C(y) + (y_T - 1) F_T(y)
\]

\[
= [y_C + (y_T - 1)] \frac{3}{\delta} \quad \text{(recall that } y_D = 0) 
\]

\[
= [y_C + (1 - y_C - 1)] \frac{3}{\delta} \quad \text{(because } y_T = 1 - y_C) 
\]

\[
= 0.
\]

This violates (2). Hence all-\(T\) is not an ESS.

This is a case of evolutionary drift.
Vector Field