

# Property Rights and Investments: An Evolutionary Approach

Luciano Andreozzi  
Università degli Studi di Cassino

March 15, 2006

## Abstract

When contracts are not enforceable, or property rights are not clearly defined, individuals might lack an incentive to carry out costly investments even when they are socially efficient. The reason for this is that part of the revenues of the investment might be reaped by other parties. This is the so called hold-up problem. Some recent contributions point out that this problem might be less dramatic in real economies than standard economic models would suggest. These models are based on standard results in stochastic evolutionary game theory. They prove that only those equilibria in which the investing agent gets a sufficiently large share of the total revenues to induce her to invest can be stochastically stable. They conclude that evolution will tend to select contracts and institutions that mitigate, or eliminate, the hold-up problem. The present paper shows that these results are not robust with respect to the introduction of individual heterogeneity. It presents a model in which individuals have different cost functions and shows that stochastically stable states are inefficient, even if they induce a positive (suboptimal) level of investment. Some consequences for the evolutionary approach to social institutions are also discussed.

**Very first draft. Please do not quote.**

## 1 Introduction

Property rights are crucial for the correct working of virtually any economic activity. When individuals are not sure whether they will be able to enjoy the fruits of their own labor because it might be confiscated by a political authority, investments that are rational from an economic point of view might fail to be undertaken. Similarly, joint ventures in which participants are required to make relation-specific investments will usually fail to work efficiently. Unless a way is found to protect the investment made by each participant from the opportunistic behavior of the others, the level of effort put in the venture will be less than optimal.

Given the ubiquity of this problem, it is not surprising that a large literature emerged on the impact of property rights (or the lack thereof) on economic activities. Part of this literature goes under the heading of hold-up problem. In its simplest form the hold-up problem can be described as follows. (See Tirole [5] for a textbook presentation.) There are two risk-neutral agents  $A$  and  $B$ .  $A$  can either invest or not in the production of a pie that has a positive value for both  $A$  and  $B$ . By hypothesis, there is no way for them to sign a binding contract before  $A$  makes his investment decision. *After* (and if) the pie has been produced,  $A$  and  $B$  contract over the division of the pie itself.  $A$  will thus undertake an investment only if it costs less than the value of what she expects to get from the bargaining. Assuming that the pie is divided equally between the two players,  $A$  will only invest if the cost of investment is less than half the value of the pie. This implies that many investments that are economically efficient (i.e. whose revenues are larger than the cost) will fail to be undertaken because  $A$  expects to be (partially) expropriated by  $B$ .

Some recent contributions, notably Ellingsen and Robles [2] and Troger [6], have developed evolutionary models whose conclusions differ markedly from the standard analysis of the hold-up problem I have just presented. They build on the evolutionary approach to bargaining due to Young [8][7]. In Young's approach, players are replaced by large populations of agents. The game is played repeatedly by pairs of individuals drawn at random from these populations. Bargaining follows the Nash demand game protocol. Each agent claims a fraction of the pie. If the claims are compatible, each player obtains what she has claimed. Otherwise, nobody gets anything.

Occasionally, before playing the game, agents are able to observe the way in which it has been played by previously by other agents. Agents will then play their best strategy against the distribution of strategies they have observed. Over time, agents converge towards one solution to the bargaining game, which becomes the conventional way of solving that particular game. For example, they might converge to a state in which agents in the  $A$  population get sixty per cent of the pie and those in  $B$  population the remaining forty percent. Once a convention has been established, it pays to each agent to stick to it. For example, if all agents in  $A$  claim sixty percent of the pie, it doesn't pay  $B$  to switch to a demand which is higher than forty percent, because in so doing she would get nothing. Of course, it would be irrational to claim less than forty percent. The first example that comes to mind of conventional contracts of this kind is the customary way of dividing the crop evenly between landlords and tenants (the so called *mezzadria*).

