Bernergard, Axel; Wärneryd, Karl

Working Paper

Finite-population 'Mass-Action' and evolutionary stability

CESifo Working Paper, No. 3378

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Bernergard, Axel; Wärneryd, Karl (2011) : Finite-population 'Mass-Action' and evolutionary stability, CESifo Working Paper, No. 3378, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/46575

CESifo Working Papers
www.cesifo.org/wp

Finite-Population "Mass-Action" and Evolutionary Stability

Axel Bernergård
Karl Wärneryd

CESIFO WORKING PAPER NO. 3378
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
MARCH 2011

CESifo *Working Paper No. 3378*

Finite-Population "Mass-Action" and Evolutionary Stability

Abstract

Nash proposed an interpretation of mixed strategies as the average pure-strategy play of a population of players randomly matched to play a normal-form game. If populations are finite, some equilibria of the underlying game have no such corresponding "mass-action" equilibrium. We show that for mixed strategy equilibria of 2 × 2 games, the requirement of such a correspondence is equivalent to neutral evolutionary stability.

JEL-Code: C720, C730.

Keywords: mass action, finite population games, evolutionary stability.

Axel Bernergård
Department of Economics
Stockholm School of Economics
Box 6501
Sweden – 11383 Stockholm
Axel.Bernergard@hhs.se

Karl Wärneryd
Department of Economics
Stockholm School of Economics
Box 6501
Sweden – 11383 Stockholm
Karl.Warneryd@hhs.se

March 6, 2011
Bernergård thanks the Knut and AliceWallenberg Foundation for financial support. Wärneryd thanks Jeroen Swinkels for discussions on this topic that took place in a previous millenium, and the Bank of Sweden Tercentenary Foundation for financial support.

1 Introduction

Did John Nash anticipate evolutionary game theory in his 1950 dissertation? Young [7] suggests he did, based on the following passage, which can be seen as introducing the random-matching population games that are the basis of Maynard Smith's (Maynard Smith and Price [4]; Maynard Smith [3]) evolutionary approach:

> We shall now take up the "mass-action" interpretation of equilibrium points. In this interpretation solutions have no great significance. It is unnecessary to assume that the participants have full knowledge of the total structure of the game, or the ability and inclination to go through any complex reasoning processes. But the participants are supposed to accumulate empirical information on the relative advantages of the various pure strategies at their disposal.
>
> To be more detailed, we assume that there is a population (in the sense of statistics) of participants for each position of the game. Let us also assume that the "average playing" of the game involves n participants selected at random from the n populations, and that there is a stable average frequency with which each pure strategy is employed by the "average member" of the appropriate population [...] Thus the assumptions we made in this "mass-action" interpretation lead to the conclusion that the mixed strategies representing the average behavior in each of the populations form an equilibrium point [...] Actually, of course, we can only expect some sort of approximate equilibrium, since the information, its utilization, and the stability of the average frequencies will be imperfect. (Nash [5].)

Nash here says nothing very explicit about any evolutionary selection dynamics leading to equilibrium play. Young, along with, e.g., Leonard [2] and Hofbauer [1], interprets the passage to be about a best-response dynamics, where players respond myopically to the current population distribution. But

		Player 2	
		s_1	s_2
Player 1	s_1	$1,1$	$0,0$
	s_2	$0,0$	$1,1$

Table 1: A coordination game.

even without this latter reading, as we shall see, there is a sense in which a notion of evolutionary stability is implicit in the discussion.

Taking the "mass-action" story literally it seems natural to study *finite* populations. But then some equilibria of the underlying game cannot be supported as "mass-action" equilibria. Consider a non-trivial symmetric mixed-strategy equilibrium of a symmetric normal-form game. Suppose we try to implement this equilibrium in a setting where a finite population of individuals, who are only allowed to play pure strategies, are randomly matched to play our underlying game. As Nash notes, "the stability of the average frequencies will be imperfect," since it cannot be the case that each individual faces the average strategy, as the individual's own strategy choice affects the average.

