The evolution of cooperation in Prisoners' Dilemma with an endogenous learning mutant

Carsten Vogt*

Faculty of Economics and Management, Otto-von-Guericke-University Magdeburg,
P.O. Box 4120, D-39016 Magdeburg, Germany

Abstract

We consider a population initially consisting of cooperators and defectors, who are engaged in a Prisoners' Dilemma. A knowledgeable mutant is introduced who memorizes other players by experience. We derive a formal model which allows us to describe the growth of the mutant's knowledge of defectors in a precise way. When meeting known defectors, the mutant is assumed to avoid exploitation by refusing interaction. Otherwise the mutant chooses the cooperative action. According to replicator dynamics there exists a fully mixed equilibrium, a two-type equilibrium with sophisticated cooperators and defectors and the defectors-only corner solution. Depending on the parameters, these equilibria are globally stable. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

During the last few decades much experimental work has been carried out in order to get an idea of how people actually behave and what the driving forces behind their behavior are. Experimental economics tells us that in a number of important situations the standard model of homo economicus fails to explain what really goes on. In particular, in dilemma games it has been shown over and over again that, contrary to the conventional wisdom of game theory, there is a significant level of cooperative behavior. [1] Moreover (and more importantly), it has been pointed out by several experimentalists that there might exist a host

---

* Tel.: +49-391-6718801; fax: +49-391-6712971.

E-mail address: carsten.vogt@ww.uni-magdeburg.de (C. Vogt)

[1] For an excellent overview see Ledyard (1995).

of different types of behavior. Many experiments show that individuals differ significantly in their willingness to contribute to the voluntary provision of a public good. Weimann (1994) concludes that there are at least three different types present: highly cooperative individuals, free riders, and weak free riders, i.e. subjects who sometimes contribute but often do not. Isaac et al. (1994) (henceforth IWW) report a very similar observation. In their paper they present the behavior of three individuals which they call typical for many of their experiments. Again, roughly stylized, there is one type that is highly defective, a second type that tends to cooperate at a high level (i.e. contributes often and at high efficiency levels) and a third type that shows intermediate contribution levels that decline over time. Given the extraordinary significance of these observations (the IWW-experiment involved 1908 subjects!), the existence of different behavioral types seems to be well evidenced.

A look at Prisoners' Dilemma experiments lends strong support to the view gained from public good experiments. A study which explicitly addresses the question of different types is Andreoni and Miller (1993). In their paper the sequential equilibrium hypothesis of Kreps et al. (1982) is tested. Apart from the question of whether people try to build up their reputation, Andreoni and Miller are also strongly interested in whether there really do exist (conditional) cooperative types [2] (in theory it suffices if one player believes the other to be a tit-for-tat player with very small probability). Andreoni and Miller examined four treatments. Within the 'Partners'-scheme, subjects played 20 runs, where each run consisted of a 10-period repeated Prisoners' Dilemma. Partners did not change within a 10-period game. 'Strangers' played 200 single-shot PDs. 'Computer50' subjects had a 50 percent chance of being paired with a computer that played tit-for-tat. 'Computer0' was the same, but the probability of being paired to the machine was only 0.1 percent. If the subject pool consisted of defectors only, and subjects learned this while playing the game, in the partners and Computer50 treatments, one would expect cooperation to break down at increasingly earlier stages of the game. On the other hand, if there really do exist players who are willing to reciprocate to cooperative behavior, this may be learned by players during the course of action. Updating their beliefs about the frequency of such conditional cooperative types may then even lead to an increase of cooperation at later stages of the game. This, indeed, is what happened. 'Partners' as well as 'Computer50's waited longer before they started to defect. Andreoni and Miller argue that this observation is "consistent with a hypothesis of altruism in which subjects continue to update their priors on the degree of altruism in the population throughout the experiment" (Andreoni and Miller, 1993, p. 579). Moreover, there was a persistent degree of cooperative behavior even in the 'Strangers' treatment, i.e. subjects cooperated throughout the repeated one-shot PD game. This striking stability led Andreoni and Miller to conjecture that subjects play a Nash equilibrium with incomplete information about 'altruism' (again, what is meant here is the presence of conditionally cooperative players, C.V.) of their opponents. This equilibrium hypothesis implies the existence of three different types: defectors, cooperators and subjects who play a mixed strategy. An inspection of the data reveals that these three types are indeed present. Moreover, the theoretical equilibrium predictions strongly agree

---

[2] Andreoni and Miller term these subjects 'altruists'. This label is at least unusual in the literature. What essentially is meant by 'altruism' here is a kind of behavior which reciprocates, i.e. that conditionally cooperates like, for example, tit-for-tat.

with the observations. Similar results, i.e. evidence of reputation building as well as of conditional cooperative types, can be found in Camerer and Weigelt (1988). The interaction of preferences for cooperation with strategic motives has been evidenced by Palfrey and Rosenthal (1988). For an overview of PD-experiments see Roth (1988) and Rapoport and Chammah (1965). For a recent study see Cooper et al. (1996).

In psychology, it is a much more accepted view that people differ with respect to what is called their value orientation. Often the observations lead to a categorization of people as being competitors, individualists, cooperators, altruists, or aggressors, with the last two categories having rarely been observed. Recently Offerman et al. (1996) argued that it might be misleading to conclude the existence of distinct types from the observations in public good games. They demand an independent measure of value orientation. They achieve this in their experiment by playing some kind of dictator game repeatedly before playing a step-level public good game. During the experiment, subjects had to decide how to divide a certain amount of money between themselves and a second player who could not react to this decision. The procedure was carried out 24 times. Their main findings are: (i) that subjects decided highly consistently; (ii) that they differed in their basic value orientations (i.e. there are different types — particularly individualists and cooperators) and (iii) that these differences prevailed in the public good game. Hence, it is clearly necessary to concede the existence of types and to take into account that this fact influences the outcomes of, for example, public good games (but we suspect that this fact matters for other games, too).

In theory, it is no problem to keep the existence of types in mind. Of course, one might build models of repeated Prisoners' Dilemma games that assume distinct types and then study their characteristics, possibly adding strategic play (so that subjects, for strategic reasons, may deviate from more or less their value orientation in the course of action). However, the main question clearly is how to explain the existence of different types. This task can hardly be accomplished in a standard game theoretical framework. One tool for tackling such a problem is evolutionary game theory. The core of evolutionary theorizing is to assume the existence of a host of different behavioral types and then consider which of these are viable, i.e. which are the survivors of selection. This, in fact, is also the approach of this paper. We will introduce a new, mutational behavioral type into a population initially consisting of defectors and cooperators alone and then look at what consequences this has for the equilibria in this modified Prisoners' Dilemma.

Traditional game theory stresses the point that cooperation in a Prisoners' Dilemma game requires some kind of punishment of defectors. The Folk Theorem results (Rubinstein, 1979, Fudenberg and Maskin, 1986) make use of minimaxing strategies. Axelrod (1981, 1984) showed in his famous computer tournaments that tit-for-tat is quite a successful strategy. A third contribution to this topic is the gang-of-four model (Kreps et al., 1982) which only requires one player to be uncertain about the identity of his opponent. In a finitely repeated Prisoners' Dilemma, if the opponent is believed to be a tit-for-tat player, cooperation occurs until close to the end of the game. Thus, reciprocity seems to be quite a powerful principle to explain cooperation.

However, for reciprocity to take place, it is necessary for players to be able to make their decisions contingent on some information about the identity of their opponent or on past history of play. Kandori (1992) is concerned with the problem of how cooperation could be sustained when interaction is infrequent, i.e. players change their partners over

time. He shows that for small communities, the Folk Theorem can be restated. This is so because in this case each member can observe the behavior of other group members. "This observation shows that changing partners itself is unimportant and the crux of the matter is information transmission among the community members" (Kandori, 1992, p. 64). For large communities, the assumption of public observability is no longer reasonable. If one drops this assumption, one has to deal with the problem that "each agent typically possesses a piece of information about the history of trades in the community which may not be known to others" (Kandori, 1992, p. 64). But the description of how this knowledge grows over time may be very difficult to formalize. At this point of theorizing, two possibilities arise. Either one tries to model the growth of knowledge in matching games — this is the approach adapted in the present paper, but we embed the problem in an evolutionary framework — or one tries to avoid this problem of modeling the development of information and concentrates on extreme cases and the conditions under which cooperation is achievable. This is Kandori's approach. He considers a situation where some local information about the opponent's previous action and his label (a signal about his type) is available. But the way this knowledge is generated has not yet been modeled. Moreover, it is assumed that the signal works perfectly, i.e. a player's type is revealed honestly.

Milgrom et al. (1990) argue that, in the case of large groups, information about a player's defective behavior in the past might be provided by an institution (e.g. a court). In this paper, we offer an alternative approach. We introduce a so-called knowledgeable mutant who is able to identify his opponents. But this knowledge does not exist as a matter of course. Instead, it has to be acquired by interaction with other players. Experience is the only way to gain some insight into the nature of other subjects and into the composition of the population. This is also the main difference to Amann and Yang (1994). They also introduce a knowledgeable mutant but they assume this type to be perfectly informed. That is, whenever their mutant meets a defector, he is detected independently of whether this player was met before or not. This view is, of course, rather optimistic because the mutant can never be exploited by non-cooperative players.

The idea of introducing a sophisticated mutant is not really new. Robson (1990) considers a model where cooperators could identify each other, form an insider group and share the gains from cooperation. There are two main differences to our model. First, in Robson's model, the generation of information remains unexplained since the availability of a signal is simply assumed. Second, his theory is unable to explain the coexistence of cooperative and defective types because the mutant strategy is the only evolutionarily stable strategy (ESS). Apart from these two shortcomings, there is a further problem. A nasty sophisticated mutant is able to erode the original ESS, so that in the long run defection has to be expected.