While Young assumes that bargaining takes place over a fixed and exogenous pie, Ellingsen and Robles [2] and Troger [6] endogenize the size of the pie assuming that it is the result of an investment decision made by one of the two agents. The result they get is striking: the two populations will gravitate around efficient equilibria in which the investing agent appropriates a share of the pie which is large enough to cover the costs she paid to bring it about. Investments will thus be undertaken whenever they are efficient. They conclude that the hold-up problem could no be so severe after all, because evolution will promote

the emergence of norms and institutions (in the form of bargaining conventions) that mitigate or eliminate it.

There is a lack of reference to any instance of existing institution in these analyses which is surprising even for theoretical papers. A likely reason is that the standard reference to the fifty-fifty crop division between landlord and tenants would be the most obvious counterexample. This custom has traditionally been criticized (for example by Marshall) for being the source of many inefficiencies in backward agrarian countries such as Italy in the XIX century. In general, there seems to be little empirical ground to the claim that evolution will promote conventional contracts which protect investments and efforts. (See Dawid and McLeod [1] for a more detailed discussion of this point.)

This paper examines this issue from the same perspective as in Ellingsen and Robles [2] and Troger [6], but reaches a different conclusion. It shows that under a seemingly minor modification of their model, the population will fail to converge to a state in which investors appropriate a sufficiently large share of their efforts, so that investment becomes individually rational. Quite on the contrary, the model shows that the system will tend to gravitate around states in which the investment level is suboptimal.

The paper proceeds as follows: Section 2 provides an informal presentation of the model and presents the main result. Section 3 introduces the definitions and the technicalities of the model. Section 4 contains a formal statement of the main result. Section 5 concludes. Proofs are collected in the Appendix.

## 2 Informal presentation and motivation

I follow as closely as possible the approach of Ellingsen and Robles [2], who in turn build upon Noldeke and Samuelson [3]. (Troger [6] presents a different approach which is closer to the original Young's models.) The reader is to refer to the original papers for most of the mathematical details.

There are two risk-neutral individuals  $A$  and  $B$ .  $A$  moves first and decides whether to invest or not in the production of a pie. If she decides not to invest, the game is over and  $A$  and  $B$  get nothing. If  $A$  invests, she pays a cost  $c \in (0, 1)$  and a pie of size one is produced. At this point in time,  $A$  and  $B$  bargain over the division of the pie. The standard hold-up argument proceeds as follows. After production has taken place, the investment cost  $c$  is sunk and therefore should not influence  $A$ 's behavior. Bargaining between two risk neutral players would then assign half of the pie to each, regardless of who produced the pie. As a consequence, production would take place only if player  $A$ 's cost is so small that investment yields a positive return even when  $A$  appropriates only half of the available pie. This requires that  $c < \frac{1}{2}$ . When this is not the case,  $A$  will not produce the pie even if production is socially efficient because it maximizes consumers' surplus. Inefficiencies will then occur whenever  $c \in (\frac{1}{2}, 1)$ .

This result is crucially based on the hypothesis that in the second stage of the game, when (and if) the pie has been produced, the only reasonable outcome of bargaining is the Nash bargaining solution. For instance, this happens

if the two players bargain according to the alternating offer protocol proposed by Rubinstein [4]. In such an approach, the game has a single subgame perfect Nash equilibrium, which approximates the Nash bargaining solution in the second stage of the game and therefore implies that no production will take place whenever  $c > \frac{1}{2}$ .

A different result would follow if, instead of the alternating offer protocol,  $A$  and  $B$  played a round of the Nash demand game. Suppose that, if production has taken place, each player claims a fraction  $x_i \in (0, 1]$  of the available pie ( $i = A, B$ ). If  $x_A + x_B \leq 1$ , each player gets the fraction of the pie she has claimed. If  $x_A + x_B > 1$  they both get nothing. Since both players are risk neutral, their payoffs are  $u_A = x_A - c$  for player  $A$  and  $u_B = x_B$  when  $A$  produces the pie and  $x_A + x_B \leq 1$ . When  $A$  does not produce a pie players get zero. When  $A$  produces and they fail to reach an agreement about the best way of dividing it (that is if  $x_A + x_B > 1$ ),  $A$  gets  $-c$  and  $B$  gets zero.

