Metroeconomica 51:3 (2000) 235–256

COMBINING RATIONAL CHOICE AND EVOLUTIONARY DYNAMICS: THE INDIRECT EVOLUTIONARY APPROACH

Manfred Königstein and Wieland Müller*
Humboldt-University Berlin

(August 1998; revised June 1999)

ABSTRACT

In this study we propose a formal framework for the indirect evolutionary approach initiated by Güth and Yaari. It allows us to endogenize preferences and to study their evolution. We define two-player indirect evolutionary games with observable types and show how to incorporate symmetric as well as asymmetric situations. We show how to apply solution concepts that are well known from game theory and evolutionary game theory to solve these games. For illustration we include two examples.

1. INTRODUCTION

In this paper we describe a special technique in modeling and analyzing human behavior: the indirect evolutionary approach (IEA). It was initiated by Güth and Yaari (1992) in a study on the evolution of reciprocal behavior and was subsequently applied to investigate, for example, the evolution of trust (Güth and Kliemt (1994)), monopolistic competition (Güth and Huck (1997)), as well as the evolution of altruism within a duopoly framework (Bester and Güth (1998)) and within ultimatum games with production (Königstein (2000)).

The IEA comprises modeling and solving an indirect evolutionary

* We wish to thank Martin Brown, Werner Güth, Steffen Huck, Jörg Oechssler and two anonymous referees for helpful comments and suggestions. Financial support from the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 373 'Quantifikation und Simulation ökonomischer Prozesse', Humboldt-Universität zu Berlin, is gratefully acknowledged. Part of this research was done while the first author was visiting the Institute for Empirical Economics at the University of Zürich.

© Blackwell Publishers Ltd 2000, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

game. We focus on two-player games. In such games a large population of players is matched in pairs. Each pair plays a two-player game. As in any other noncooperative two-player game, each player chooses a strategy and the strategy profile determines the utilities of both players. In addition, each strategy profile determines a fitness vector, with fitness being a measure of reproductive (evolutionary) success. A key idea of the IEA is that utility and evolutionary success need not coincide; while individuals may act rationally by choosing strategies that maximize utility, this may not maximize their evolutionary success. If individuals differ in evolutionary success, the personal characteristics of more successful individuals will spread within the population faster than the characteristics of the less successful ones. This leads to a dynamic process that determines the long-run distribution of personal characteristics within a society. The characteristics are modeled as parameters of the utility functions. Along with studying the evolution of preferences the analysis also determines the long-run distribution of strategy choices that are associated with the evolving preferences.

In solving an indirect evolutionary game two classes of familiar solution concepts are combined: solution concepts of noncooperative game theory (Nash equilibrium, subgame perfect equilibrium etc.) and static or dynamic solution concepts of evolutionary game theory (ESS, LESS, (asymptotically) stable fixed points of evolutionary dynamics etc.).[1]

The IEA has emerged via applications. So far, only little work has been done on establishing a formal conceptual framework. Here we take a step in this direction. We define the class of two-player indirect evolutionary games with observed types, which is the class of games studied in most applications to date, and we present a unifying description of symmetric as well as asymmetric games. We give two examples: a symmetric one where we apply a static evolutionary solution concept (ESS) and an asymmetric one where we apply a dynamic solution concept (asymptotically stable fixed points of the replicator dynamics). We conclude with a methodological discussion in which we address some potential criticisms of the IEA and point to extensions. One possible extension could be the substitution of the perfect rationality assumption by some kind of boundedly rational learning process. Furthermore one could replace the evolutionary dynamic by some other

[1] ESS and LESS refer to the concepts of an evolutionarily stable strategy and a limit evolutionarily stable strategy, respectively.

© Blackwell Publishers Ltd 2000

kind of social dynamic, e.g. imitation. The formalism of the IEA is also suited for such applications. These extensions might be fruitful.

2. TWO-PLAYER INDIRECT EVOLUTIONARY GAMES WITH OBSERVABLE TYPES

2.1 Game models

In an indirect evolutionary game Γ there is a large population of individuals (throughout we will actually assume an infinite population) who are repeatedly matched in pairs to play a two-person game which we refer to as the *base game* of Γ. Player i ($i = 1, 2$) chooses a strategy s_i in order to maximize individual utility u_i and bequeaths his type t_i (with u_i depending on t_i) to the next generation depending on evolutionary success r_i. Formally, we define Γ as follows.

Definition 1. A two-player indirect evolutionary game with observable types is described by the 8-tuple

$$\Gamma \equiv ((S_i, T_i, u_i, r_i)_{i=1,2})$$

- S_i denotes a nonempty set consisting of player i's pure strategies s_i. The tuple $s = (s_1, s_2)$ represents a pure strategy vector. The set $S = \{(s_1, s_2)|s_i \in S_i\} = S_1 \times S_2$ is the set of all pure strategy vectors.
- T_i is a nonempty set of possible types of player i (type space). A single element of set T_i is denoted by t_i. The tuple $t = (t_1, t_2)$ consisting of both players' types is referred to as a type vector, and we will write $T = \{(t_1, t_2)|t_i \in T_i\} = T_1 \times T_2$ for the set of all type vectors. In this study a player's type is assumed to be a parameter of his utility function. So, the term 'type' essentially refers to a type of preferences. The types are assumed to be observable.
- u_i denotes player i's utility function. It is a mapping u_i: $S \times T \rightarrow \mathbb{R}$.
- r_i is player i's evolutionary success (fitness function). It is a mapping r_i: $S \rightarrow \mathbb{R}$.
- For a given type vector t the components S_i and u_i define a game $G^{\mathrm{B}}(t)$:

  $$G^{\mathrm{B}}(t_1, t_2) \equiv ((S_i, u_i)_{i=1,2})|_{(t_1,t_2)}$$

  which we will refer to as a base game of Γ. Accordingly $G^{\mathrm{B}} = \{G^{\mathrm{B}}(t)|t \in T\}$ denotes the set of base games. Modeling an indirect

© Blackwell Publishers Ltd 2000

evolutionary game basically means embedding a set of base games into an evolutionary system.

- Throughout the paper we investigate evolution in a single population of infinite size.[2] Each individual of the population is randomly matched with one other individual to play the base game $G^{B}(t_1, t_2)$. They play only once and then bequeath their respective type to their 'children' with reproductive success of i according to r_i. In the next period the population of children is matched and plays the game. This process continues forever.
- The assumption of a single-population model implies that an individual can be assigned to either of the two roles $i = 1$ or $i = 2$ in the base game. So, an individual should actually be endowed with a type vector (t_1, t_2) instead of a one-dimensional type t_i. This will be assumed within asymmetric games. Within symmetric games an individual is sufficiently characterized by t_i.