In a mixture of evolutionary and rationalistic arguments, Robert Frank (1987) proposed a model where a type signal is available which, in general, one cannot perfectly rely on. He starts his work with the statement that the problem of achieving cooperative outcomes in Prisoners' Dilemma situations can mainly be viewed as a commitment problem, as has already been pointed out by Schelling (1960). If it were possible for the players to commit themselves to a conditionally cooperative strategy, they would be able to avoid the defective outcome. But, of course, there immediately arises the problem of credibility of such commitments. Frank's favorite candidate for a workable commitment device is human conscience coupled with strong emotions, like the feeling of guilt whenever a promise

is being broken. Technically speaking, there might be internal negative payoffs involved which alter the payoff matrix of the PD game. But, as Frank puts it himself, "merely having a conscience does not solve the commitment problem; one's potential trading partner must also know about it" (Frank, 1987, p. 594). This leads him to the assumption that there exists an observable signal which allows subjects to identify their potential partners as more or less trustworthy. In addition, this signal must (at least to some degree) be independent of individual control. The major point to be made here is that there must be some kind of information which allows the identification of the opponent's behavioral type. This indeed constitutes the main similarity to our model. However, while Frank more or less assumes the existence of such information, justified only by an implicit evolutionary argument, we will give a systematic derivation of this kind of knowledge.

Our work is also related to a paper by Guttman (1996). He shows that mechanical tit-for-tat types might coexist with purely rational, optimizing actors in an evolutionary equilibrium when optimizing is costly. This resembles our main result, namely a stable coexistence of naive defectors, naive cooperators, and sophisticated cooperators.

Let us sum up the idea of this work. We strongly believe that a satisfactory theory of human behavior not only has to explain cooperation per se but, moreover, has to give an explanation of the coexistence of different behavioral types. We take this as a stylized fact from several experimental observations which we try to explain within an evolutionary framework in which a sophisticated mutant is introduced. Extending previous work, our aim is to develop a simple model of endogenous learning for evolutionary matching games. In doing so we also provide a model which principally shows how reciprocity within larger groups might work.

The paper is organized as follows. In Section 2.1 we give the main results of our learning model. A formula is derived which allows the precise calculation of the probability of meeting a known defector. Section 2.2 deals with the existence and stability of equilibria in the extended evolutionary Prisoners' Dilemma. Particularly, we are able to show the existence of a globally stable interior population state. Section 3 contains some concluding remarks.

2. The Prisoners' Dilemma with a knowledgeable mutant

We consider a population whose members initially show two behavioral patterns. Either they are cooperative players, or they are defectors, i.e. players who never show any cooperative behavior. Let us label the first ones 'Doves' and the latter ones 'Hawks'. In this paper we are following an direct evolutionary game-theoretical approach [3], which means that we

---

[3] Besides this branch of theorizing there exists the so called indirect evolutionary approach, which aims to bridge the gap between the strong rationality assumptions of conventional economic theorizing and the pure adaptational behavior in direct evolutionary models. Players there are not modelled as automata but choose their actions in the basic games fully rational. Thus, standard solution concepts of game theory can be applied to solve this game. However, it is assumed that the game 'is embedded in an evolutionary process in which some of the parameters or rules of the games change' (Güth and Kliemt (1998), p. 380) The main idea is that rational choices taken in the basic game and the rules of the game influence each other interdependently. The long run behavior of the resulting dynamic process can then be analyzed with tools borrowed from (direct) evolutionary game theory. See for this branch of the literature Güth and Yaari (1992), Güth and Kliemt (1995), Güth and Kliemt (1996), Güth (1997), Güth et al. (1998), Güth and Kliemt (1998), Guttman (1996).

Table 1
An original Prisoners' Dilemma

	Dove	Hawk
Dove	3	0
Hawk	6	1

model the players as automata. Behavior is programmed for a certain period of time during which the agents cannot deviate from their prescribed strategies, so they are not optimizing at all.

This paper deals with the well-known Prisoners' Dilemma as given in Table 1. As usual, we assume that, at each point of time, the members of the population are chosen randomly and are sorted into pairs. If two Doves are chosen and play the game, they cooperate mutually. If two Hawks meet, they both defect and end up in the socially undesirable outcome. The most favorable possibility for a Hawk is to be paired with a Dove. In this case he can exploit the cooperative player.

As is well known, in the normal form game $\{H, H\}$ is the unique, dominant Nash-equilibrium. In an evolutionary setting, it is easy to see that the only equilibrium is the one in which the population entirely consists of Hawks. This is so because Hawk behavior always yields the higher payoff, and so the Doves are driven to extinction. But matters may change if one introduces other patterns of behavior into the population. So let us assume that a new mutant type appears on the scene who is able to learn something about his environment.

The introduction of a knowledgeable type (in short, the K-type) into the PD game has also been proposed by Yang (1994) and Amann and Yang (1994) in two unpublished working papers. In their models, they assume the K-type to be able to detect whether his opponent is a Hawk or a Dove. In order to get the information about his opponent, the mutant has to sacrifice some amount of fitness. However, in bearing the investigation-costs, he gets perfect information about his opponent. This seems to be a very strong assumption and there are doubts as to whether their optimistic results — the existence and stability of a fully mixed equilibrium — would hold under more realistic assumptions. The investigation of these topics is the primary purpose of this paper.

The importance of endowing players in economic models with the ability to learn has been stressed by several authors (Selten, 1991; Börgers, 1996). In the models mentioned above, the question of how the K-type acquires his knowledge is left open, i.e. a learning model is missing. In addition, it seems to be somewhat artificial to assume that the mutant gets the information about his opponent's player-type before any interaction has taken place.

Therefore, we need a model of endogenous learning. [4] In particular, we think a convincing model of learning has to take into account that people learn by experience, including the possibility of bad experience.

---

[4] As Yang puts it himself, "a systematic derivation of information provision for K-type via specific adaptations may be desirable, as a better, less ad hoc, motivation" (Yang, 1994, p. 20). This is exactly the point at which our work starts.

2.1. The learning model

Knowledge in our paper is modeled as a memory which is built up during play. This means the K-type is randomly matched to some other member of the population and, after interaction has taken place, he 'stores' precisely the information concerning this individual in his brain. In the subsequent rounds of the game he is then able to detect whether his opponent is someone whom he has met in some earlier round or someone who, up to this point in time, is unknown to him. In other words, once he has met an individual, the K-type is able to identify this individual in later rounds when he might possibly meet this actor again.

We assume the K-type-mutant has an initial tendency to play cooperatively. The mutant uses the cooperative strategy against all players who are also cooperative (i.e. Doves, or later in time when reproduction has taken place, perhaps other K-types) and against Hawks as long as he does not perceive them to be defectors. This means the mutant always behaves cooperatively, except in those cases where a Hawk is known to him. When meeting a known defector, the K-type is assumed to refuse interaction, as is explained in more detail later on. Therefore, all we need is a model of how the K-type's knowledge of Hawks grows.

Suppose then that there are $t = 0, 1, \ldots, T < \infty$ matching rounds, where all members of the population are randomly paired off. During this drawing-time, the composition of the population remains fixed. One could think of T as the lifetime of a generation. Changes in population shares only occur from generation to generation, as is explained in more detail later on. For the moment, it is enough to know that during one generation, the shares of Hawks, Doves, and K-types are constant.

Formally, we have to deal with a combinatorial problem, since the population can be viewed simply as an urn. However, the problem is non-standard in the following sense: combinatorial models usually deal with urns of constant composition. Although we keep population shares fixed during the life-time of a generation, from the perspective of the K-type the composition of the urn changes. This is simply due to the fact that the mutant learns. Knowledge about Hawks is generated by interactions, i.e. by drawings from the urn. Fig. 1 gives a brief sketch of the fast-growing complexity, even for the first few rounds of the game from the perspective of the K-type.

Since the game-tree grows very fast, the figure only describes the first four draws $ts = 1, \ldots, 4$, and for $t=4$ only one branch is completed. Clearly, one can see that already after $t=1$ the urn is no longer one which contains only two 'element'. Suppose, for example, the K-type is matched to a Hawk in $t=1$. After that, in $t=2$, it is entirely possible that he will meet precisely this Hawk again (denoted H1 in the figure), or that he meets one of the Hawks unknown to him up to this point of time (H-1), or that he meets a Dove (D). Hence, from the point of view of the mutant, the urn splits into three elements, namely Doves, known Hawks and unknown Hawks, where the shares of the latter two change during the drawing procedure. What we are interested in, and what is needed later to calculate expected payoffs, is the probability of the K-type meeting one of the members of the set of known Hawks at an arbitrary point in time $1 < t \leq T$ The solution to this problem is given in the following theorem.

Proposition 1. *Consider a population consisting of strictly cooperative automata (Doves) and strictly defective automata (Hawks). Let x_H denote the population share of the Hawks,*

Fig. 1. The changing urn for the K-type.

which remains fixed during the drawing procedure, and n+1 the constant absolute size of the population. Let a single K-type mutant be randomly matched to the population. Then the probability for the K-type to meet a known Hawk at some arbitrary point in time t+1 is given as

$$\Pr(H_K(t+1)) = x_H \left(1 - \left(\frac{n-1}{n} \right)^t \right) \tag{1}$$

Proof. Let $\{H_K(t)\}$, $\{H_U(t)\}$ denote the sets of defectors which are known, respectively, unknown at t and let $\{D(t)\} = \{D\}$ denote the set containing the Doves. Let $H_K(t)$, $H_U(t)$, $D(t) = D$ denote the corresponding numbers of players contained in those sets at t. Let $I = \{1, \dots, n\}$ be the set of all players in the population and let $i(t) \in I$ denote the player who is matched to the mutant at t.