One can easily see that when production is followed by the Nash demand game a multitude of subgame perfect equilibria exist, some of which involve production while others do not. To see this, consider that any pair of strategies  $(x, 1 - x)$  for the second stage of the game forms a subgame perfect equilibrium in which  $A$  invests in the first stage, provided that  $x \geq c$ . All these equilibria are socially efficient (provided that  $c < 1$ ), although they differ in the way in which surplus is divided between  $A$  and  $B$ . Beside these, there are other subgame perfect equilibria in which production does not take place, because in the second stage  $B$  expects  $A$  to claim a fraction of the pie  $x < c$ , and he optimally (given his expectations) claims a fraction of the pie  $1 - x > 1 - c$ .

Inefficient equilibria in which production does not take place are less compelling than equilibria in which  $A$  does not invest. Inefficient equilibria are sustained by  $A$ 's believing that  $B$  will claim a share of the pie larger than  $(1 - c)$ . Such a belief would only be rational if  $B$  expected  $A$  to produce the pie and claim less than  $c$ . This strategy yields a negative payoff for sure to  $A$  and hence it is strictly dominated (in the normal form version of the game) by the strategy consisting in not producing the pie at all. Once strictly dominated strategies are eliminated, only Nash equilibria in which production does take place survive.

Efficient equilibria are also the only ones that are compatible with the forward induction principle. To see this, consider that if  $A$  has opted for investing, either he made a mistake, or he hoped to coordinate on one of the equilibria in which he gets a positive payoff. Since the only equilibria which guarantee this are those in which  $A$  appropriates at least a fraction  $c$  of the total pie,  $B$  can make sense of  $A$ 's decision to invest only assuming that he is going to claim at least  $c$ .

Not surprisingly, the evolutionary approach to this game proposed by Ellingsen and Robles [2] and Troger [6] shows that evolution will sustain production in this class of games. These results exploit (via different routes) the fact that evolutionary models support forward induction to a larger extent than standard models based on rationality alone, a result originally obtained by Noldke and Samuelson [3]. The role of "evolution" in this context is to coordinate

agents' expectations on one among the many equilibria of the bargaining game which can sustain investment.

These results are based on the following, straightforward piece of reasoning. As the game is repeated over time, each agent in both populations updates her beliefs about the behavior of agents in the other population on the basis of her observations of the current play of the game. However, expectations concerning the behavior at information sets that are never reached (given the current states of the two populations) cannot be updated and therefore are subject to change randomly, a process usually referred to as *drift*. (See the next section for technical details)

Now take one of the inefficient subgame perfect equilibria in which  $A$  agents do not invest, because they expect agents in population  $B$  to claim a share of the pie larger than  $(1 - c)$ . Since production does not take place, what agents in population  $A$  and  $B$  would claim in case a pie were produced is not observed, beliefs are not updated and they are prone to drift. Because of drift,  $B$  agents might become convinced that, in case a pie were produced, agents  $A$  would demand a share  $x$  larger than  $c$ . They would then be willing to claim a share  $(1 - x)$  of the pie, which is smaller than  $c$ .

When this happens, it takes a single  $A$  agent to switch (by mistake, given her expectations) to production, to reveal that, in fact,  $B$  players are willing to claim  $(1 - x)$ . Following this event,  $A$  agents will switch to a best reply to their new expectations, which consists in producing the pie and claiming  $x$ . All  $B$  players will then start claiming  $(1 - x)$  and a new convention is selected. This argument shows that drift plus a single mutation can bring the system from any equilibrium in which production does not take place, to a state in which it does.