An indirect evolutionary game thus combines two familiar types of theoretical models: a noncooperative game and an evolutionary system. Each base game $G^{B}(t_1, t_2)$ is a standard noncooperative game with strategy spaces S_1, S_2 and with t_1 and t_2 being parameters of the players' utility functions (which are commonly known). T_i is the set of possible utility parameters and represents the mutation space of the evolutionary system. The population share of individuals of type t_i which is present in period $\tau + 1$ depends on evolutionary fitness r_i in period τ and on mutations. We shall say more about the evolutionary dynamics below. A key idea of the IEA is that, in general, utility and evolutionary success are not the same; i.e. the functions u_i and r_i differ. This should not be viewed as a critical assumption. Rather, it is the *specific purpose* of the IEA to investigate evolution of behavior when the motives (preferences) that drive individual decisions differ from the forces that determine long-run survival of motives (preferences) within a society.

Another special feature of indirect evolutionary games is that r_i does not depend on t, at least not directly. Remember that t_i is the characteristic of an individual that is the object of mutation and inheritance within this model. That fitness does not directly depend on it may be surprising from the perspective of usual evolutionary games.

[2] It will be obvious later how to extend the indirect evolutionary approach to two-population models.

© Blackwell Publishers Ltd 2000

However, we will show below that the assumption of utility-maximizing players leads to strategy choices s_i that depend on t. Since r_i is defined on S, and s is a function of t, it follows that evolutionary success does ultimately depend on types. This will become clear when we now describe how to solve an indirect evolutionary game.

2.2 Solution concepts

Within an indirect evolutionary game a player i is endowed with a utility function u_i and an individual preference parameter t_i. His choice variable is s_i.[3] In accordance with standard economic theory we assume that i will choose s_i in order to maximize his utility; i.e. we assume that players are rational. Within strategic games this assumption implies that the chosen strategy profile is a Nash equilibrium. We denote by $s^*(t)$ a Nash equilibrium of $G^B(t_1, t_2)$ which is considered the unique solution of the game. Saying that $s^*(t)$ is considered the unique solution of the game implicitly assumes that if the base game exhibits multiple equilibria then some kind of equilibrium refinement[4] or equilibrium selection theory[5] is applied to determine a unique solution.

In determining their equilibrium strategies the players take into account utility but not evolutionary success. However, we can determine the evolutionary success of each player i as

$$r_i^*(t_1, t_2) \equiv r_i(s_1^*(t), s_2^*(t))$$

i.e. by evaluating the evolutionary success function r_i at equilibrium strategies. In the terminology of dynamic optimization the function r_i^* is a value function. We will refer to it as the 'indirect evolutionary success function'. Indirect evolutionary success r_i^* depends only on the players' types. These types are inherited from generation to generation and may mutate within the limits of the type space T_i. Note that at this stage we have all the ingredients of a 'usual' evolutionary game G^E which is described by the type spaces T_i and the indirect evolutionary success functions r_i^*:

[3] We assume that the players choose pure strategies in order to simplify our exposition. Thus, we assume that a solution in pure strategies exists. However, it will be obvious how to extend the approach to allow for mixed strategies.

[4] For details see van Damme (1991).

[5] See Harsanyi and Selten (1988).

© Blackwell Publishers Ltd 2000

$$G^{\mathrm{E}} \equiv ((T_i, r_i^*)_{i=1,2}).$$

We will refer to the models usually investigated in evolutionary game theory[6] as 'direct evolutionary games' in order to distinguish them from an indirect evolutionary game Γ. Accordingly, G^{E} will be referred to as the 'direct evolutionary game that is associated with Γ'.

In direct evolutionary games, evolutionary success depends directly on types. Similarly, one could have started here by modeling the dependence of r_i^* on t_1 and t_2, instead of deriving G^{E} from the indirect evolutionary game Γ—which is what we do. By doing so we motivate the specification of r_i^* by some underlying structure that allows for rational strategy choices s_i based on inherited types. We view this as an advantage of the IEA, and we return to this point in the discussion below.

We spent this effort on explaining differences and similarities of direct and indirect evolutionary games in order to show that we do not have to develop new solution concepts to solve the evolutionary part (G^{E}) of an indirect evolutionary game Γ. Rather, we can simply apply one of the static or dynamic solution concepts that are well known for direct evolutionary games[7] to analyze G^{E}. We show how to do this below.

But first, we generally characterize the solution of an indirect evolutionary game as follows.

Definition 2: A solution of a two-player indirect evolutionary game Γ as defined above consists of

1. a type vector $t^* = (t_1^*, t_2^*)$ (or a set of type vectors) which is the solution of the game G^{E} associated with Γ and
2. the equilibrium strategy choices $s_i^*(t_1^*, t_2^*)$ associated with t^* (or the set of equilibrium strategies that are associated with the solution set of types).

Up to the point where G^{E} is established by determining r_i^* both symmetric and asymmetric indirect evolutionary games proceed in the same manner. The above definition describes the solution for both kinds of games. In the following we discuss some special concerns in the analysis of symmetric versus asymmetric games.

6 See for example the textbooks by Weibull (1995) and Vega–Redondo (1996) or the paper by Hammerstein and Selten (1994).
7 See for example Weibull (1995).

© Blackwell Publishers Ltd 2000

2.3 Symmetric games

In this section we show more specifically how to solve an indirect evolutionary game by applying the concepts of an evolutionarily stable strategy (ESS) to a symmetric game Γ, or by looking for stable fixed points of replicator dynamics. We will show later how to solve asymmetric games. The ESS concept belongs to the class of static solution concepts whereas the second approach is a dynamic solution method.

A symmetric indirect evolutionary game Γ is defined as above with the additional restrictions $S_1 = S_2$, $T_1 = T_2$ as well as

$$u_1(s_1, s_2, t_1, t_2) = u_2(s_2, s_1, t_2, t_1)$$

and

$$r_1(s_1, s_2) = r_2(s_2, s_1).$$

We can solve the base games $G^{\mathrm{B}}(t_1, t_2)$ by determining $s^*(t)$ for all $t = (t_1, t_2) \in T$. Accordingly, the indirect evolutionary success function

$$r_i^*(t_1, t_2) = r_i(s_1^*(t), s_2^*(t))$$

together with T_i $(i = 1, 2)$ defines the symmetric direct evolutionary game

$$G^{\mathrm{E}} = ((T_i, r_i^*)_{i=1,2})$$

that is associated with Γ. Within symmetric games the players' roles 1 or 2 are meaningless and it is customary in evolutionary game theory to drop the role index i.