Nash's "mass-action" idea therefore carries within it its own refinement or stability notion. The fact that an individual's strategy choice in the finite population game holds information itself affecting that choice acts as a perturbation of the strategy distribution at the equilibrium. In the following we shall see that requiring stability in the face of such perturbations is, indeed, closely related to Maynard Smith's [3] concept of evolutionary stability. Specifically, we define a non-artifactuality criterion for equilibria of symmetric 2-player games and show that it is equivalent to neutral stability for completely mixed strategies in 2×2 games.

2 Finite Population Games: Examples

Consider the game in Table 1. It has three equilibria, one where both players play s_1, one where both players play s_2, and a mixed-strategy equilibrium in which the players play each of their pure strategies with probability .5.

Now consider the same game played by many pairs of players, randomly matched from an uncountably infinite population. Such a game has one equilibrium where all players play s_1, one where all players play s_2, and one in which half of the players play s_1 and the other half play s_2.

The latter equilibrium, which implements in the population game version the mixed-strategy equilibrium of the original game, has nothing that corresponds to it in a game played by a *finite* population. To see this, suppose the number of players is $n \geq 2$, with n even. Suppose half of the players play s_1, the other half s_2. Then an s_1-player will be matched with another s_1-player with probability $(n-2)/2(n-1)$, and with an s_2-player with probability $n/2(n-1)$, which implies that his expected payoff is

$$\frac{n-2}{2(n-1)} \cdot 1 + \frac{n}{2(n-1)} \cdot 0.$$

If instead he played s_2, his expected payoff would be

$$\frac{n-2}{2(n-1)} \cdot 0 + \frac{n}{2(n-1)} \cdot 1,$$

which is strictly greater. A similar disincentive holds for any s_2-player. Hence for no finite n is there an equilibrium in which half of the players play s_1 and the other half play s_2. That there is such an equilibrium in the infinite-population case is an artifact of that special setting.

In the Battle-of-the-Sexes game of Table 2, on the other hand, there is a unique symmetric equilibrium in which each player plays s_1 with probability 2/3. This equilibrium does have "mass-action" equivalents. Suppose 2/3 of the n players in the population game play s_1, the rest s_2. Each s_1-player is then playing a best reply, since his expected payoff is

$$\frac{2n-3}{3(n-1)} \cdot 0 + \frac{n}{3(n-1)} \cdot 2,$$

		Player 2	
		s_1	s_2
Player 1	s_1	$0,0$	$2,1$
	s_2	$1,2$	$0,0$

Table 2: The Battle of the Sexes.

which is strictly greater than

$$\frac{2n-3}{3(n-1)} \cdot 1 + \frac{n}{3(n-1)} \cdot 0,$$

his expected payoff if instead he played s_2. Similarly, s_2-players are also playing best replies.

It also so happens that the equilibrium mixed strategy of the second example is an *evolutionarily stable strategy* in the sense of Maynard Smith [3], whereas that of the first example is not. Maynard Smith's idea was that in order to survive evolutionary selection, a strategy when employed by all players of a large population who are randomly matched in pairs to play symmetric 2-player games should be stable against a small invasion of players doing something different. As we have seen, in finite populations the fact that no single player can face the population distribution, or average strategy, without distortion, in effect introduces small "mutations" around the equilibrium. We now go on to study this relationship in more detail.

3 Non-Artifactuality

Let G be a symmetric, 2-player, finite, normal form game with common pure strategy set S. Let Σ be the set of mixed strategies of G, where, if $\sigma \in \Sigma$, $\sigma(s)$ is the probability assigned by σ to the pure strategy s. The payoff function $u : S \times S \to \mathbb{R}$ is extended in the standard fashion to mixed strategies. Symmetry means that if one player is playing s and the other s', then the s-player's payoff is $u(s,s')$ and the other's is $u(s',s)$.

We now want to capture the idea from the previous section of equilibria of G that have something corresponding to them in a finite population game with random matching. The following definition seems to do this, while at the same time abstracting from problems that have to do with the fact that in finite populations, an average strategy can only involve probabilities that are rational numbers.