Destabilizing efficient equilibria is not so easy, though. Take one of the equilibria in which all agents in population  $A$  claim a fraction  $x > c$  and agents in  $B$  claim  $1 - x$ . One can easily see that these states are not subject to drift because, since production does take place, agents' claims are revealed at each interaction. Efficient equilibria in which production takes place are thus more stable than those in which  $A$  players do not produce. (Formally, equilibria corresponding to no production are not locally stable, while those corresponding to production are.)

The obvious weak part of this result is that it relies so heavily on drift. To make this argument work one must assume that agents within each population are exact clones one of another. This assumption is needed because it implies that there is a single threshold level (equal to  $c$ ) such that if agents  $A$  expect to get a share of the pie larger than  $c$  they will all invest in production, while none of them would invest if she expects to get less than  $c$ . So  $B$ 's expectations about  $A$ 's behavior will drift if  $x < c$ , and will not if  $x > c$ . In more realistic settings, however, agents in population  $A$  will show a certain degree of heterogeneity. Some of them will be more efficient (low  $c$ ), others will be less efficient (high  $c$ ). This difference will be reflected on the fact that  $A$  agents will have a different threshold over which production becomes profitable: a share of the pie  $x$  might be large enough to induce efficient  $A$  agents to produce, while being too small to make production profitable for less efficient  $A$  agents.

This paper investigates the effects of introducing this kind of heterogeneity in the  $A$  population. To do this, while retaining the assumption that all agents are risk neutral, I stipulate that agents in population  $A$  are divided into two groups: efficient agents whose cost  $c_L \in (0, 1)$  is low and inefficient agents whose cost  $c_H \in (c_L, 1)$  is high. Notice that the hypothesis that even the less efficient agents  $A$  have a cost smaller than 1 insures that production would be socially efficient both for efficient and for not efficient agents. The only efficient states are thus those in which *all*  $A$  agents invest. The main result, presented in Section 4, shows that it is not generally true that the only stable states are those in which all  $A$  agents invest in production. There are locally stable states in which only the most efficient ones produce. Furthermore, our model shows that the two populations will tend to gravitate around a state in which the least efficient agents in  $A$  will not produce. (Formally, states in which *all*  $A$  agents produce fail to be stochastically stable, despite being locally stable.)

To have a feeling of how the model works, suppose that  $c_L < \frac{1}{2} < c_H < 1$ . It follows that efficient  $A$  agents are willing to invest in production even if they expect to get only half of the pie, while inefficient agents find production worthwhile only if they appropriate more than one half. Suppose agent's beliefs are coordinated on the equilibrium in which the pie is split equally between  $A$  and  $B$ . In such a situation, only efficient  $A$  players would produce a pie, which is an inefficient arrangement. However, this state is not subject to drift, because a pie is produced by some of the agents in population  $A$ , and hence agents' beliefs concerning the bargaining stage of the game are constantly updated. The main result I shall prove is that (under mild technical conditions), when  $c_L < \frac{1}{2}$ , the state in which the pie is divided equally and only efficient  $A$  agents invest in the production of a pie is the only stochastically stable state.

### 3 Technical details

There are two large populations  $A$  and  $B$  of agents numbering  $N_A$  and  $N_B$  individuals each. Population  $A$  is divided into two subpopulations: there are  $N_{AL}$  efficient players whose cost of producing a pie is  $c_L$  and  $N_{AH}$  inefficient players whose cost of producing a pie is  $c_H$  ( $N_A = N_{AL} + N_{AH}$ ). The pie has value one for all  $A$  and  $B$  agents.

In keeping with evolutionary models of this kind, I will assume that there is an equal number of each type of agent, so that  $N_{AL} = N_{AH} = N_B$ . This hypothesis is introduced only for its expository convenience. Other hypotheses (for example,  $N_B = N_A = 2N_{AL} = 2N_{AH}$ ) could be used as well, getting qualitatively similar results.