The important point in analyzing a symmetric game is that the evolutionary solution has to be derived only for one of the two roles. So, within symmetric games an individual is characterized by a single type rather than a type vector. We therefore assume without loss of generality that the evolutionary success of an individual of type t^k when matched with an individual of type t^l is given by

$$r^*(t^k, t^l) = r_1^*(t^k, t^l)$$

with $t^k, t^l \in T_1$. Accordingly $r^*(t^k, t^k)$ is the evolutionary success of a t^k-individual when matched with a t^k-individual.[8]

© Blackwell Publishers Ltd 2000

An ESS of the indirect evolutionary game Γ is given by an ESS t^{k*} of the associated game G^{E}. The ESS concept requires that the following conditions hold:

$$r^*(t^{k*}, t^{k*}) \geqslant r^*(t^k, t^{k*}) \text{ for all } t^k \in T_1 \tag{2.1}$$

and

$$\begin{aligned} &r^*(t^{k*}, t^k) > r^*(t^k, t^k) \\ &\text{for all } t^k \neq t^{k*} \text{ that satisfy } r^*(t^{k*}, t^{k*}) = r^*(t^k, t^{k*}). \end{aligned} \tag{2.2}$$

To see that the solution of a symmetric game Γ fits into the formalism of a solution for general indirect evolutionary games as defined above just note that $t^* = (t_1^*, t_2^*)$ is given by (t^{k*}, t^{k*}) and that the corresponding solution strategies are $s_1^*(t^k, t^k)$ and $s_2^*(t^k, t^k)$.

ESS is a static solution concept, but one might equally well apply dynamic solution concepts to solve G^{E} (and thus Γ). As an example we briefly sketch how to use the continuous-time replicator dynamics[9] (see Weibull (1995)) within a symmetric indirect evolutionary game. Again, we simplify the notation and consider an individual's type as given by $t^k \in T_1 = \{t^1, \ldots, t^n\}$; i.e., T_1 is now assumed to be a finite set of n different types.

Let the state of the population at time τ be given by $x(\tau) = (x_{t^1}(\tau), \ldots, x_{t^n}(\tau))$ with $\sum_{k=1}^n x_{t^k}(\tau) = 1$ where $x_{t^k}(\tau)$ is the population share of individuals who are endowed with type t^k at time τ. The dynamics for the population share x_{t^k} are given by

$$\dot{x}_{t^k} = [e_k Ax - xAx]x_{t^k} \tag{2.3}$$

with

---

[8] Of course, the definitions of r^*, t^k and t^l (thereby dropping the role index) are not necessary. However, readers who are familiar with direct evolutionary games may appreciate seeing the customary notation.

[9] Note that the replicator dynamics described here operate on a finite set of types (or strategies). This is the standard modeling approach in the literature (see for example Weibull (1995) or Vega-Redondo (1996)). Replicator dynamics operating on mixed strategies are discussed by for example Zeeman (1981), Akin (1982), Thomas (1985) and Bomze (1991). For recent research regarding evolutionary dynamics on continuous action spaces, see for example Binmore and Seymour (1995), Friedman and Yellin (1996), Hopkins and Seymour (1996) and Oechssler and Riedel (1998).

© Blackwell Publishers Ltd 2000

$$A = \begin{pmatrix} r^*(t^1, t^1) & \cdots & r^*(t^1, t^n) \\ \vdots & \ddots & \vdots \\ r^*(t^n, t^1) & \cdots & r^*(t^n, t^n) \end{pmatrix} \tag{2.4}$$

where $\dot{x}_{t^k}$ denotes the time derivative of x_{t^k} and e_k is the kth unit vector. Again, without loss of generality, $r^*(t^k, t^l) \equiv r_1^*(t^k, t^l)$ represents the evolutionary success of a t^k-individual when paired with a t^l-individual (for all $k, l = 1, \ldots, n$). One may now look for population states $x^*(\tau)$ which are stable or asymptotically stable fixed points of the dynamics (2.3).[10]

We have shown in these examples how to solve G^{E} (and thus Γ) via a static or a dynamic solution concept. Other solution concepts for indirect evolutionary games suggest themselves: since G^{E} is formally no different than in usual direct evolutionary games one can transfer other static solution concepts and define, for example, neutral evolutionarily stable strategies (NESS), limit evolutionarily stable strategies (LESS) or evolutionarily stable sets for indirect evolutionary games. Furthermore, all dynamic solution concepts which are common in evolutionary game theory can be applied to analyze indirect evolutionary games as well. On the other hand, solving an indirect evolutionary game faces the same problems as for direct evolutionary games, and the choice of a solution concept will depend on the specifics of the game G^{E} at hand. Importantly though, deriving G^{E} from Γ does not raise any issues regarding the solution of G^{E} other than what is known for direct evolutionary games.

2.4 Example 1: Symmetry and ESS

As an example of a symmetric two-player indirect evolutionary game we investigate whether firm owners who care not only for profit but also for consumer surplus may survive evolution. The base game is modeled as a duopoly market. Consider two duopolists (players 1 and 2) playing a Cournot game on a homogeneous market. Their quantity choices are s_1

[10] Roughly, a fixed point $x^*(\tau)$ (i.e. a point at which the right-hand side of the equations in (2.3) vanish) is stable if a small perturbation of the population mixture cannot lead far away from $x^*(\tau)$, and the fixed point is asymptotically stable if it is stable and if any sufficiently small perturbation is followed by a movement back to $x^*(\tau)$.

© Blackwell Publishers Ltd 2000

and s_2 with $s_i \in S_i = [0, \frac{1}{2}]$. We assume a linear demand function which is suitably normalized such that player i's profit π_i is given by

$$\pi_i(s_i, s_j) = (1 - s_i - s_j)s_i$$

for $i, j \in \{1, 2\}$, $i \neq j$. Player i's preferences are described by the following utility function:

$$u_i(s_i, s_j, t_i) = t_i \pi_i(s_i, s_j) + (1 - t_i)C(s_i, s_j)$$

where

$$C(s_i, s_j) = \int_0^{s_1+s_2} (1 - y)\mathrm{d}y - (1 - s_i - s_j)(s_i + s_j)$$

is net consumer surplus and with $t_i \in [\frac{1}{2}, 1]$. Thus, player i is not only concerned with his own profit but also with the welfare of consumers. Specifically, i's utility is a weighted average of his own profit and consumers' surplus. The weight t_i is an individual preference parameter (i's type). It is observable and is the object of evolution; i.e. t_i will spread according to the evolutionary success of player i.[11] Note that the specification of the utility function allows for preferences that are usually assumed in economics. Namely, for $t_i = 1$ player i only cares about his own profit.