Consider an arbitrary point in time t. At t, three things can happen: the mutant meets a known Hawk with probability $\Pr(H_K(t)) \equiv \Pr(i(t) \in \{H_K(t)\})$, an unknown Hawk with probability $\Pr(H_U(t)) \equiv \Pr(i(t) \in \{H_U(t)\}) = x_H - \Pr(H_K(t))$, or a Dove with probability $(1 - x_H)$. Given the first event occurs, the probability of meeting a known Hawk at $t+1$ remains unchanged. This is simply due to the fact that no further Hawk has been learned at t. Thus, $H_K(t+1) = H_K(t)$. The probability of meeting a known Hawk at $t+1$, given a previously known Hawk was met at t is then simply given as

$$\Pr(H_K(t+1) \mid H_K(t)) \equiv \Pr(i(t+1) \in \{H_K(t+1)\} \mid i(t) \in \{H_K(t)\})$$
$$= \Pr(i(t+1) \in \{H_K(t)\} \mid i(t) \in \{H_K(t)\}) = \Pr(H_K(t))^2$$

If at t the second event occurred, then two possibilities for meeting a known Hawk at $t+1$ can be distinguished. First, the mutant could meet exactly the same Hawk that he met at t again at $t+1$. This happens with probability $1/n$, where n is the number of individuals excluding the mutant. Formally, $\Pr(i(t+1) = i(t) \mid i(t) \in \{H_U(t)\}) = \frac{1}{n}\Pr(H_U(t)) =$

$\frac{1}{n}(x_H - \Pr(H_K(t)))$. Second, given he meets at t, a Hawk unknown up to this point and does not meet this individual again at $t+1$, the probability for meeting a known Hawk at $t+1$ remains constant. To be more precise we are interested in the probability $\Pr(i(t+1) \in \{H_K(t+1)\} \backslash i(t) \mid i(t) \in \{H_U(t)\})$. Clearly, $\{H_K(t+1)\} \backslash i(t) \in \{H_U(t)\} = \{H_K(t)\}$ and $\#\{H_K(t+1)\} \backslash i(t) \in \{H_U(t)\} = H_K(t)$. Thus,

$$\Pr(i(t+1) \in \{H_K(t+1)\} \backslash i(t) \mid i(t) \in \{H_U(t)\})$$
$$= \Pr(i(t+1) \in \{H_K(t)\} \mid i(t) \in \{H_U(t)\})$$
$$= \Pr(H_K(t))\Pr(H_U(t)) = \Pr(H_K(t))(x_H - \Pr(H_K(t)))$$

The probability of meeting a known Hawk at $t+1$, given an unknown Hawk was met at t is, therefore, given as

$$\Pr(H_K(t+1) \mid H_U(t)) \equiv \Pr(i(t+1) \in \{H_K(t+1)\} \mid i(t) \in \{H_U(t)\})$$
$$= \Pr(i(t+1) = i(t) \mid i(t) \in \{H_U(t)\})$$
$$+ \Pr(i(t+1) \in \{H_K(t+1)\} \backslash i(t) \mid i(t) \in \{H_U(t)\})$$
$$= (x_H - \Pr(H_K(t))) \left(\frac{1}{n} + \Pr(H_K(t)) \right)$$

Finally, if a Dove was met at t, the probability of meeting a known Hawk at $t+1$ also remains unchanged since the number of known Hawks remains constant, i.e. $H_K(t+1) = H_K(t)$. Thus,

$$\Pr(H_K(t+1) \mid D(t)) \equiv \Pr(i(t+1) \in \{H_K(t+1)\} / i(t) \in \{D\})$$
$$= \Pr(i(t+1) \in \{H_K(t)\} / i(t) \in \{D\}) = \Pr(H_K(t))(1 - x_H)$$

Hence, the total probability of meeting a known Hawk at $t+1$, given the immediate possible histories of the game, i.e. given t, can be expressed as

$$\Pr(H_K(t+1)) = \Pr(H_K(t+1) \mid H_K(t)) + \Pr(H_K(t+1) \mid H_U(t))$$
$$+ \Pr(H_K(t+1) \mid D(t)) = \Pr(H_K(t))^2$$
$$+ (x_H - \Pr(H_K(t))) \left(\Pr(H_K(t)) + \frac{1}{n} \right) + (1 - x_H)\Pr(H_K(t))$$

which can be simplified to

$$\Pr(H_K(t+1)) = \left(1 - \frac{1}{n} \right) \Pr(H_K(t)) + \frac{1}{n} x_H.$$

Since the above relationship between the probability of meeting a known Hawk at t and the probability of meeting a known Hawk at $t+1$ holds for an arbitrary t, we can substitute $\Pr(H_K(t))$ with an identical relationship for $\Pr(H_K(t))$ and $\Pr(H_K(t-1))$. Formally, this means substituting

$$\Pr(H_K(t)) = \left(1 - \frac{1}{n} \right) \Pr(H_K(t-1)) + \frac{1}{n} x_H$$

into the above expression for $\Pr(H_K(t+1))$, which leads to

$$\Pr(H_K(t+1)) = \left(1 - \frac{1}{n}\right)^2 \Pr(H_K(t-1)) + \frac{1}{n}x_H\left(1 - \frac{1}{n}\right) + \frac{1}{n}x_H$$

Substituting again allows the evolving structure to be made more obvious. Inserting

$$\Pr(H_K(t-1)) = \left(1 - \frac{1}{n}\right)\Pr(H_K(t-2)) + \frac{1}{n}x_H$$

yields

$$\Pr(H_K(t+1)) = \left(1 - \frac{1}{n}\right)^3 \Pr(H_K(t-2))$$

$$+ \frac{1}{n}x_H\left(1 - \frac{1}{n}\right)^2 + \frac{1}{n}x_H\left(1 - \frac{1}{n}\right) + \frac{1}{n}x_H$$

Clearly, it is possible to eliminate all probability-terms, except the one for the first draw. But the probability of meeting a known Hawk in the first round is surely zero. The remaining terms are simply a geometric series of the expression $(1-1/n)$. If we use the method of backward iteration just mentioned, we obtain

$$\Pr(H_K(t+1)) = \Pr(H_K(\tau = 0))\left(1 - \frac{1}{n}\right)^t + \frac{1}{n}x_H\sum_{\tau=0}^{t-1}\left(1 - \frac{1}{n}\right)^\tau.$$

As we said, $\Pr(H_K(\tau = 0)) = 0$. Thus,

$$\Pr(H_K(t+1)) = \frac{1}{n}x_H\sum_{\tau=0}^{t-1}\left(1 - \frac{1}{n}\right)^\tau = \frac{1}{n}x_H\sum_{\tau=0}^{t-1}\left(\frac{n-1}{n}\right)^\tau.$$

Now,

$$\sum_{\tau=0}^{t-1}\left(\frac{n-1}{n}\right)^\tau = n\left(1 - \left(\frac{n-1}{n}\right)^t\right)$$

thus,

$$\Pr(H_K(t+1)) = x_H\left(1 - \left(\frac{n-1}{n}\right)^t\right)$$

□

Note that this result does not only hold for the case of a single K-type-mutant in the first generation but is rather general. Any randomly chosen member of a K-type-sub-population with an absolute size bigger than 1 faces exactly the same situation as the single mutant. All that matters is the absolute size of the population and the share of the Hawk's sub-population x_H. But the latter is the same, regardless of whether there is only one K-type or more. The only difference is that the sub-population of cooperative players now not only consists of Doves, but also includes K-types. However, this fact in no way influences the derivation of the formula above.

2.2. Existence and stability of equilibria

Since we are now endowed with a formal model of learning, we can return to the second purpose of this paper: to study the effects of learning in the evolutionary Prisoners' Dilemma. So far we have stressed the question of how the K-type acquires his knowledge. The answer has been given as the formula in the main result, which is none other than the law of growth of the K-type's knowledge. Now, we have to counsel the K-type on how to use the acquired information. A quite natural response to this question is to assume that the Knowledge-type would like to avoid being exploited by known Hawks. So let us assume that the K-type rejects interaction with a known Hawk (he simply goes away) and that in this case both take an outside option, just as in Frank's model, which yields them λ, with $1 < \lambda < 3$.

Amann and Yang assumed that the K-type had to bear some 'investigation costs'. Since we are in a situation with a fully endogenous model of learning, there is no space left for such costs. On the other hand, it is quite natural to assume that the ability of learning cannot be achieved without cost. Biological research has shown that evolutionary innovations cannot be taken for granted. There are several examples which show that whenever an 'innovation' arises, there is always a price to be paid. Recent anthropological research claims that the high-speed growth of the human brain two and a half million years ago, which led to an augmentation of the cranium, necessitated a shortening of the gestation period. Thus, human offspring now needed a longer period and higher degree of care. This could be interpreted as a price which had to be paid for the intellectual development of mankind. Therefore, we assume that the K-type has to sacrifice a fixed amount of fitness c, $0 < c < 1$, in order to gain his new ability. This is represented in our payoff matrix shown in Table 2.