Definition 1 *A symmetric strategy profile (σ,σ) of G is* non-artifactual *if there is $\delta > 0$ such that for all $\varepsilon \in (0,\delta)$, all $s \in S$ such that $\sigma(s) > 0$, and all $s' \in S$, it holds that*

$$(\sigma(s)-\varepsilon)u(s,s)+\sum_{s''\neq s}\sigma(s'')u(s,s'') \geq (\sigma(s)-\varepsilon)u(s',s)+\sum_{s''\neq s}\sigma(s'')u(s',s'').$$

Observation 1 *If (σ,σ) is non-artifactual, then (σ,σ) is an equilibrium.*

Proof. Suppose (σ,σ) is not an equilibrium. Then there must be some pure strategy s in the support of σ that is not a best reply to σ. That is, there exist $s \in S$ with $\sigma(s) > 0$ and $s' \in S$ such that $u(s,\sigma) < u(s',\sigma)$. The inequality in the definition of non-artifactuality is equivalent to

$$u(s,\sigma)-u(s',\sigma) \geq \varepsilon\left(u(s,s)-u(s',s)\right).$$

Since the left-hand side is strictly negative, this inequality is violated for all $\varepsilon > 0$ sufficiently small. □

We shall therefore in the following refer to non-artifactual strategy profiles and non-artifactual equilibria interchangeably.

The following lemma provides a convenient alternative characterization of non-artifactuality.

Lemma 1 *Let (σ,σ) be an equilibrium of G. Then (σ,σ) is non-artifactual if and only if $u(s',s) \geq u(s,s)$ for all $s' \in S$ with $u(s',\sigma) = u(\sigma,\sigma)$ and all $s \in S$ with $\sigma(s) > 0$.*

Proof. The idea here is that if $u(s',\sigma) = u(\sigma,\sigma)$, then s' and s are both best replies to σ. If also $u(s,s) > u(s',s)$, then s' must be doing better than s against the other strategies $s'' \neq s$ in the support of σ. Then it is profitable for an s-player to deviate to s' since he faces more of the other strategies and less of s. Conversely, if ε is small enough, then it is only profitable to deviate to a strategy s' that is a best reply to σ and such that $u(s',s) < u(s,s)$.

To see this formally, suppose that $u(s',s) \geq u(s,s)$ for all $s' \in S$ with $u(s',\sigma) = u(\sigma,\sigma)$ and all $s \in S$ with $\sigma(s) > 0$. Let $\mu > 0$ be such that $\mu \geq u(s,s) - u(s',s)$ for all $s', s \in S$. Let $\delta > 0$ be such that

$$u(\sigma,\sigma) - u(s',\sigma) > \delta\mu \tag{1}$$

for all $s' \in S$ with $u(s',\sigma) < u(\sigma,\sigma)$. As shown below, this δ has the desired property.

Let $s \in S$ be such that $\sigma(s) > 0$ and let $s' \in S$ be arbitrary. We have to show that, for all $\varepsilon \in (0,\delta)$,

$$(\sigma(s) - \varepsilon)u(s,s) + \sum_{s'' \neq s'} \sigma(s'')u(s,s'') \geq (\sigma(s) - \varepsilon)u(s',s) + \sum_{s'' \neq s} \sigma(s'')u(s',s''). \tag{2}$$

After rearranging, (2) may be written

$$\sum_{s'' \in S} \sigma(s'')u(s,s'') - \sum_{s'' \in S} \sigma(s'')u(s',s'') \geq \varepsilon \left(u(s,s) - u(s',s)\right),$$

from which follows that

$$u(s,\sigma) - u(s',\sigma) \geq \varepsilon \left(u(s,s) - u(s',s)\right),$$

or, equivalently,

$$u(\sigma,\sigma) - u(s',\sigma) \geq \varepsilon \left(u(s,s) - u(s',s)\right).$$

The last equivalence uses that $u(s,\sigma) = u(\sigma,\sigma)$ since (σ,σ) is an equilibrium and $\sigma(s) > 0$. We thus have to show that, for all $\varepsilon \in (0,\delta)$, it holds that

$$u(\sigma,\sigma) - u(s',\sigma) \geq \varepsilon \left(u(s,s) - u(s',s)\right). \tag{3}$$