Given preference parameters t_1, t_2 for the two players, the strategy spaces S_i and the utility functions u_i define the base game $G^{\mathrm{B}}(t_1, t_2)$ of an indirect evolutionary game Γ. Furthermore, by identifying $T_i = [\frac{1}{2}, 1]$ we have a type space, so that the only missing component for an indirect evolutionary game is the evolutionary success function r_i. We assume that

$$r_i(s_i, s_j) \equiv \pi_i(s_i, s_j)$$

So, evolutionary success is given by monetary success. From an economist's perspective this is certainly a natural assumption: while individuals may entertain various kinds of subjective preferences as captured by u_i, the long-run survival of a preference type and the associated strategy

[11] Within this model u_i does not (directly) depend on t_j. This is just a special case of the class of functions $u_i(s_i, s_j, t_i, t_j)$ that was assumed in the general description of indirect evolutionary games we gave above.

© Blackwell Publishers Ltd 2000

choices depend on their monetary consequences. We think that measuring evolutionary success by monetary payoff is a very useful specification of an indirect evolutionary game and we will discuss it later.

Having defined all components of the indirect evolutionary game the solution is derived as follows. Maximizing u_i with respect to s_i gives the system of first-order conditions

$$\frac{\partial}{\partial s_i} u_i(s_i, s_j, t_i) = s_i(1 - 3t_i) + s_j(1 - 2t_i) + t_i = 0$$

for $i = 1, 2$, which can be solved for equilibrium strategies $s_i^*(t)$:[12]

$$s_i^*(t) = s_i^*(t_i, t_j) = \frac{t_i t_j + t_j - t_i}{5t_i t_j - t_j - t_i}$$

for $i, j = 1, 2$. Note that $s_i^*(t_i, t_j) \geqslant 0$ for $t_i, t_j \in [\frac{1}{2}, 1]$.

The functions $s_i^*(t)$ characterize a unique solution for every base game $G^{\mathrm{B}}(t)$ of the indirect evolutionary game Γ. Substituting s_1 and s_2 in $r_i(s_1, s_2)$ by $s_1^*(t)$ and $s_2^*(t)$ gives the following indirect evolutionary success function $r_i^*(t_i, t_j)$:

$$r_i^*(t_i, t_j) = r_i(s_i^*(t), s_j^*(t)) = \frac{(3t_i t_j - t_j - t_i)(t_i t_j + t_j - t_i)}{(5t_i t_j - t_j - t_i)^2}$$

for $i, j = 1, 2$. The type spaces T_i together with r_i^* define the symmetric direct evolutionary game G^{E} that is associated with Γ. Without loss of generality we consider $r^*(t^k, t^l) = r_1^*(t^k, t^l)$ as a t^k-individual's evolutionary success when matched with a t^l-individual (with $t^k, t^l \in T_1$). In order to derive an ESS t^{k*} (see Maynard Smith (1982)) we first have to solve the following first-order condition:

$$\frac{\partial}{\partial t^k} r^*(t^k, t^l) = 0$$

for

$$t^k = \frac{t^l(4t^l - 1)}{8(t^l)^2 - 5t^l + 1}.$$

[12] Since $\partial^2 u_i(s_i, s_j, t_i)/\partial s_i^2 = 1 - 3t_i$, the second-order condition for a maximum is satisfied if $t_i > \frac{1}{3}$, which holds by definition.

© Blackwell Publishers Ltd 2000

Setting $t^k = t^l = t^{k*}$ and solving the resulting quadratic equation with respect to t^{k*} results in two candidates for an ESS of G^{E} (and thus of Γ): $t^{k*} = [9 + \sqrt{(17)}]/16$ or $t^{k*} = [9 - \sqrt{(17)}]/16$. Since $[9 - \sqrt{(17)}]/16 < \frac{1}{3}$ the second candidate is not feasible given the definition of T_1. So, only the candidate $t^{k*} = [9 + \sqrt{(17)}]/16$ remains. The stability requirement (2.1), i.e. $r^*(t^{k*}, t^{k*}) \geqslant r^*(t^k, t^{k*})$, is equivalent to $[16t^k - 9 - \sqrt{(17)}]^2 \geqslant 0$ which is always satisfied. Moreover, since $r^*(t^{k*}, t^{k*}) = r^*(t^k, t^{k*})$ is satisfied only for $t^k = t^{k*}$ the stability requirement (2.2) is also fulfilled. Thus we have the following.

Proposition 1. In the symmetric indirect evolutionary game as defined above $t^{k*} = [9 + \sqrt{(17)}]/16 \approx 0.82$ is the unique evolutionarily stable strategy.

Proposition 1 says that within our model of evolution based on duopoly interaction only those types of firm owners survive evolution who care for consumer welfare. Egoistic preferences—as they are assumed throughout most of economic theorizing—would die out in such markets.

2.5 *Asymmetric games*

We will now show how to solve asymmetric two-player indirect evolutionary games with observed types. These are games Γ in which one or several of the following inequalities hold:

$$S_1 \neq S_2$$

$$T_1 \neq T_2$$

$$u_1(s_1, s_2, t_1, t_2) \neq u_2(s_2, s_1, t_2, t_1)$$

or

$$r_1(s_1, s_2) \neq r_2(s_2, s_1).$$

We keep the assumption of a single-population model. Each individual of the population will be paired with one other individual. One of the paired individuals is assigned role 1 while the other is assigned role 2. It is assumed that each of the two role assignments is equally likely.