Alternatively, one could imagine that memory needs some training in order to serve the information which is needed. Loosely speaking, the mutant has to undergo some daily brain-jogging in order to keep his mind functioning. Perhaps a record file has to be created in order to avoid the loss of memory. The idea is similar to that suggested by authors like Conlisk (1980) or Guttman (1996), namely that optimizing behavior is costly. Indeed, the K-type can be viewed as a rational actor in the sense of Guttman since he plays a best reply given his actual knowledge.

Now consider a population of infinite size. More precisely, imagine a population that consists of an infinite number of finitely-sized subgroups. Let $n+1$ denote the absolute size of each subgroup where n is an odd integer. We assume that interactions take place between subjects of one subgroup, but not between subjects of two different groups. Now, imagine that the members of all subgroups are paired randomly within their groups and that this procedure is repeated $T+1$ times, where T is a finite number. We could think of $T+1$ as

Table 2
The Prisoners' Dilemma with a knowledgeable mutant

	Dove	Hawk (known)	Hawk (unknown)	K-type
Dove	3	0	0	3
Hawk (known)	6	1	1	λ
Hawk (unknown)	6	1	1	6
K-Type	3−c	λ−c	−c	3−c

the finite lifetime of a generation. Before players die they leave offspring according to their 'success' which will be explained more precisely below. What we are interested in is the question of how the shares of the different behavioral types (D, H and K) develop over time. To give an answer to this question, some preliminary work has to be done.

Let Q denote the set of all possible compositions of a subgroup and let Q_j denote an element of Q, i.e. Q_j is a triple (H_j, D_j, K_j), such that $H_j + D_j + K_j = n + 1$. Now let q denote the number of all possible triples (H_j, D_j, K_j) whose elements sum up to $n+1$. Clearly, since $n+1$ is finite, the number $q = \#\{(H_j, D_j, K_j) \mid H_j + D_j + K_j = n + 1\}$ also has to be finite. Thus, $Q = \{Q_1, \dots, Q_q\}$. Now consider any given population state (x_H, x_D, x_K), $\sum_i x_i = 1$, $i = H, D, K$. Obviously, any given population state is being realized via some distribution on Q. Let f_{Q_j} denote the relative frequency of triple Q_j, $\sum_{j=1}^{q} f_{Q_j} = 1$. Then, $(x_H, x_D, x_K) = \sum_{j=1}^{q} (Q_j/n+1) f_{Q_j}$. Recall our assumption of an infinitely-sized population. Let r_{Q_j} denote the absolute frequency of triple Q_j. The total population size is, therefore, $\sum_{j=1}^{q} Q_j r_{Q_j} = \sum_{j=1}^{q} (n + 1) r_{Q_j}$. Thus, for every Q_j with $f_{Q_j} > 0$, r_{Q_j} equals infinity (n and q are definitely finite).

We are now prepared to state the main argument. Since every Q_j with $f_{Q_j} > 0$ has an infinite number of realisations, we can rely on the law of large numbers. This implies that the average payoff received by a type within all subgroups that belong to this type Q_j equals the expected payoff. Summing up expected payoffs over all $j = 1, \dots, q$ types of triples and weighting them according to their relative frequency f_{Q_j} thus yields a type's total average payoff. Formally, let $\bar{\pi}_{Q_j}^i$ denote the average payoff received by type i within all subgroups which belong to the type Q_j. Then, $\bar{\pi}_{Q_j}^i = E\pi_{Q_j}^i$ and $\bar{\pi}^i = \sum_{j=1}^{q} \bar{\pi}_{Q_j}^i f_{Q_j} = \sum_{j=1}^{q} E\pi_{Q_j}^i f_{Q_j}$. Then we assume that the development of population shares is given by the familiar continuous time replicator dynamic

$$\dot{x}_i = x_i(\bar{\pi}^i - \bar{\pi}), \quad i = H, D, K \tag{2}$$

where $\bar{\pi}$ is the population average payoff. Thus, strategies which do better than the average increase over time while strategies which persistently yield less than the average payoff are driven to extinction.

In Lemma 1 we present the results concerning the existence of equilibrium population states. The fixed points of the replicator dynamics are defined by $\dot{x}_i = 0$, $\forall i$. Note that not all fixed points are Nash equilibria. While this is true for interior fixed points, on the boundary of the simplex Δ_3 there may arise corner solutions which cannot be justified as Nash equilibria (cp. e.g. Friedman, 1991; Weibull, 1995). Since all locally stable equilibria also have to be Nash equilibria (cp. Bomze, 1986), we will concentrate directly on the latter.

Lemma 1. *(Existence) Consider the evolutionary Prisoners' Dilemma with a knowledge-able mutant and suppose that the development of population shares is given by replicator dynamics. Then*

(i) for

$$c < \frac{(n + 1)\lambda(1 - k)^2(\lambda - 6) + \lambda(1 - k)(5 + 3n)}{n(2 + (1 - k)(\lambda - 6))}$$

the interior fixed point

$$(x_H^*, x_K^*, x_D^*) = \left\{ \frac{n}{n+1} \frac{c}{\lambda(1-k)}, \frac{n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(\lambda-6)(1-k)^2}, 1 - x_H^* - x_K^* \right\}$$

exists;

(ii) for

$$\frac{(n+1)\lambda(1-k)[(1-k)(\lambda-6)+5] - 2\lambda n(1-k)}{n[(1-k)(\lambda-6)+2]} < c < \frac{n+1}{n}\lambda(1-k)$$
$$-\frac{3+n}{n} - 1$$

or

$$\frac{(n+1)\lambda(1-k)[(1-k)(\lambda-6)+5] - 2\lambda n(1-k)}{n[(1-k)(\lambda-6)+2]} > c > \frac{n+1}{n}\lambda(1-k)$$
$$-\frac{3+n}{n} - 1$$

the two-type fixed point

$$(x_H^*, x_K^*, x_D^*) = \left\{ 1 - x_K^*, \frac{\lambda(1-k) - (1+c) \cdot n/(n+1) - 3/(n+1)}{((1-k)(2\lambda-6)+2)}, 0 \right\}$$

is a Nash equilibrium;

(iii) for

$$c > \lambda(1-k)\frac{n+1}{n} - \frac{n+3}{n}$$

the defectors-only fixed point $(x_H^*, x_K^*, x_D^*) = \{1, 0, 0\}$ *is a Nash equilibrium.*

Proof: see Appendix.

Proposition 2. *(Stability): (i) for*

$$c < \min \left\{ \frac{(n+1)\lambda(1-k)^2(\lambda-6) + \lambda(1-k)(5+3n)}{n(2+(1-k)(\lambda-6))}, \right.$$
$$\left. (1-k)\left(2\lambda\frac{(n+2)}{n} - 3\frac{(n+3)}{n} \right) \right\}$$

the interior fixed point is globally stable.

(ii) For

$$\frac{(n+1)\lambda(1-k)[(1-k)(\lambda-6)+5] - 2\lambda n(1-k)}{n[(1-k)(\lambda-6)+2]} < c < \frac{n+1}{n}\lambda(1-k)$$
$$-\frac{3+n}{n} - 1,$$

the two-type NE is globally stable.

(iii) For

$$c > \lambda(1 - k)\frac{n+1}{n} - \frac{n+3}{n}$$

the defectors-only equilibrium is globally stable.

Proof: see Appendix.

3. Concluding remarks

In deriving an endogenous model of learning, we have been able to justify the introduction of a knowledgeable mutant into the Prisoners' Dilemma game. In our paper, knowledge no longer falls like manna from heaven. Instead, we give a full and precise description of how the learnable mutant gains his knowledge. This is done in a very simple and intuitive way, namely by allowing the mutant to simply store the face of each opponent he meets. All he has to do in order to get informed about his opponents is to play the game — as simple as that. Despite the somewhat mechanistic look of our model we think it to have a 'realistic' appeal because in all everyday forms of interaction, simple experience is the source of knowledge. Furthermore, the knowledgeable mutant need not carry out complicated reasoning in order to condition his decision to cooperate or to reject, as would be necessary in any model of Bayesian flavor, where some a priori beliefs about the type distribution would have to be updated. Despite low cognitive requirements, our stupid-learning mutant succeeds. Indeed, it is worth mentioning that this kind of learning inevitably arises in combination with memory, and is, therefore, one of the simplest and most elementary forms of learning one can imagine. However, this rudimentary learning already leads to remarkable results. The occurrence of a knowledgeable mutant does not only favor sophisticated cooperators, but also naive cooperators. This is the case if the costs of sophistication are not too high. For higher evolutionary costs, c, there is still a stable two-type equilibrium consisting of defectors and sophisticated cooperators. If, and only if, the costs of evolutionary innovation are much too high, i.e. $c > \lambda(1 - k)[(n + 1)/n] - [(n + 3)/n]$, the original PD outcome is yielded. Thus, only by introducing the K-type, the chances for the survival of cooperative behavior have been raised to a remarkable extent.

Of course one might object that memory is modeled to work perfectly, i.e. every defector previously met is stored in the brain and will never be forgotten. But it is easy to incorporate some kind of forgetfulness into the model, e.g. by introducing some depreciation of the acquired stock of knowledge. Note that, even in this case, the essence of our results remains. As long as there is a sufficiently large net effect of learning, cooperators will survive.