In each period  $t = \{1, 2, \dots\}$  all possible combinations of agents from the two populations form and play a match of the bargaining plus production game. In order to work with a finite strategy space, I shall make the standard assumption that the pie is divisible up to a fraction  $\delta = \frac{1}{n}$ , where  $n$  is the number of possible divisions. So each player's strategy space in the second stage of the game is

$\Delta = \{\delta, 2\delta, \dots, 1 - \delta, 1\}$ .<sup>1</sup> For mathematical convenience I shall also assume that  $n$  is an even number, so that each player can always claim half of the pie.

Each agent's state is characterized by a combination of a strategy and a belief. For a  $B$  agent a strategy is simply a share  $x$  of the pie to claim in the bargaining stage of the game. For an  $A$  agent a strategy is a decision whether to produce or not a pie and the share of it to claim in the bargaining.

Let  $\nu(\cdot)$  be the beliefs of agents  $A$  concerning the behavior of agents in population  $B$ . Similarly, agents in population  $B$  have beliefs  $\sigma(\cdot)$  over the demands of agents in population  $A$ . Both  $\sigma(\cdot)$  and  $\nu(\cdot)$  are probability distributions over  $\Delta$ . Notice that  $B$ 's beliefs are not conditioned on  $A$ 's cost, because we shall assume that while the outcome of the game is observed by other players, payoffs are not. This implies that  $B$  agents can observe whether an  $A$  agent has produced a pie and (if he did) how much of it it claimed for himself. However, they cannot observe how much it cost him to produce it.

Let  $\theta$  be the state of the population.  $\theta$  specifies how many agents in both populations have every possible combination of belief and strategy.  $z(\theta)$  is the distribution across the terminal nodes of the game, given the state of the population  $\theta$ .

The state  $\theta$  evolves over time on the basis of adaptation and random mutation. At the beginning of each period, each agent in both populations receives with a fixed probability the opportunity to revise his strategy and his beliefs. In this case we say that she receives a learning draw. An agent who receives a learning draw observes  $z(\theta)$  and adjusts his beliefs accordingly. Notice that since all agents who receive the learning draw observe the same distribution  $z(\theta)$  over outcomes, they will all update in the same manner. Notice also that adjustments in beliefs takes place only in those states  $\theta$  in which at least one  $A$  agent produces a pie, because only in this event the bargaining stage of the game is reached. In the other cases, the learning draw has no effect on agents' beliefs. After adjusting his beliefs, an agent chooses a best reply to them. If there are multiple best replies, one of them is chosen at random.

Beside learning, the state  $\theta$  evolves through mutation. Every period each agent has a probability  $\varepsilon$  to mutate. An agent who mutates chooses a combination of belief and strategy at random. When mutation produces a change in a strategy corresponding to a node that is not reached under the current state  $\theta$ , or when it changes a belief concerning an unreached node, we say that the state *drifts*.

The combination of learning and mutation generates a Markov chain over the states  $\theta$ . When  $\varepsilon > 0$ , this Markov chain is irreducible because every possible transition among all states have positive probability. Therefore, there is a stationary distribution  $\mu(\varepsilon)$  which the system will approach regardless of initial conditions. We are interested in the stationary distribution for very small values of the noise  $\varepsilon$ . Formally, we are interested in the limit distribution  $\mu^* = \lim_{\varepsilon \rightarrow 0} \mu(\varepsilon)$ . Those states that receive a positive probability in  $\mu^*$  are

---

<sup>1</sup>Notice that we are introducing the standard assumption that agents cannot claim nothing. They must claim at least  $\delta$ .

called *stochastically stable*.

Finally, we say that a state  $\theta$  is an *equilibrium* if there is no alternative state  $\theta'$  that can be reached from  $\theta$  without mutation. The *basin of attraction* of an equilibrium is the set of states  $B(\theta)$  from which there is a positive probability that the system can reach  $\theta$  without mutations.