Within asymmetric games it is important to read s_i as role i's strategy

© Blackwell Publishers Ltd 2000

to distinguish it from an individual j's strategy ($j = 1, 2$) which shall be defined as the individual j's behavior strategy s^j:

$$s^j = (s_1^j, s_2^j).$$

Here s_1^j (s_2^j) is individual j's strategy when assigned player role 1 (role 2). Analogously, one can define individual j's type vector t^j:

$$t^j = (t_1^j, t_2^j).$$

The vectors s^j and t^j will become important when we analyze G^{E} which is the direct evolutionary game associated with Γ (see below). However, we drop the superscript index for the individual whenever this should not cause confusion. In solving the asymmetric game Γ we solve

$$G^{\mathrm{B}}(t_1, t_2) \equiv ((S_i, u_i)_{i=1,2})|_{(t_1,t_2)}$$

for all $t = (t_1, t_2) \in T$. The solution is $s^*(t) = (s_1^*(t), s_2^*(t))$ for all $t \in T$.

The equilibrium strategy profile $s^*(t)$ can be inserted into $r_i(s)$ to give the indirect evolutionary success function $r_i^*(t) = r_i(s^*(t))$. Accordingly the game

$$G^{\mathrm{E}} \equiv ((T_i, r_i^*)_{i=1,2})$$

associated with Γ is now a direct evolutionary game with role asymmetry. In solving G^{E} (and thus Γ) we therefore have to apply methods that are suited for asymmetric games (see for example Weibull (1995, pp. 64) or Selten (1980)). We consider two individuals (players) $j = 1, 2$ who are characterized by their type vectors $t^1 = (t_1^1, t_2^1)$ and $t^2 = (t_1^2, t_2^2)$, respectively. We call $(t^1 \times t^2)$ a pairing and (t_1^1, t_2^2) the match in which player 1 is assigned to role 1 and player 2 is assigned to role 2. Accordingly, (t_1^2, t_2^1) is the match in which player 1 has role 2 and player 2 has role 1. We can distinguish the equilibrium strategy profiles (of the base games $G^{\mathrm{B}}(t)$) for both matches as

$$s^*(t_1^1, t_2^2) \qquad \text{and} \qquad s^*(t_1^2, t_2^1)$$

and furthermore the indirect evolutionary success of each player in each match:

© Blackwell Publishers Ltd 2000

$$r_1^*(t_1^1,\ t_2^2), \qquad r_2^*(t_1^1,\ t_2^2)$$

and

$$r_1^*(t_1^2,\ t_2^1), \qquad r_2^*(t_1^2,\ t_2^1).$$

The indirect evolutionary success of player 1, respectively player 2, in the pairing $(t^1 \times t^2)$ will thus be determined as the expected success of both matches:[13]

$$r^{1*}(t^1 \times t^2) \equiv \tfrac{1}{2} r_1^*(t_1^1,\ t_2^2) + \tfrac{1}{2} r_2^*(t_1^2,\ t_2^1)$$

and

$$r^{2*}(t^1 \times t^2) \equiv \tfrac{1}{2} r_1^*(t_1^2,\ t_2^1) + \tfrac{1}{2} r_2^*(t_1^1,\ t_2^2).$$

Note that $r^{1*}(t^1 \times t^2) = r^{2*}(t^2 \times t^1)$; thus, by assuming that each match is equally likely and determining evolutionary success as the expected success of both matches we have essentially removed the asymmetry. Without loss of generality we can now interpret $r^*(t^k \times t^l) \equiv r^{1*}(t^1 \times t^2)$ with $t^k,\ t^l \in T = T_1 \times T_2$ as the evolutionary success of an individual who is endowed with type vector t^k within a population of t^l-individuals.[14] We are now ready to look for a solution of G^{E} according to, for example, the notion of ESS or some other static or dynamic solution concept. Thus, a solution of Γ (when a unique solution exists) will be a type vector $t^{k*} = (t_1^{k*},\ t_2^{k*})$ (i.e. the solution of G^{E}) together with the associated solution strategies $s_1^*(t_1^{k*},\ t_2^{k*})$ and $s_2^*(t_1^{k*},\ t_2^{k*})$. This illustrates that the solution of asymmetric games fits into the general definition of a solution for indirect evolutionary games as given above.

We want to remark that the formal requirement that an individual is endowed with a type vector t^j does not necessarily imply role-dependent types nor that T_1 and T_2 have to be regarded as two separate mutation spaces where a mutation of the role 1 type is independent of a mutation of the role 2 type. For instance, a quite plausible restriction is to require $t_1^j = t_2^j$, i.e. an individual's type is independent of its role. It was applied by Königstein (2000) for example in a study on the evolution of altruism

[13] This is the usual assumption in direct evolutionary games with role asymmetry.

[14] Note that t^k, t^l now represent elements of T while in symmetric games these variables were used to denote elements of T_1.

© Blackwell Publishers Ltd 2000

within an asymmetric bargaining game. The restriction that the degree of altruism is independent of an individual's role within a bargaining procedure seems reasonable. Formally, such a restriction means that $t^j \in T^j \subseteq T = T_1 \times T_2$; i.e. the mutation space $T^j = \{t^j | t^j \in T, t_1^j = t_2^j\}$ for individual types t^j is a subspace of T.

2.6 Example 2: Asymmetry and replicator dynamics

The following example serves two purposes: it shows how to handle asymmetric games and it shows how dynamic solution concepts can be applied. We consider the game shown in figure 1 and, following Gale *et al.* (1995), we refer to it as the 'ultimatum minigame'. In this game 'role 1' first proposes how to divide a pie of size $c > 0$.[15] The choice labeled F represents a fair offer, inducing an equal split of c. The choice U (unfair offer) results in an uneven split of c given that 'role 2' subsequently chooses A (accept). In this case, 1 earns $(1 - \varepsilon)c$ while 2 gets εc with $0 < \varepsilon < \frac{1}{2}$. If, instead, 2 chooses R (reject), then both roles earn nothing. These payoffs are monetary payoffs and they are assumed

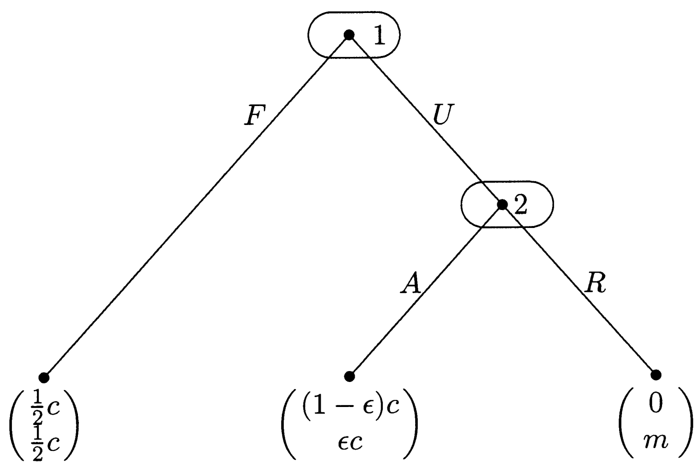

Figure 1: Ultimatum minigame

[15] We use the term 'role *i*' instead of 'player *i*' here to avoid confusion; remember that a player in an indirect evolutionary game is an individual person out of a large population who gets assigned either of the two roles.