Since (σ,σ) is an equilibrium there are two possibilities. Either we have $u(s',\sigma) = u(\sigma,\sigma)$, or $u(s',\sigma) < u(\sigma,\sigma)$. If we have $u(s',\sigma) < u(\sigma,\sigma)$, then (1) holds and this implies that (3) holds for all $\varepsilon \in (0,\delta)$. If we have $u(s',\sigma) = u(\sigma,\sigma)$, then, by assumption, $u(s',s) \geq u(s,s)$, and (3) holds trivially since the left hand side is 0 and the right hand side is non-positive. This completes the "if" part of the proof.

To prove the converse implication, suppose that there exists $s', s \in S$, such that $u(s',\sigma) = u(\sigma,\sigma)$, $\sigma(s) > 0$, and $u(s',s) < u(s,s)$. For these s', s, the inequality (3) does not hold for any $\varepsilon > 0$ since the left hand side is 0 and the right hand side is positive, so (σ,σ) is not non-artifactual. □

Corollary 1 *Suppose σ is a completely mixed strategy such that (σ,σ) is an equilibrium of G. Then (σ,σ) is non-artifactual if and only if each $s \in S$ is a worst reply to itself.*

Proof. The strategy profile (σ,σ) is such that $\sigma(s) > 0$ for all $s \in S$ and $u(s',\sigma) = u(\sigma,\sigma)$ for all $s' \in S$. Hence Lemma 1 implies that (σ,σ) is non-artifactual if and only if $u(s',s) \geq u(s,s)$ for all $s, s' \in S$. □

Corollary 2 *Suppose σ is a completely mixed strategy such that (σ,σ) is an equilibrium of the 2×2 game G. Then (σ,σ) is non-artifactual if and only if s_1 is a best reply to s_2 and s_2 is a best reply to s_1.*

Proof. This follows from Corollary 1. □

Proposition 1 *If G is a 2×2 game, then it has a non-artifactual equilibrium.*

Proof. If s_1 is a best reply to s_1, then (s_1, s_1) is a non-artifactual equilibrium. If s_2 is a best reply to s_2, then (s_2, s_2) is a non-artifactual equilibrium. Suppose that s_1 is not a best reply to s_1 and that s_2 is not a best reply to s_2. Then, since a symmetric equilibrium must exist, there is a competely mixed strategy σ that is a best reply to itself. Since s_2 is a best reply to s_1 and s_1 is a best reply to s_2 Corollary 2 implies that (σ,σ) is non-artifactual. □

		Player 2	
		s_1	s_2
Player 1	s_1	$1,1$	$1,1$
	s_2	$1,1$	$1,1$

Table 3: No ESS.

4 Evolutionary Stability

We next relate non-artifactuality to evolutionary stability. Maynard Smith (e.g., Maynard Smith [3]) defined an evolutionarily stable strategy as follows.

Definition 2 *A strategy σ of G is an* evolutionarily stable strategy (ESS) *if for all $\sigma' \in \Sigma$, $\sigma' \neq \sigma$, there is $\delta > 0$ such that for all $\varepsilon \in (0,\delta)$ it holds that*

$$(1-\varepsilon)u(\sigma,\sigma)+\varepsilon u(\sigma,\sigma') > (1-\varepsilon)u(\sigma',\sigma)+\varepsilon u(\sigma',\sigma').$$

The ESS notion involves the idea that a strategy is stable if when it is played by the entire population, any small group of invading mutants playing some other strategy would do strictly worse in the perturbed population. Although Maynard Smith originally did not propose any explicit dynamics to support ESS, it has been shown to correspond to asymptotically stable states of the replicator dynamics, a model of asexual reproduction (Taylor and Jonker [6]).

Not every game has an ESS, however. In the 2×2 game of Table 3 every symmetric pair of strategies is non-artifactual, but the game has no ESS. Maynard Smith also proposed the following weakening of the ESS criterion, which only requires that a strategy should do at least as well as that played by any small invading group of mutants.