## 4 Results

It will facilitate our discussion to partition the set of equilibria in three classes, labeled  $P_L$ ,  $P_H$  and  $NP$ . Intuitively, in  $NP$  there are all states  $\theta$  that are equilibria and in which none of the agents in  $A$  invests to produce a pie. In  $P_L$  there are all equilibria in which only the most efficient agents in  $A$  (whose cost is  $c_L$ ) invest, and in  $P_H$  there are all equilibria in which all agents in  $A$  produce a pie regardless their cost. Notice that the only efficient equilibria are those in  $P_H$ .

Formally, these three classes can be defined as follows:

Let  $v$  be an agent  $A$ 's beliefs concerning the demands of  $B$  agents. Let  $x^*$  be the optimal share of the pie to claim in the bargaining stage of the game, given expectations  $\nu$ . Let  $\pi^*$  be the expected size of the pie in the bargaining stage of the game for an  $A$  agent who claims  $x^*$  when her expectations are  $\nu$ . Let  $c_i$  be the cost of the  $A$  agent under consideration ( $i = H, L$ ). If  $\pi^* < c_i$ , the  $A$  agent will not invest in the first stage of the game regardless of their cost. Let  $v^*$  be the set of beliefs such that  $\pi^* < c_L$ . All states  $\theta$  such that every  $A$  agent has a belief in the set  $v^*$  are equilibria, regardless of the strategies and beliefs of  $B$  players. All such equilibria belong to the set  $NP$ . Notice that all states in  $NP$  yield the same outcome of the game, namely  $A$  does not produce a pie. They only differ in the expectations of  $A$  and  $B$ 's players and their strategies in the bargaining stage of the game.

To introduce  $P_L$  and  $P_H$  we need some more terminology. Let  $x_L$  be the smallest value of  $x \in \Delta$ , such that  $x > c_L$ . Similarly,  $x_H$  is the smallest value of  $x \in \Delta$  such that  $x > c_H$ . Among all possible divisions  $\Delta$ ,  $x_L$  is the most favorable to  $B$ , which still gives to the most efficient agents  $A$  an incentive to invest. Similarly,  $x_H$  is the convention which is most favorable to  $B$  and still gives an inefficient  $A$  agent an incentive to invest.

We make the following technical assumption:

**Assumption 1** *We assume that  $x_L < x_H$ .*

This assumption rules out the cases in which the same share  $x = x_L = x_H$  of the pie induces both efficient and inefficient players to invest. When it does not hold, one can easily prove that the model reduces to those discussed in the literature, in which evolution selects an efficient equilibrium. To see that this is a relatively innocent assumption, notice that it will hold whenever  $\delta$  is sufficiently small, because by hypothesis  $c_L$  is strictly smaller than  $c_H$ .

An equilibrium  $\theta$  belongs to  $P_L$  if agents  $A$  and  $B$  expectations and strategies are coordinated on one of the Nash equilibria of the bargaining game  $(x, 1 - x)$ ,

such that  $x_H > x \geq x_L$ . In these equilibria, efficient  $A$  agents invest, non efficient  $A$  agents do not invest. Finally, an equilibrium is in  $P_H$  when  $A$  and  $B$  beliefs are coordinated on one of the Nash equilibria  $(x, 1 - x)$ , such that  $x \geq x_H$ . In this case, both types of  $A$  agents invest.