© Blackwell Publishers Ltd 2000

to be the measure of evolutionary success. However, each individual has a utility function of the following kind: in the case of rejection the role 2 utility is t_2 with $t_2 \in \{\underline{m}, \overline{m}\}$ and $0 \leqslant \underline{m} \leqslant \overline{m}$, whereas the role 1 utility is $t_1 = 0$ (i.e. it is equal to monetary payoff for all individuals). For all other cases utility is equal to monetary (and thus evolutionary) payoff. Thus figure 1 shows utility payoffs. The parameters $\underline{m}$ and $\overline{m}$ can be thought of as measuring 2's 'feeling of revenge' when rejecting an unfair offer: if $t_2 = \overline{m}$ this means that the individual acting in role 2 has a strong feeling of revenge; otherwise it has a weak feeling of revenge ($t_2 = \underline{m}$). To make things interesting we will assume furthermore $\underline{m} < \varepsilon c < \overline{m}$.

With these definitions we have essentially modeled an asymmetric two-player indirect evolutionary game with observed types $\Gamma = ((S_i, T_i, u_i, r_i))_{i=1,2}$. For clearer exposition we will write down its components explicitly:

$$S_1 = \{F, U\} \qquad S_2 = \{A, R\}$$

$$T_1 = \{0\} \qquad T_2 = \{\underline{m}, \overline{m}\}$$

$$(u_1, u_2) = \begin{cases} (\tfrac{1}{2}c, \tfrac{1}{2}c) & \text{for} \quad (s, t) \in \{(F, A), (F, R)\} \times \{(0, \underline{m}), (0, \overline{m})\} \\ ((1-\varepsilon)c, \varepsilon c) & \text{for} \quad (s, t) \in \{(U, A)\} \times \{(0, \underline{m}), (0, \overline{m})\} \\ (0, t_2) & \text{for} \quad (s, t) = \{(U, R)\} \times \{(0, \underline{m}), (0, \overline{m})\} \end{cases}$$

$$(r_1, r_2) = \begin{cases} (\tfrac{1}{2}c, \tfrac{1}{2}c) & \text{for} \quad s \in \{(F, A), (F, R)\} \\ ((1-\varepsilon)c, \varepsilon c) & \text{for} \quad s = (U, A) \\ (0, 0) & \text{for} \quad s = (U, R) \end{cases}$$

To solve the game Γ we first solve the associated base games $G^B(t_1, t_2)$ for the equilibrium strategies.[16] Rationality (i.e. utility maximization given individual types) implies the following subgame perfect equilibrium strategies:

$$(s_1^*(t_1, t_2), s_2^*(t_1, t_2)) = \begin{cases} (U, A) & \text{for} \quad t_2 = \underline{m} \\ (F, R) & \text{for} \quad t_2 = \overline{m} \end{cases} \tag{2.5}$$

[16] Since T_1 is a singleton we do not need t_1 as an identifier of a generic element of T_1. But, for the sake of exposition, we will use it nevertheless.

© Blackwell Publishers Ltd 2000

This describes the solution for every base game $G^{B}(t_1, t_2)$. By evaluating $r_i(s)$ at $s^*(t)$ we get the asymmetric direct evolutionary game $G^{E} = ((T_i, r_i^*)_{i=1,2})$ with the indirect evolutionary success functions

$$(r_1^*(t_1, t_2), r_2^*(t_1, t_2)) = \begin{cases} ((1-\varepsilon)c, \varepsilon c) & \text{for} \quad t_2 = \underline{m} \\ (\frac{1}{2}c, \frac{1}{2}c) & \text{for} \quad t_2 = \overline{m} \end{cases} \qquad (2.6)$$

Since G^{E} is asymmetric we consider player $j = 1, 2$ as being endowed with type vector $t^j = (t_1^j, t_2^j) = (0, t_2^j)$. Pairing $(t^1 \times t^2)$ results in one of the two possible matches $(t_1^1, t_2^2) = (0, t_2^2)$ or $(t_1^2, t_2^1) = (0, t_2^1)$ with each match being equally likely. The 'ultimatum minigame with random role assignment' is illustrated in figure 2.

The indirect evolutionary success of player 1 in a pairing with player 2 $(r^{1*}(t^1 \times t^2))$ is described by the following matrix where rows refer to t^1 and columns refer to t^2:

	$(0, \underline{m})$	$(0, \overline{m})$
$(0, \underline{m})$	$\frac{1}{2}c$	$\frac{1}{4}(1+2\varepsilon)c$
$(0, \overline{m})$	$\frac{1}{4}(3-2\varepsilon)c$	$\frac{1}{2}c$

(2.7)