A further possible objection is that the ability to learn is restricted to the good guys. So why not allow Hawks to learn, too? At first sight, it seems as though it makes no difference whether Hawks are assumed to be able to learn or not. Acquiring knowledge about opponents does not lead to an advantage for defectors. Imagine a Hawk being matched to a K-type. Even if the defector knows that he faces a mutant he will take the outside option. But this is only half of the truth. A sophisticated defector is able to avoid the worst outcome whenever meeting a known defector. Thus, sophistication on the other hand creates an advantage for defectors, too. It is far from obvious whether the results of our three-types model are robust

against the introduction of a nasty mutant. We investigated this question in a separate paper (Vogt, 1998). Our main finding is that the results of the three-types model can roughly be restated. Within appropriate parameters the interior three-types equilibrium is again stable. Interestingly, there is no completely mixed equilibrium which includes the nasty mutant. Most of the new arising fixed points including the sophisticated defector turn out to be unstable. The only exception is a symbiosis of naive and sophisticated defectors which might be stable for appropriate parameters.

One last objection might concern the assumption that the cooperative mutant rejects interaction when meeting a known Hawk. How should this assumption be justified? We do not really know much about human behavior. But one of the best documented and most widespread preferences of human beings is the one for being treated fairly (which of course in no way implies that people themselves always behave in a fair manner — the opposite is often true). In ultimatum games, people often reject offers they should accept if they were fully rational. This is one of the most robust results of bargaining experiments.[5] For this reason we think the behavioral assumption of exploitation aversion is well justified.

Acknowledgements

Appendix A

A.1. Derivation of payoffs

Before we start dealing with the questions of existence and stability of population states, we have to derive expected payoffs for the three types D, H and K. We will start with the naive cooperator D. What is the expected payoff for a Dove within some triple Q_j? Remember that the size of a subgroup equals $n+1$. D_j, H_j, K_j denote the number of Doves, Hawks and K-types within a subgroup which belongs to the class of triples Q_j. The naive cooperator gains three when he meets another naive cooperator or when he is matched to a K-type while his earnings are zero when being exploited by a defector. Thus, the expected payoff for a Dove with respect to triple Q_j is given as

$$E\pi_{Q_j}^D = 3\left(\frac{D_j - 1}{n} + \frac{K_j}{n}\right) \tag{A.1}$$

Let

$$\frac{D_j}{n+1} \equiv x_j^D, \quad \frac{K_j}{n+1} \equiv x_j^K, \quad \frac{H_j}{n+1} \equiv x_j^H$$

---

[5] For an overview, see Roth (1995). The pioneering work in this field was Güth et al. (1982). For a more detailed study see e.g. Ochs and Roth (1989).

Then the expected payoff can be written as

$$E\pi_{Q_j}^D = 3\frac{1}{n}(n+1)\left(\frac{D_j}{n+1} + \frac{K_j}{n+1} - \frac{1}{n+1}\right) \Leftrightarrow E\pi_{Q_j}^D$$

$$= 3\frac{n+1}{n}(x_j^D + x_j^K) - 3\frac{1}{n} \tag{A.2}$$

We get the total expected payoff for D if we sum up over all triples $Q = \{Q_1, ..., Q_q\}$:

$$E\pi^D = \sum_{j=1}^{q} E\pi_{Q_j}^D f_{Q_j} = 3\frac{n+1}{n}\sum_{j=1}^{q}(x_j^D + x_j^K)f_{Q_j} - 3\sum_{j=1}^{q}\frac{1}{n}f_{Q_j}$$

$$= 3\frac{n+1}{n}(x_D + x_K) - 3\frac{1}{n}.$$

Since $x_D + x_K = 1 - x_H$, we get

$$E\pi^D = 3(1 - x_H) + 3\frac{1}{n}(1 - x_H) - 3\frac{1}{n} = 3(1 - x_H) - 3\frac{1}{n}x_H \tag{A.3}$$

Next turn to the K-type. Let $\Pr(H_K^j(t+1))$ denote the probability to meet a known Hawk at point of time $t+1$ with respect to Q_j. Note that this probability grows as t becomes larger. Thus, the expected payoff also changes while time passes by and we have got to calculate the average expected payoff. Let $\bar{P}r(H_K^j(T+1))$ denote the average probability to meet a known defector when there are $T+1$ matching rounds, i.e.

$$\bar{P}r(H_K^j(T+1)) = \frac{1}{T+1}\sum_{t=0}^{T}\Pr(H_K^j(t+1)).$$

The K-type gets three if he is matched to other cooperative guys (naive or sophisticated), he receives the payoff of the outside option λ when he meets a known defector and he has to bear the costs of sophistication c in all cases. Now let us first deal with the expected payoff for some given point of time $t+1$. This payoff then is simply given as

$$E\pi_{Q_j}^K(t+1) = 3\frac{1}{n}(D_j + K_j - 1) + \lambda\frac{1}{n}(H_K^j(t+1)) - c \Leftrightarrow E\pi_{Q_j}^K(t+1)$$

$$= 3\frac{n+1}{n}\left(\frac{D_j + K_j - 1}{n+1}\right) + \lambda\frac{n+1}{n}\frac{(H_K^j(t+1))}{n+1} - c \tag{A.4}$$

Clearly, the expected frequency of meeting a known Hawk at $t+1$ is derived from the learning model as $\Pr(H_K^j(t+1))$. Thus, we can write

$$E\pi_{Q_j}^K(t+1) = 3\frac{n+1}{n}(x_j^D + x_j^K) - 3\frac{1}{n} + \lambda\frac{n+1}{n}\Pr(H_K^j(t+1)) - c \tag{A.5}$$

The total expected payoff for $t+1$ is then received by summing over all triples Q_j:

$$E\pi^K(t+1) = \sum_{j=1}^{q} E\pi^K_{Q_j} f_{Q_j} = 3\frac{n+1}{n} \sum_{j=1}^{q} \left(x^D_j + x^K_j - \frac{1}{n+1}\right) f_{Q_j}$$

$$+\lambda\frac{n+1}{n} \sum_{j=1}^{q} \Pr(H^j_K(t+1)) f_{Q_j} - c\sum_{j=1}^{q} f_{Q_j} \qquad (A.6)$$

The second sum needs a closer inspection. Clearly, the formula from the learning model is directly applicable to a subgroup belonging to some triple Q_j. The probability to meet a known Hawk is simply given as

$$\Pr(H^j_K(t+1)) = x^H_j \left(1 - \left(\frac{n-1}{n}\right)^t\right). \qquad (A.7)$$

Thus,

$$\sum_{j=1}^{q} \Pr(H^j_K(t+1)) f_{Q_j} = \sum_{j=1}^{q} x^H_j \left(1 - \left(\frac{n-1}{n}\right)^t\right) f_{Q_j}$$

$$= \left(1 - \left(\frac{n-1}{n}\right)^t\right) \sum_{j=1}^{q} x^H_j f_{Q_j}$$

$$= x_H \left(1 - \left(\frac{n-1}{n}\right)^t\right) = \Pr(H_K(t+1)) \qquad (A.8)$$

which equals our original result from the learning model (Proposition 1) for a population which is not disaggregated into an infinity of subgroups. Thus, the K-types expected payoff for $t+1$ is given as

$$E\pi^K(t+1) = 3\frac{n+1}{n} \left(x_D + x_K - \frac{1}{n+1}\right) + \lambda\frac{n+1}{n} \Pr(H_K(t+1)) - c \qquad (A.9)$$

Now let us calculate the average expected payoff for K, i.e. sum over all $t = 1, \ldots, T$ and divide by $T+1$:

$$\frac{1}{T+1} \sum_{t=0}^{T} E\pi^K(t+1) = 3\frac{1}{T+1}\frac{n+1}{n} \sum_{t=0}^{T} \left(x_D + x_K - \frac{1}{n+1}\right)$$

$$+\lambda\frac{n+1}{n}\frac{1}{T+1} \sum_{t=0}^{T} \Pr(H_K(t+1)) - \frac{1}{T+1} \sum_{t=0}^{T} c \qquad (A.10)$$

Again, the second sum needs a closer inspection:

$$\frac{1}{T+1} \sum_{t=0}^{T} \Pr(H_K(t+1)) = \frac{1}{T+1} \sum_{t=0}^{T} x_H \left(1 - \left(\frac{n-1}{n}\right)^t\right)$$

$$= x_H - x_H \frac{1}{T+1} \sum_{t=0}^{T} \left(\frac{n-1}{n}\right)^t.$$

Now consider

$$\sum_{t=0}^{T}\left(\frac{n-1}{n}\right)^{t}=1+\left(\frac{n-1}{n}\right)+\cdots+\left(\frac{n-1}{n}\right)^{T}.$$

The sum of this geometric series simply calculates as

$$n\left(1-\left(\frac{n-1}{n}\right)^{T+1}\right).$$

Thus,

$$\frac{1}{T+1}\sum_{t=0}^{T}E\pi^{K}(t+1)=3\frac{n+1}{n}\left(x_{D}+x_{K}-\frac{1}{n+1}\right)$$

$$+\lambda\frac{n+1}{n}x_{H}\left(1-\frac{n}{T+1}\left(1-\left(\frac{n-1}{n}\right)^{T+1}\right)\right)-c \quad \text{(A.11)}$$

With

$$\frac{n}{T+1}\left(1-\left(\frac{n-1}{n}\right)^{T+1}\right)\equiv k\left(n,T\right)\equiv k$$

we can write total average expected payoff for K as

$$E\pi^{K}=3\frac{n+1}{n}\left(x_{D}+x_{K}-\frac{1}{n+1}\right)+\lambda\frac{n+1}{n}x_{H}(1-)-c$$

$$=3\frac{n+1}{n}(1-x_{H})-3\frac{1}{n}+\lambda\frac{n+1}{n}x_{H}(1-)-c \quad \text{(A.12)}$$