Definition 3 *A strategy σ of G is a* neutrally stable strategy (NSS) *if for all $\sigma' \in \Sigma$ there is $\delta > 0$ such that for all $\varepsilon \in (0,\delta)$ it holds that*

$$(1-\varepsilon)u(\sigma,\sigma)+\varepsilon u(\sigma,\sigma') \geq (1-\varepsilon)u(\sigma',\sigma)+\varepsilon u(\sigma',\sigma').$$

		Player 2	
		s_1	s_2
Player 1	s_1	$1,1$	$1,1$
	s_2	$1,1$	$2,2$

Table 4: A counterexample.

Proposition 2 *Suppose σ is an NSS of the 2×2 game G. Then (σ,σ) is non-artifactual.*

Proof. If $\sigma(s_1)=1$ or $\sigma(s_2)=1$, then the implication holds trivially since all symmetric pure strategy equilibria are non-artifactual. Assume by way of contradiction that there exists a completely mixed strategy σ such that σ is an NSS but (σ,σ) is not non-artifactual. By Corollary 2, we then either have that s_1 is not a best reply to s_2 or that s_2 is not a best reply to s_1. Assume, without loss of generality since the strategies can be relabelled, that s_1 is not a best reply to s_2, i.e., that we have that

$$u(s_1,s_2)<u(s_2,s_2). \tag{4}$$

Since (σ,σ) is a completely mixed equilibrium, all strategies σ' are best replies to σ. Since σ is an NSS, it follows that

$$u(\sigma',\sigma')\leq u(\sigma,\sigma')$$

for all σ'. In particular, it holds for $\sigma'=s_2$ that

$$u(s_2,s_2)\leq u(\sigma,s_2)=\sigma(s_1)u(s_1,s_2)+\sigma(s_2)u(s_2,s_2). \tag{5}$$

Using (4) in (5) yields

$$u(s_2,s_2)<\sigma(s_1)u(s_2,s_2)+\sigma(s_2)u(s_2,s_2)=u(s_2,s_2).$$

Since the inequality $u(s_2,s_2)<u(s_2,s_2)$ cannot hold, we conclude that no such σ can exist. □

It is *not* the case that every non-artifactual equilibrium of a 2×2 game involves a neutrally stable strategy. In the game of Table 4, (s_1, s_1) is non-artifactual, but s_1 is not an NSS since it can be invaded by s_2.

As the following result shows, however, if we consider only completely mixed strategies the notions of non-artifactuality and neutral stability coincide in 2×2 games.

Proposition 3 *Suppose σ is a completely mixed strategy such that (σ,σ) is non-artifactual in the 2×2 game G. Then σ is an NSS.*

Proof. From Corollary 2, we have that s_1 is a best reply to s_2 and that s_2 is a best reply to s_1. To prove the proposition it is sufficient to show that this implies that

$$u(\sigma',\sigma') \leq u(\sigma,\sigma') \tag{6}$$

for all σ'.

To see that this holds, let σ' be an arbitrary mixed strategy, and define

$$x := \sigma'(s_1) - \sigma(s_1) = \sigma(s_2) - \sigma'(s_2).$$

The inequality (6) can be written as

$$\sigma'(s_1)u(s_1,\sigma') + \sigma'(s_2)u(s_2,\sigma') \leq \sigma(s_1)u(s_1,\sigma') + \sigma(s_2)u(s_2,\sigma'),$$

or, equivalently,

$$x\,u(s_1,\sigma') - x\,u(s_2,\sigma') \leq 0.$$

Expanding the expressions $u(s_1,\sigma')$ and $u(s_2,\sigma')$ yields

$$x\sigma'(s_1)u(s_1,s_1) + x\sigma'(s_2)u(s_1,s_2) - x\sigma'(s_1)u(s_2,s_1) - x\sigma'(s_2)u(s_2,s_2) \leq 0.$$

After replacing $\sigma'(s_1)$ with $x+\sigma(s_1)$ and $\sigma'(s_2)$ with $\sigma(s_2)-x$, the inequality may be written

$$x\,(x+\sigma(s_1))(u(s_1,s_1) - u(s_2,s_1)) + x\,(\sigma(s_2)-x)(u(s_1,s_2) - u(s_2,s_2)) \leq 0.$$

		Player 2		
		s_1	s_2	s_3
Player 1	s_1	$1,1$	$2,0$	$0,2$
	s_2	$0,2$	$1,1$	$2,0$
	s_3	$2,0$	$0,2$	$1,1$

Table 5: A 3×3 counterexample.