The first step in proving our main theorem is the following:

**Proposition 2** *Equilibria in  $NP$  are not locally stable.*

**Proof.** Consider an equilibrium  $\theta \in NP$ . Since none of the  $A$  agents is producing a pie,  $B$ 's expectations about  $A$ 's behavior in the second stage of the game drift in consequence of mutations. Suppose  $B$ 's expectations drift to a point in which all  $B$  agents expect  $A$  agents to claim a share of the pie  $x > c_L$ . (From any state  $\theta \in NP$  this can be achieved by a series of one mutation transitions.) From this point, it takes a single mutation from an  $A$  to get to the basin of attraction of an equilibrium in  $P_L$ . To see this, suppose that when all  $B$  agents expect  $A$  agents to claim  $x > c_L$ , an  $A$  agent makes a mistake and produces a pie. (Notice that this is a mistake, because  $A$  agents have no way of knowing  $B$ 's beliefs.) All  $A$  agents who receive the learning draw would now observe that  $B$  players are claiming  $(1 - x)$ , and will thus switch to  $x$ . Learning alone can thus push the system to the convention  $(x, 1 - x)$ , which proves that we are within the basin of attraction of an equilibrium  $\theta \in P_L$ . ■

We are then ready for the main result of the paper.

**Proposition 3** *Let  $x^* = \max(\frac{1}{2}, x_L)$ . There is a single stochastically stable state in which agents' beliefs are coordinated on the equilibrium  $x^*$  and only the most efficient  $A$  agents produce a pie.*

While the proof is fairly involved, the intuition is straightforward. In fact, it is an immediate consequence of the evolutionary approach to bargaining pioneered by Young [7] [8]. The gist of the proof can be summarized as follows. From Proposition 2 above we know that only equilibria that are in  $P_L$  and  $P_H$  are locally stable. All we have to do is to construct a minimum resistance tree connecting these equilibria.

To illustrate the procedure, imagine that only local mistakes are possible: an agent who claims  $x$  can only switch (by mistake) to  $x + \delta$  or  $x - \delta$ . Let  $p^+$  be the number of  $A$  players who must switch to  $x + \delta$  in order to induce  $B$  agents to switch from  $(1 - x)$  to  $(1 - x - \delta)$ . Simple algebra suffices to show that  $p^+ = N \frac{\delta}{1-x}$  if  $x \in P_S$  and  $p^+ = 2N \frac{\delta}{1-x}$  if  $x \in P_L$ . The reason for this difference is that when  $x \in P_L$  only efficient agents produce a pie, while when  $x \in P_H$  all  $A$  agents produce a pie. As a consequence,  $B$  agents receiving a learning draw will observe a larger number of outcomes of the bargaining process when the existing convention is  $x \in P_L$ , rather than when the existing convention is  $x \in P_H$ .

Similarly, to move the population from  $(x, 1 - x)$  to  $(x - \delta, 1 - x + \delta)$  takes  $q^+ := N \frac{\delta}{x}$   $B$  agents to mutate from  $(1 - x)$  to  $(1 - x + \delta)$ . The picture below represents  $p^+$  and  $q^+$  when  $N = 30$ ,  $\delta = \frac{1}{10}$ ,  $c_S = 0.3$   $c_L = 0.75$ ). Dots represents equilibria. Where  $p^+ < q^+$ , transitions are easier towards larger

values of  $x$ . Where  $p^+ > q^+$  transitions are easier towards smaller values of  $x$ . The minimum resistance tree is thus rooted at  $x = \frac{1}{2}$ , so that the stochastically stable state is a state in which the pie is divided evenly between  $A$  and  $B$ , and only efficient  $A$  players invest. (This is because  $c_L < \frac{1}{2}$ .)

The picture below represents  $p^+$  and  $q^+$  for the same values of  $N$  and  $\delta$  as above, but now efficient  $A$  agents have a larger cost:  $c_L = 0.65$ . By the same reasoning as above, the stochastically stable state is  $(0.7, 0.3)$ . In this state  $A$  agents appropriate a fraction 0.7 of the produced pie, which gives efficient  $A$  agents an incentive to invest because  $0.7 > 0.65$ . However, less efficient  $A$  agents would invest if they could appropriate at least a fraction equal to .8 of the total pie, but such a convention is not stochastically stable. In fact, it takes less mutations to move from  $(.8, .2)$  to  $(.7, .3)$  than vice versa.