Finally, we turn to the Hawks. They receive six when meeting a naive cooperator or when they meet a K-type to whom the Hawk under consideration is still unknown, they get the value of the outside option in case they are matched to a K-type who knows the Hawk, and they get the worst payoff (one) when meeting another Hawk. First, note that the probability to be a known Hawk when meeting a K-type is simply given as

$$\frac{\Pr(H_{K}^{j}(t+1))}{x_{j}^{H}}.$$

Accordingly, the probability to be an unknown Hawk is

$$\frac{x_{j}^{H}-\Pr(H_{K}^{j}(t+1))}{x_{j}^{H}}=1-\frac{\Pr(H_{K}^{j}(t+1))}{x_{j}^{H}}.$$

Thus, the expected payoff when being a known Hawk for triple Q_{j} is

$$E\pi_{H_{K}}^{j}(t+1)=\left(6\frac{D_{j}}{n}+\frac{H_{j}-1}{n}+\lambda\frac{K_{j}}{n}\right)\frac{\Pr(H_{K}^{j}(t+1))}{x_{j}^{H}}, \quad \text{(A.13)}$$

otherwise it is given as

$$E\pi_{H_U}^j(t+1) = \left(6\frac{D_j}{n} + \frac{H_j - 1}{n} + 6\frac{K_j}{n}\right)\left(1 - \frac{\Pr(H_K^j(t+1))}{x_j^H}\right). \tag{A.14}$$

Summing up these two expressions we get

$$
\begin{aligned}
E\pi_H^j(t+1) &= 6\frac{1}{n}(D_j + K_j) + \frac{1}{n}(H_j - 1) + (\lambda - 6)\frac{1}{n}K_j\frac{\Pr(H_K^j(t+1))}{x_j^H} \\
&= 6\frac{n+1}{n}(x_j^D + x_j^K) + \frac{n+1}{n}\left(x_j^H - \frac{1}{n+1}\right) \\
&\quad + (\lambda - 6)\frac{n+1}{n}x_j^K\frac{\Pr(H_K^j(t+1))}{x_j^H} \tag{A.15}
\end{aligned}
$$

Summing over all triples yields

$$
\begin{aligned}
E\pi_H(t+1) &= 6\frac{n+1}{n}\sum_{j=1}^{q}(x_j^D + x_j^K)f_{Q_j} + \frac{n+1}{n}\sum_{j=1}^{q}x_j^H f_{Q_j} - \frac{1}{n}\sum_{j=1}^{q}f_{Q_j} \\
&\quad + (\lambda - 6)\frac{n+1}{n}\sum_{j=1}^{q}x_j^K\frac{\Pr(H_K^j(t+1))}{x_j^H}f_{Q_j} \tag{A.16}
\end{aligned}
$$

Now look at the last sum:

$$
\begin{aligned}
\sum_{j=1}^{q}x_j^K\frac{\Pr(H_K^j(t+1))}{x_j^H}f_{Q_j} &= \sum_{j=1}^{q}x_j^K\frac{1}{x_j^H}x_j^H\left(1 - \left(\frac{n-1}{n}\right)^t\right)f_{Q_j} \\
&= \left(1 - \left(\frac{n-1}{n}\right)^t\right)\sum_{j=1}^{q}x_j^K f_{Q_j} \\
&= x_K\left(1 - \left(\frac{n-1}{n}\right)^t\right) \tag{A.17}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
E\pi_H(t+1) &= 6\frac{n+1}{n}(x_D + x_K) + \frac{n+1}{n}x_H - \frac{1}{n} \\
&\quad + (\lambda - 6)\frac{n+1}{n}x_K\left(1 - \left(\frac{n-1}{n}\right)^t\right) \tag{A.18}
\end{aligned}
$$

Now, averaging over all $t = 0, \dots, T$ and again exploiting the fact that $x_D + x_K = 1 - x_H$, leads to

$$E\pi_H = 6 - 5x_H + \frac{5}{n}(1 - x_H) + (\lambda - 6)(1 - k)\frac{n+1}{n}x_K \tag{A.19}$$

A.2. Existence

Proof of Lemma 1. There exist three corner fixed points. The pure Dove fixed point can never be a NE because, in a pure Dove population, Dove is not a best reply to itself (Hawk is always better). The pure K-type fixed point can never be a NE because, in a 100 percent K-type population, it is always better to be a naive cooperator who does not have to bear the evolutionary cost of sophistication.

Expected payoffs are given by (A.3), (A.12) and (A.19). The interior fixed point then simply results from equating all average payoffs and the condition for existence results from $x_H^* + x_K^* < 1$. In more detail:

Equating (A.3) and (A.12) immediately yields

$$x_H^* = \frac{n}{n+1} \frac{c}{\lambda(1-k)} \tag{A.20}$$

Equating (A.19) and (A.3) yields

$$x_K = \frac{2x_H - 3}{(\lambda - 6)(1-k)} - \frac{2}{(n+1)(\lambda - 6)(1-k)} \tag{A.21}$$

Inserting (A.20) into (A.21) we obtain

$$
\begin{aligned}
x_K^* &= \frac{2(n/(n+1))[c/\lambda(1-k)] - 3}{(\lambda - 6)(1-k)} - \frac{2}{(n+1)(\lambda - 6)(1-k)} \\
&= \frac{2cn - 3(n+1)\lambda(1-k)}{(n+1)\lambda(\lambda - 6)(1-k)^2} - \frac{2\lambda(1-k)}{(n+1)\lambda(\lambda - 6)(1-k)^2} \\
&= \frac{n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(\lambda - 6)(1-k)^2} \tag{A.22}
\end{aligned}
$$

To be an interior state in Δ_3 equilibrium shares x_K^* and x_H^* must sum to less than 1. Thus,

$$
\begin{aligned}
x_K^* + x_H^* &= \frac{cn}{(n+1)\lambda(1-k)} + \frac{n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(1-k)^2(\lambda - 6)} \\
&= \frac{cn(1-k)(\lambda - 6) + n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(1-k)^2(\lambda - 6)} < 1 \\
&\Leftrightarrow c < \frac{(n+1)\lambda(1-k)^2(\lambda - 6) + \lambda(1-k)(5+3n)}{n(2 + (1-k)(\lambda - 6))} \tag{A.23}
\end{aligned}
$$

Clearly, for an interior equilibrium it must also be fulfilled that

$$
\begin{aligned}
x_K^* &= \frac{n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(1-k)^2(\lambda - 6)} > 0 \Leftrightarrow 2cn - 3\lambda n(1-k) \\
&- 5\lambda(1-k) < 0 \Leftrightarrow c < (1-k)(5+3n)\frac{\lambda}{2n} \tag{A.24}
\end{aligned}
$$

It is easily shown that (A.23) is always more restrictive than (A.24). Thus, if (A.23) is fulfilled the interior Nash Equilibrium exists.

Consider now the Hawks-only fixed point. Payoffs at (1,0,0) result in 1 for Hawks, 0 for Doves and $\lambda(1-k)[(n+1)/n] - (3/n) - c$ for K-types. So Hawk is a best reply to itself as long as

$$E\pi_H|_{(1,0,0)} = 1 > \lambda(1-k)\frac{n+1}{n} - \frac{3}{n} - c$$

$$= E\pi_K|_{(1,0,0)} \Leftrightarrow c > \lambda(1-k)\frac{n+1}{n} - \frac{n+3}{n} \tag{A.25}$$

By a similar argument, together with the requirement $x_K^* > 0$, one receives the condition for the two-types fixed point to be a NE. More detailed:

Payoffs for Hawks and Doves evaluated at $x_H = 1 - x_K$ result as

$$E\pi_H|_{x_H=1-x_K} = 1 + \frac{1}{n}x_K(5(1+n) + (n+1)(1-k)(\lambda - 6)) \tag{A.26}$$

$$E\pi_K|_{x_H=1-x_K} = \lambda(1-k)\frac{n+1}{n} - \frac{3}{n} - c$$

$$+x_K\frac{1}{n}(3(n+1) - \lambda(n+1)(1-k)) \tag{A.27}$$

Now equate (A.26) and (A.27):

$$E\pi_H|_{x_H=1-x_K} = E\pi_K|_{x_H=1-x_K} \Leftrightarrow x_K^*$$

$$= \frac{\lambda(1-k) - (1+c)n/(n+1) - 3/(n+1)}{((1-k)(2\lambda - 6) + 2)} \tag{A.28}$$

The two types fixed point exists if $0 < x_K^* < 1$.

$$x_K^* > 0 \Leftrightarrow \frac{\lambda(1-k) - (1+c)(n/n+1) - (3/n+1)}{((1-k)(2\lambda - 6) + 2)} > 0$$

(i) $((1-k)(2\lambda-6)+2) > 0$

$$x_K^* > 0 \Leftrightarrow c < \frac{n+1}{n}\lambda(1-k) - \frac{3+n}{n} \tag{A.29}$$

$$x_K^* < 1 \Leftrightarrow c > \frac{n+1}{n}[(1-k)(6-\lambda) - 2] - \frac{3+n}{n} \tag{A.30}$$

(ii) $((1-k)(2\lambda-6)+2) < 0$

$$x_K^* > 0 \Leftrightarrow c > \frac{n+1}{n}\lambda(1-k) - \frac{3+n}{n} \tag{A.31}$$

$$x_K^* < 1 \Leftrightarrow c < \frac{n+1}{n}[(1-k)(6-\lambda) - 2] - \frac{3+n}{n} \tag{A.32}$$

Next we turn to the question whether the two types fixed point can be a Nash equilibrium. Thus, we have to decide whether a small fraction of Doves is able to erode the two types population states.