The cell entries represent player 1's evolutionary success in a pairing

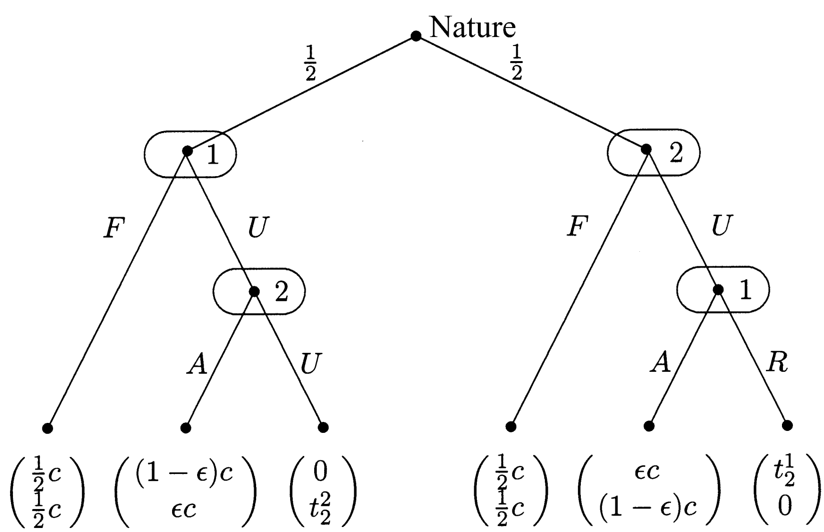

Figure 2: Ultimatum minigame with random role assignment

© Blackwell Publishers Ltd 2000

with 2. To show how to get these formulae we derive $r^{1*}(t^1 \times t^2) = r^{1^*}((0, \underline{m}) \times (0, \overline{m}))$, i.e. the evolutionary success of player 1 with $t_2^1 = \underline{m}$ when paired with player 2 with $t_2^2 = \overline{m}$. The other payoffs can be calculated similarly. Note first that player 1's role 1 success $r_1^{1*}(t_1^1, t_2^2)$ is determined in match (t_1^1, t_2^2) by the strategy profile $(s_1^{1*}, s_2^{2*}) = (F, R)$ (see (2.5)); the result is $r_1^{1*} = \frac{1}{2}c$ (see (2.6)). Player 1's role 2 payoff r_2^1 is determined in match (t_1^2, t_2^1) by the strategy profile $(s_1^{2*}, s_2^{1*}) = (U, A)$ (see (2.5), and note that in this model it is the type of the role 2 player that determines the equilibrium strategies); the result is $r_2^{1*} = \varepsilon c$ (see (2.6)). Therefore $r^{1*}((0, \underline{m}) \times (0, \overline{m})) = \frac{1}{2}r_1^{1*} + \frac{1}{2}r_2^{1*} = \frac{1}{4}(1 + 2\varepsilon)c$.

To illustrate the use of a dynamic solution concept we apply the replicator dynamics to solve G^{E} (and thereby Γ). Let the state of the population at time τ be $x(\tau) = (x_{\underline{m}}(\tau), x_{\overline{m}}(\tau)) = (x_{\underline{m}}(\tau), 1 - x_{\underline{m}}(\tau))$ where $x_{\underline{m}}(\tau)$ is the population share of (individuals endowed with) type vectors $(0, \underline{m})$ at time τ.[17] With the payoff matrix given in (2.7) the replicator dynamics (2.3) become

$$\begin{aligned} \dot{x}_{\underline{m}} &= \tfrac{1}{4}c(1 - 2\varepsilon)x_{\underline{m}}(x_{\underline{m}} - 1) \\ \dot{x}_{\overline{m}} &= -\dot{x}_{\underline{m}} \end{aligned} \tag{2.8}$$

For all $x_{\underline{m}}(0) > 0$ the solution of the system of ordinary differential equations in (2.8) is given by

$$x_{\underline{m}}(t) = \frac{1}{1 + c_1 \exp(at)} \qquad x_{\overline{m}}(t) = 1 - x_{\underline{m}}(t)$$

where $c_1 = [1 - x_{\underline{m}}(0)]/x_{\underline{m}}(0) \geqslant 0$ and $a = \frac{1}{4}c(1 - 2\varepsilon) > 0$ since $0 < \varepsilon < \frac{1}{2}$.

The population states $x = (1, 0)$ and $x = (0, 1)$ are the only fixed points of system (2.8). Since for every $x_{\underline{m}}(0) > 0$ it holds that $c_1 > 0$, it follows that $\lim_{t\to\infty} x_{\underline{m}}(t) \to 0$, i.e. the population state (0, 1) is globally stable. Thus, if initially only a single type vector is present in the population, this state does not change over time as long as no other type vector appears via mutation. In every mixed population the $(0, \overline{m})$-individuals are more successful than the $(0, \underline{m})$-individuals. Therefore the population share of the latter will converge to 0 over time. In the long

[17] Note here that for convenience we use $\underline{m}$ and $\overline{m}$, respectively, as a shorthand for $(0, \underline{m})$ and $(0, \overline{m})$, respectively.

© Blackwell Publishers Ltd 2000

run the population will consist only of $(0, \overline{m})$ individuals, i.e. of players that exhibit strong feelings of revenge when rejecting an unfair offer. Furthermore, the associated strategy choices are such that only fair offers occur and that unfair offers will be rejected. We summarize these results by formulating the following proposition.

Proposition 2. The population state $x = (x_{\underline{m}}, x_{\overline{m}}) = (0, 1)$ is globally stable in the replicator dynamics, i.e. for all completely mixed initial conditions the population will consist only of $(0, \overline{m})$-individuals for $t \to \infty$.

Note that Huck and Oechssler (1999) also investigate a version of the ultimatum minigame using the indirect evolutionary approach. They assume a finite population of players interacting in small groups and relax the assumption of observable types. Under specific assumptions regarding the dynamics used they show that if the maximal group size is sufficiently small almost all proposers will offer the fair split in the long run (for arbitrary initial conditions). So, their result is in line with the one above.

3. DISCUSSION

We summarize that the IEA allows us to model endogenous preferences and to study their evolution.[18] The reproduction of preferences depends on the strategy choices which they induce. Therefore, evolutionary success depends indirectly rather than directly on the preference types. We have presented a unifying description of two-player indirect evolutionary games with observed types, which clarifies the conceptual links between different applications one finds in the literature. The IEA is a combination of existing methods of modeling and analyzing human behavior. We have shown how to model such games and how the solution concepts known from game theory and (direct) evolutionary game theory can readily be applied.

The process of evolution is not necessarily to be interpreted as biological evolution. One might just as well think of it as social evolution: ‘Memes (ideas, learning rules, behavioral norms, etc.) are just as much the object of evolutionary pressures as genes, but memes

[18] There are other approaches to explaining the change of preferences. For instance, see Bowles (1998) for a recent review of models and evidence on the impact of economic institutions on preferences. We see this work as complementary.

© Blackwell Publishers Ltd 2000

multiply through *imitation* [italics in original] rather than physical replication' (Binmore (1988, p. 16), paraphrasing Dawkins (1976)). This interpretation of evolution may lead to applications of the IEA which we find quite appealing: namely, to model the development of social norms within a society, and to show how this depends on the strategy choices which the norms induce, e.g. a norm may be thought of a special kind of preference. Within social interaction, individuals are not programmed to choose certain actions independent of their own norms and those of other interacting persons, but may derive choices taking these norms into account. Thus, in the short run, human behavior depends on existing norms. However, in the long run, the norms themselves may change. The spreading of new (and old) norms within a society may depend on their material (e.g. monetary) success. In our opinion it is especially this view upon the dynamics of social behavior which renders the IEA a useful modeling tool: the combination of rational behavior and social evolutionary dynamic.