Collecting the x^2 and x terms and writing this as a polynomial in x yields

$$\begin{aligned} &x^2(u(s_1,s_1)-u(s_2,s_1)+u(s_2,s_2)-u(s_1,s_2))\\ +x(\sigma(s_1)(u(s_1,s_1)-u(s_2,s_1))&+\sigma(s_2)(u(s_1,s_2)-u(s_2,s_2)))\leq 0. \end{aligned}$$

The coefficient of the x term is

$$\sigma(s_1)u(s_1,s_1)+\sigma(s_2)u(s_1,s_2)-\sigma(s_1)u(s_2,s_1)-\sigma(s_2)u(s_2,s_2)$$

which is equal to $u(s_1,\sigma)-u(s_2,\sigma)$. Since both s_1 and s_2 are best replies to σ, this number is 0 and the x term disappears. We can conclude that to complete the proof it is sufficient to show that the coefficient of the x^2 term is non-positive. That is, it is sufficient to show that

$$u(s_1,s_1)-u(s_2,s_1)+u(s_2,s_2)-u(s_1,s_2)\leq 0.$$

This inequality does indeed hold, since s_2 is a best reply to s_1 and s_1 is a best reply to s_2. □

As we go beyond the family of 2×2 games, it is no longer the case that neutral stability of a completely mixed strategy implies non-artifactuality. The Rock-Scissors-Paper game of Table 5 has a unique NSS such that each of the three pure strategies are played with equal probability, but a pair of such strategies is not a non-artifactual equilibrium. For instance, in the finite population game an s_1-player meets another s_1-player with probability less than $1/3$, but each of s_2 and s_3 with probability greater than $1/3$, so s_2 is a better reply.

		Player 2		
		s_1	s_2	s_3
Player 1	s_1	0,0	6,6	0,2
	s_2	6,6	0,0	0,2
	s_3	2,0	2,0	2,2

Table 6: Another 3×3 counterexample.

Nor is it the case that completely mixed non-artifactual equilibria necessarily involve neutrally stable strategies in the 3×3 case. Consider the game of Table 6. All equilibria of this game are non-artifactual since each pure strategy is a worst reply to itself. In particular, the equilibrium (σ, σ) with $\sigma(s_1) = \sigma(s_2) = \sigma(s_3) = 1/3$ is non-artifactual. Let σ' be the strategy with $\sigma'(s_1) = \sigma'(s_2) = 1/2$. Then $u(\sigma', \sigma') = (6/4) + (6/4) = 3$. We also have that

$$u(\sigma, \sigma') = \frac{1}{3}u(s_1, \sigma') + \frac{1}{3}u(s_2, \sigma') + \frac{1}{3}u(s_3, \sigma') = \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 2 < 3.$$

That is, we have $u(\sigma, \sigma') < u(\sigma', \sigma')$. Since we also have $u(\sigma', \sigma) = u(\sigma, \sigma)$, σ cannot be an NSS.

References

[1] Josef Hofbauer. From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS. *Selection*, 1:81–88, 2000.

[2] Robert J Leonard. Reading Cournot, reading Nash: The creation and stabilisation of the Nash equilibrium. *Economic Journal*, 104:492–511, 1994.

[3] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.

[4] John Maynard Smith and G R Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

[5] John F Nash. Non-cooperative games. PhD dissertation, Princeton University, 1950.

[6] Peter D Taylor and Leo B Jonker. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.

[7] H Peyton Young. Commentary: John Nash and evolutionary game theory. *Games and Economic Behavior*, 71:12–13, 2011.