$$E\pi_H|_{x_K^*} = 1 + \frac{\lambda(1-k) - (1+c)n/(n+1) - 3/(n+1)}{n((1-k)(2\lambda-6)+2)}(5(1+n)$$

$$+(n+1)(1-k)(\lambda-6)) \tag{A.33}$$

$$E\pi_D|_{x_K^*} = 3\frac{n+1}{n}\frac{\lambda(1-k) - (1+c)n/(n+1) - 3/(n+1)}{n((1-k)(2\lambda-6)+2)} - 3\frac{1}{n} \tag{A.34}$$

A comparison of these expressions yields after tedious calculations

(i) $((1-k)(2\lambda-6)+2)>0$

$$E\pi_H|_{x_K^*} > E\pi_D|_{x_K^*} \Leftrightarrow c > \frac{(n+1)\lambda(1-k)[(1-k)(\lambda-6)+5] - 2\lambda n(1-k)}{n[(1-k)(\lambda-6)+2]}$$

$$\tag{A.35}$$

(ii) $((1-k)(2\lambda-6)+2)<0$

$$E\pi_H|_{x_K^*} > E\pi_D|_{x_K^*} \Leftrightarrow c < \frac{(n+1)\lambda(1-k)[(1-k)(\lambda-6)+5] - 2\lambda n(1-k)}{n[(1-k)(\lambda-6)+2]}$$

$$\tag{A.36}$$

Thus, for case (i) (A.29) and (A.35) define the parameter range for which the two types fixed point is a Nash equilibrium while (A.31) and (A.36) constitute case (ii). This completes the proof on existence. □

A.3. Stability

Before proving global stability we will derive the conditions for local asymptotic stability. The results are given in the Lemma below.

Lemma 2. *(Local asymptotic stability): Let*

$$c_1 \equiv \frac{(n+1)\lambda(1-k)^2(\lambda-6) + \lambda(1-k)(5+3n)}{n(2+(1-k)(\lambda-6))},$$

$$c_2 \equiv (1-k)\left(2\lambda\frac{(n+2)}{n} - 3\frac{(n+3)}{n}\right), \quad c_3 \equiv \lambda(1-k)\frac{n+1}{n} - \frac{n+3}{n}.$$

(i) For $c < \min\{c_1, c_2\}$ the interior Nash equilibrium is locally stable.
(ii) For $c_3 > c > c_1$ the two types Nash equilibrium is locally stable.
(iii) For $c > c_3$ the Hawk only equilibrium is locally stable.

Proof. (ad i) The trace of the Jacobian at the interior fixed point is calculated as

$$\text{Tr } J = \frac{(1-k)[n(3-2\lambda) - 4\lambda + 9] + cn}{(1-k)(n+1)(1-k)(6-\lambda)} 2\frac{c}{\lambda} \tag{A.37}$$

The determinant looks a little bit more complicated. It is given as

$$\text{Det } J = (2cn - \lambda(1-k)(3n-5))\frac{c}{\lambda^2}$$

$$\cdots \frac{\begin{array}{c}\lambda^2(n+1)(1-k)^2 - \lambda(3n(1-k(3-2k))+1+k(7-6k)) \\ +c(n(4-6k-\lambda(1-k)))\end{array}}{n(\lambda-6)(k-1)^3(n+1)} \tag{A.38}$$

A standard result from the theory of dynamical systems tells us that a fixed point is locally stable if the determinant of the Jacobian is positive while the trace has to be negative.

Since $0 \leq k < 1$ and $\lambda < 3$, $n(\lambda-6)(k-1)^3(n+1) > 0$. Now the equilibrium share of the K-types was given as

$$x_K^* = \frac{n(2c - 3\lambda(1-k)) - 5\lambda(1-k)}{(n+1)\lambda(\lambda-6)(1-k)^2}.$$

Clearly,

$$x_K^* > 0 \Leftrightarrow n(2c - 3\lambda(1-k)) - 5\lambda(1-k) < 0.$$

Thus,

$$(2cn - \lambda(1-k)(3n-5)) > 0$$

and

$$\begin{aligned}\text{Det } J > 0 &\Leftrightarrow \lambda^2(n+1)(1-k)^2 - \lambda(3n(1-k(3-2k)) \\ &\quad + 1 + k(7-6k)) + c(n(4-6k-\lambda(1-k))) > 0 \\ &\Leftrightarrow c < \frac{(n+1)\lambda(1-k)^2(\lambda-6) + \lambda(1-k)(5+3n)}{n(2+(1-k)(\lambda-6))}\end{aligned} \tag{A.39}$$

Obviously, (A.39) is identical to (A.23). Thus, the determinant is positive whenever the interior fixed point exists. Next look at the trace of the Jacobian.

$$\text{Tr } J < 0 \Leftrightarrow c < (1-k)\left(2\lambda\frac{(n+2)}{n} - 3\frac{(n+3)}{n}\right) \tag{A.40}$$

Either of these conditions may be binding thus c has to be less than the minimum of the two critical values.

(ad ii) To proof part two of Lemma 2 it is more convenient to calculate directly the eigenvalues of the Jacobian. They are given as

$$\varepsilon_1 = \frac{(n+1)[6\lambda k^2 - \lambda^2(1-k)^2] - cn[(\lambda-6)(1-k)+2] + \lambda[3n(1-3k)+1-7k]}{n[(2\lambda-6)(1-k)+2]} \tag{A.41}$$

$$\varepsilon_2 = (\lambda(1-k)(n+1) - 3 - n(1+c))\frac{n(3-c)+1-(n+1)(\lambda(1-k)+6k)}{n[(2\lambda-6)(1-k)+2]} \tag{A.42}$$

First, consider ε_i. Two cases have to be distinguished. If $[(2\lambda - 6)(1 - k) + 2] < 0$ (which implies $(\lambda - 6)(1 - k) + 2 < 0$), then

$$\varepsilon_1 < 0 \Leftrightarrow c > \frac{(n + 1)\lambda(1 - k)[(1 - k)(\lambda - 6) + 5] - 2\lambda n(1 - k)}{n[(1 - k)(\lambda - 6) + 2]} \equiv c_1. \qquad \text{(A.43)}$$

Remember that for $[(2\lambda - 6)(1 - k) + 2] < 0$ the two types fixed point is a Nash equilibrium if and only if $c < c_1$ (cp. (A.36)). Thus, whenever the Nash equilibrium in this case exists it is unstable. If $[(2\lambda - 6)(1 - k) + 2] > 0$, two further subcases have to be considered.

(a) $(\lambda - 6)(1 - k) + 2 < 0$:

$$\varepsilon_1 < 0 \Leftrightarrow c < c_1. \qquad \text{(A.44)}$$

This again contradicts the existence of a Nash equilibrium for this parameter region.

(b) $(\lambda - 6)(1 - k) + 2 > 0$:

$$\varepsilon_1 < 0 \Leftrightarrow c > c_1 \qquad \text{(A.45)}$$

which clearly is identical to (A.35). Thus, in this parameter region the Nash equilibrium exists and the first eigenvalue is negative.

Similar reasoning for the second eigenvalue yields that it may be negative in the case $[(2\lambda - 6)(1 - k) + 2] > 0$. Indeed, this happens if

$$\frac{n + 1}{n}\lambda(1 - k) - \frac{3 + n}{n} > c > \frac{n + 1}{n}[(1 - k)(6 - \lambda) + 2] - \frac{3 + n}{n} \qquad \text{(A.46)}$$

Clearly, the second part of this inequality is identical to (A.30). Since the condition for the two types fixed point to be a Nash equilibrium is at least as restrictive as (A.30) we conclude that both eigenvalues are negative if

$$c_3 > c > c_1. \qquad \text{(A.47)}$$

(ad iii) The determinant of the Jacobian evaluated at the Hawk only fixed point is obtained as

$$\text{Det } J = -\frac{3 + n}{n}\left[\frac{(n + 1)\lambda(1 - k)}{n} - \frac{n(1 + c) + 3}{n}\right] > 0$$
$$\Leftrightarrow c > \frac{n + 1}{n}\lambda(1 - k) - \frac{3 + n}{n} \qquad \text{(A.48)}$$

The trace is calculated as

$$\text{Tr } J = -\frac{3 + n}{n} + \lambda\frac{(n + 1)(1 - k)}{n} - \frac{n(1 + c) + 3}{n} < 0$$
$$\Leftrightarrow c > \frac{n + 1}{n}\lambda(1 - k) - \frac{2n + 6}{n} \qquad \text{(A.49)}$$

It is easy to show that (A.48) is always more restrictive than (A.49). This completes the proof on local stability.

Before proving global stability, some further results need to be stated.   □

Lemma 3. *Consider any 2×2 matrix A. If Tr A>0 then at least one eigenvalue of A is strictly positive or both complex conjugated eigenvalues have positive real parts.*

Proof. The proof is straightforward. □

Definition. (Permanence): A given Replicator Dynamic is said to be permanent if there exists some $\delta > 0$ such that all paths $v(\cdot)$ with $v_i(0) > 0$ for some $i = 1, 2, \ldots, m$ satisfy $\liminf_{t \to \infty} v_i(t) > \delta$

Theorem 1. *(Hofbauer and Sigmund, 1987): Let p be a fixed point in the interior of the simplex Δ_n and let p satisfy $p \cdot Ax > x \cdot Ax$ for all fixed points $x \in bd\Delta_n$. Then the Replicator Dynamic is permanent.*

Proof. See Vega–Redondo (1996), pp. 65–66. □

Theorem 2. *(Hofbauer and Sigmund, 1987): Let $v(\cdot) \subset \mathrm{int}\,\Delta_n$ be any path of a permanent Replicator Dynamic. It satisfies*

$$\lim_{T \to \infty} \frac{1}{T} \int_{t=0}^{T} v(t)\, \mathrm{d}t = v^*$$

where v^* is the unique interior fixed point.