One might wonder whether the change of preferences or norms could also be modeled as rational choice rather than evolution. We acknowledge that this may be possible. But there are phenomena (like changes in fashion, tastes, corporate culture etc.) that seem better described, for instance, as imitation processes rather than rational choice. Furthermore, evolutionary solution concepts and rational choice concepts do not coincide, in general.

On the other hand, from the perspective of direct evolutionary game theory one might wonder about substituting the rational choice part of the IEA by another evolutionary process. Namely, rational choice based on preference types as it is captured by the equilibrium strategy $s_i^*(t)$ can be thought of as a behavioral rule which is parametrized by t. Instead of deriving $s_i^*(t)$, as we did here, one could start out by modeling a space of such rules, which would be the usual direct evolutionary approach. Huck and Oechssler (1999) show for a simple example how this can be achieved. However, they also show that the solution of the direct evolutionary model and the solution of the indirect evolutionary game do not necessarily coincide. Moreover, we view the incorporation of rational choice as a structural advantage of the IEA compared with pure evolution. After all, human beings are endowed with a cognitive system that allows for behavioral adjustments based on reasoning.

It may nevertheless be fruitful to think of relaxing the assumptions underlying the rational choice part of solving an indirect evolutionary game. For instance, following Binmore (1988) one might attempt 'to

© Blackwell Publishers Ltd 2000

model the thinking processes of players *explicitly'* (p. 10, italics in original). This leads to a dynamic process by which equilibrium may be reached, and which Binmore refers to as 'eductive libration' (p. 11). He distinguishes this process from that of 'evolutive libration', which represents the equilibrating processes assumed by the direct evolutionary approach. He argues that there are two kinds of environments (eductive or evolutive) in which the one or the other process applies. Assuming rational choice, as we do here, and consequently applying game theoretic equilibrium concepts can be thought of as one (extreme) form of eductive libration; one might very well think of other specifications. Furthermore, the IEA suggests that eductive and evolutive environments may not exclude each other. Rather, both types of libration processes may be present at the same time. Specifically, the IEA might be extended by substituting the rational choice part by eductive libration.

In the examples above, we assumed that evolutionary success is represented by monetary payoff. This is not required to apply the IEA. The formalism allows for any other specification of the evolutionary success function. However, all applications we are aware of used monetary payoff as the success measure. Furthermore, since income levels can be observed relatively easily within a society this seems a natural specification, especially within economic models.

Finally, we want to point to two important restrictions of our presentation: we assumed infinite populations and observability of preference types. Evolution in finite populations and/or imperfect observability of types requires a richer description of an indirect evolutionary game than we allowed. For applications of such models see, for example, Huck and Oechssler (1999) or Güth and Kliemt (1994).

REFERENCES

Akin E. (1982): "Exponential families and game dynamics", *Canadian Journal of Mathematics*, 34, pp. 374–405.

Bester H., Güth W. (1998): "Is altruism evolutionarily stable?", *Journal of Economic Behavior and Organisation,* 34 (2), pp. 193–209.

Binmore K. (1988): "Modeling rational players, Part II", *Economics and Philosophy*, 4, pp. 9–55.

Binmore K. and Seymour R. (1995): "Stationary distributions of noisy replicator dynamics in the ultimatum game", mimeo, University College London.

Bomze I. (1991): "Cross entropy minimization in uninvadable states of complex populations", *Journal of Mathematical Biology*, 30, pp. 73–87.

Bowles S. (1998): "Endogenous preferences: the cultural consequences of markets and other economic institutions", *Journal of Economic Literature*, 36 (1), pp. 75–111.

© Blackwell Publishers Ltd 2000

van Damme E. (1991): *Stability and Perfection of Nash Equilibrium*, 2nd edn, Springer, Berlin.
Dawkins R. (1976): *The Selfish Gene*, Oxford University Press, Oxford.
Friedman D., Yellin J. (1996): "Evolving landscapes for population games", mimeo, University of California at Santa Cruz.
Gale J., Binmore K. G., Samuelson L. (1995): "Learning to be imperfect: the ultimatum game", *Games and Economic Behavior*, 8, pp. 56–90.
Güth W., Huck S. (1997): "A new justification of monopolistic competition", *Economics Letters*, 57, pp. 177–182.
Güth W., Kliemt H. (1994): "Competition or co-operation—on the evolutionary economics of trust, exploitation and moral attitudes", *Metroeconomica*, 45, pp. 155–187.
Güth W., Yaari M. (1992): "An evolutionary approach to explain reciprocal behavior in a simple strategic game", in Witt U. (ed.): *Explaining Process and Change—Approaches to Evolutionary Economics*, The University of Michigan Press, Ann Arbor, MI, pp. 23–34.
Hammerstein P., Selten R. (1994): "Game theory and evolutionary biology", in (Aumann R. J., Hart S. (eds)): *Handbook of Game Theory*, vol. 2, North-Holland, Amsterdam, pp. 929–993.
Harsanyi J. C., Selten R. (1988): *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, MA.
Hopkins E., Seymour R. (1996): "Price dispersion: an evolutionary approach", mimeo, University College London.
Huck S., Oechssler J. (1999): "The indirect evolutionary approach to explaining fair allocations", *Games and Economic Behavior*, 28, pp. 13–24.
Königstein M. (2000): "Efficiency and evolution of social preferences and prosocial behavior", in: Equity Efficiency and Evolutionary Stability in Bargaining games with Joint Production, Springer-Verlag.
Maynard Smith J. (1982): *Evolution and the Theory of Games,* Cambridge University Press, Cambridge.
Oechssler J., Riedel F. (1998): "Evolutionary dynamics on infinite strategy spaces", Discussion Paper 68, SFB 373, Humboldt-Universität Berlin.
Selten R. (1980): "A note on evolutionarily stable strategies in asymmetric animal conflicts", *Journal of Theoretical Biology*, 84, pp. 93–101.
Thomas B. (1985): "Evolutionarily stable sets in mixed-strategist models", *Theoretical Population Biology*, 28, pp. 332–341.
Vega-Redondo F. (1996): *Evolution, Games, and Economic Behaviour*, Oxford University Press, Oxford.
Weibull J. W. (1995): *Evolutionary Game Theory*, MIT Press, Cambridge, MA.
Zeeman E. (1981): "Dynamics of the evolution of animal conflicts", *Journal of Theoretical Biology*, 89, pp. 249–270.

Institute for Economic Theory III
Spandauer Strasse 1
10178 Berlin
Germany
E-mail: mkoenig@wiwi.hu-berlin.de; and wmueller@wiwi.hu-berlin.de

© Blackwell Publishers Ltd 2000