Proof. See Vega–Redondo (1996) pp. 67–68. □

Proof of Proposition 2. Now, we can start proving Proposition 2 and show that the Nash equilibria for the parameter regions given in Lemma 2 are not only locally, but globally stable. To prove this, it would be useful to clarify the question of whether there can coexist several locally stable NE. Concerning existence, two cases can occur, $c_1 < c_3$ or $c_3 < c_1$. In the first case it turns out that $c_2 > c_3$ and in the second case the opposite is true. The possible constellations can be illustrated in the following way:
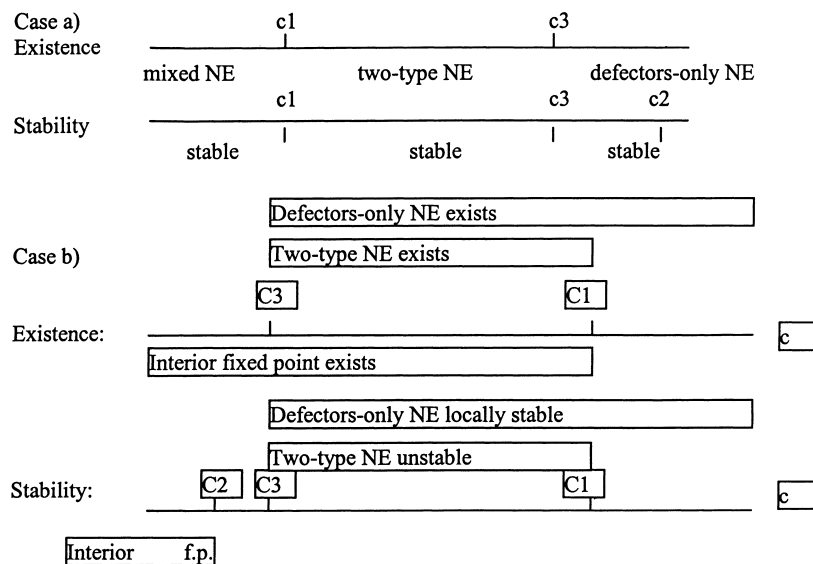
Consider case (a). For $c > c_1$ no interior fixed point exists. From Poincaré–Bendixson we know that any limit set of a regular dynamic system in the plane has to be a fixed point or a limit cycle. Hofbauer and Sigmund (1988) have shown that the replicator dynamics for dimensions $n \leq 3$ cannot have any isolated cycles in the interior of the simplex. So in case (a) the only possible limit cycle would be the boundary. When it is reached, either we are in the parameter region where the two-type NE is locally stable or the defectors-only NE is locally stable. Clearly, the boundary could not be a limit cycle, because this would contradict the existence of locally stable fixed points. Since stable states do not coexist, for $c_1 < c < c_3$ the two-type NE is globally stable and for $c > c_3$ the Hawks-only NE is globally stable. Now consider case (b). Clearly, for $c > c_2$ the trace of the Jacobian at the interior fixed point becomes positive. By Lemma 3 this implies that the fixed point becomes unstable. Then: (i) $c_2 < c < c_3$: any interior solution for $t \rightarrow \infty$ reaches $bd\Delta_3$ which is the unique limit set in this case. (ii) $c > c_3$: the defectors only NE is the only locally stable NE in this parameter region, hence it is globally stable (all interior solution paths still tend to the boundary by the same argument as above). Finally consider the case where the interior fixed point is locally stable, i.e. $c < \min\{c_1, c_2\}$. In both cases no other stable fixed points coexist. Since the limit set consists only of fixed points or limit cycles, again by the Hofbauer–Sigmund argument the only stable limit cycle which could exist would be the boundary of the simplex. But it is straightforward to show that the conditions of Theorem 2 are fulfilled if $c < \min\{c_1, c_2\}$, i.e. the Replicator Dynamic is permanent for the interior fixed point with respect to the three corner fixed points and the one edge fixed point. Therefore, the unique interior fixed point is globally stable. $\square$

References

Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated Prisoners' Dilemma: experimental evidence. Economic Journal 103, 570–585.

Amann, E., C.-L. Yang, 1994. The robustness of a sophisticated mutant in an evolutionary model of cooperation, discussion paper No. 94-03, University of Dortmund.

Axelrod, R., 1981. The emergence of cooperation among egoists. American Political Science Review 75, 306–318.

Axelrod, R., 1984. The Evolution Of Cooperation. Basic Books, New York.

Bomze, I.M., 1986. Non-cooperative, two-person games in biology: a classification. International Journal of Game Theory 15, 31–57.

Börgers, T., 1996. On the relevance of learning and evolution to economic theory. Economic Journal 106, 1374–1385.

Camerer, C., Weigelt, K., 1988. Experimental tests of the sequential equilibrium reputation model. Econometrica 56, 1–36.

Conlisk, J., 1980. Costly optimizers versus cheap imitators. Journal of Economic Behavior and Organization 1, 275–293.

Cooper, R., de Jong, D.V., Forsythe, R., Ross, T.W., 1996. Cooperation without reputation: experimental evidence from prisoner's dilemma games. Games and Economic Behavior 12, 187–218.

Frank, R.H., 1987. If homo economicus could choose his own utility function, would he want one with a conscience? American Economic Review 77, 593–604.

Friedman, D., 1991. Evolutionary games in economics. Econometrica 59, 637–666.

Fudenberg, D., Maskin, E., 1986. The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–554.

Güth, W., 1997. Sequential versus independent commitment — An indirect evolutionary analysis of bargaining rules. Humboldt University Discussion Paper – Economics Series, 106.

Güth, W., Kliemt, H., 1995. Evolutionarily stable co-operative commitments. Humboldt University Discussion Paper, Economics Series 53.

Güth, W., Kliemt, H., 1996. Towards a completely indirect evolutionary approach — A note. Humboldt University Discussion Paper, Economics Series 82.

Güth, W., Kliemt, H., 1998. The indirect evolutionary approach: bridging the gap between rationality and adaptation. Rationality and Society 10, 377–399.

Güth, W., Kliemt, H., Peleg, B., 1998. Co-evolution of preferences and information in simple games of trust. Humboldt University Discussion Paper, Economics Series 123.

Güth, W., Schmittberger, R., Schwarze, R., 1982. An experimental study of ultimatum bargaining. Journal of Economic Behavior and Organization 3, 367–388.

Güth, W., Yaari, M., 1992. An evolutionary approach to explaining reciprocal behavior in a simple strategic game. In: Witt, U. (Ed.), Explaining Process and Change. The University of Michigan Press, Ann Arbor.

Guttman, J., 1996. Rational actors, tit-for-tat types, and the evolution of cooperation. Journal of Economic Behavior and Organization 29, 27–56.

Hofbauer, J., Sigmund, K., 1988. The Theory of Evolution and Dynamical Systems. Cambridge University Press, Cambridge.

Isaac, R.M., Walker, J.M., Williams, A., 1994. Group size and the voluntary provision of public goods: experimental evidence utilizing very large groups. Journal of Public Economics 54, 1–36.

Kandori, M., 1992. Social norms and community enforcement. Review of Economic Studies 59, 63–80.

Kreps, D., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated Prisoners' dilemma. Journal of Economic Theory 27, 245–252.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook Of Experimental Economics. Princeton University Press, Princeton, NJ.

Milgrom, P., North, D., Weingast, B., 1990. The role of institutions in the revival of trade: the law merchant, private judges and champagne fairs. Economics and Politics 2, 1–23.

Ochs, J., Roth, A., 1989. An experimental study of sequential bargaining. American Economic Review 79, 355–384.

Offerman, T., Sonnemans, J., Schram, A., 1996. Value orientations, expectations and voluntary contributions in public goods. Economic Journal 106, 817–845.

Palfrey, T.R., Rosenthal, H., 1988. Private incentives in social dilemmas: the effects of incomplete information and altruism. Journal of Public Economics 35, 309–332.

Rapoport, A., Chammah, A.M., 1965. Prisoners' Dilemma. University of Michigan Press, Ann Arbor.

Robson, A., 1990. Efficiency in evolutionary games: Darwin, Nash and the secret handshake. Journal of Theoretical Biology 144, 376–396.

Roth, A.E., 1988. Laboratory experimentation in economics: a methodological overview. Economic Journal 98, 974–1031.

Roth, A.E., 1995. Bargaining experiments. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook Of Experimental Economics. Princeton University Press, Princeton, NJ.

Rubinstein, A., 1979. Equilibrium in supergames with the overtaking criterion. Journal of Economic Theory 21, 1–9.

Schelling, T., 1960. The Strategy Of Conflict. Harvard University Press, Cambridge.

Selten, R., 1991. Evolution, learning and economic behavior. Games and Economic Behavior 3, 3–24.

Vogt, C., 1998. Endogenous learning mutants in an evolutionary prisoner's dilemma, mimeo.

Weibull, J., 1995. Evolutionary Game Theory. MIT Press, Cambridge.

Weimann, J., 1994. Individual behavior in a free riding experiment. Journal of Public Economics 54, 185–200.

Yang, C.-L., 1994. Evolution of cooperation in prisoners' dilemma game with a knowledgeable mutant, discussion paper No. 93-09, University of Dortmund.