The main conclusion of the above reasoning is that the population will tend to gravitate around a state in which only the most efficient  $A$  players will invest, which is inefficient.

## 5 Conclusions

The results presented in the previous section cast doubt about the ability of evolutionary forces to shape efficient contracts. The main reason for the difference between my results and those presented in Ellingsen and Robles [2] and Troger [6], is that here it is assumed that the same contract (e.g. the fifty-fifty division) applies to a large variety of different cases. When there is large variation among individuals' productivity, the same contract might provide incentives to invest to some agents and not others. For example, the fifty-fifty division will be stochastically stable, even if other divisions (more favorable to the investing agents) would be more efficient. Incidentally, this seems to be in line with the empirical research on conventional contracts in agriculture. For example, Young and Burke [10] notice that "fifty-fifty is used in almost all farms despite the fact that there is a wide range of soil qualities on which it is used." (11-12) In general, their study shows that there is no evidence that conventional contracts are tailored to the characteristics of the single agents, like the models so far discussed in the literature would suggest. The present paper shows that when this is the case, the emerging conventional contract is likely to be inefficient.

## 6 Appendix. Proofs

We must construct the minimum resistance tree connecting all locally stable equilibria. Because of Proposition 2, they are all belong either to  $P_L$  or to  $P_H$ . Since only local mistakes are possible, for each convention  $x$  the system can only move to  $x + \delta$  and  $x - \delta$ . To move the system to  $x + \delta$  it takes a number  $p^+(x)$  of mutations in population  $A$ . Similarly, to move the system to  $x - \delta$  it takes a number  $q^+(x)$  of mutations in population  $B$ . With respect to the existing analysis of evolutionary models of bargaining, such as Binmore Samuelson and Young [10], we must take into consideration the fact that in some locally stable equilibria (those belonging to  $P_H$ ) all  $A$  agents invest so that agents who receive the learning draw will observe a number  $N_A N_B = (N_{AL} + N_{AH})N_B$  of pie divisions, while in other equilibria (belonging to  $P_L$ ) only efficient  $A$  agents will produce a pie so that learning agents will observe only  $N_{AL}N_B$  bargaining. Beside this difference, most of our analysis is very close to the one presented in Binmore, Samuelson and Young [10].

Consider first  $q^+(x)$ , that is the number of mutations that involve population  $B$ . The following lemma is a direct consequence of Young's original evolutionary approach to bargaining.

**Lemma 4** *Let  $\theta$  be an equilibrium in  $P_L$  or  $P_H$  in which the selected convention is  $(x, 1 - x)$ . The minimum number of agents in  $B$  population who must mutate to reach the basin of attraction of the convention  $(x - \delta, 1 - x + \delta)$ , is  $q^+(x) := \frac{\delta N_B}{x}$ .*

**Proof.** Suppose the system is at equilibrium  $(x, 1 - x)$ . Suppose  $q$  agents in  $B$  population mutate from claiming  $1 - x$  to claiming  $1 - x + \delta$ . In the bargaining stage of the game, an  $A$  agent who claims  $x$  gets a payoff equal to  $x(N_B - q)$ . If she switches to  $x - \delta$ , she gets a payoff  $(x - \delta)N_B$ . Playing  $x - \delta$  is a best reply provided that  $x(N_B - q) < (x - \delta)N_B$ , which requires  $q > \frac{\delta N_B}{x} := q^+(x)$ . ■

Consider now how mutations in population  $A$  can take the system from a convention  $(x, 1 - x)$  to convention  $(x + \delta, 1 - x - \delta)$ . This number will depend upon the original convention belonging to  $P_L$  or  $P_H$ .

**Lemma 5** *Let  $\theta$  be an equilibrium in  $P_L$  in which the selected convention is  $(x, 1 - x)$ . It takes  $p_L^+(x) := \frac{\delta N_{AL}}{1 - x}$  mutations in population  $A$  to enter the basin of attraction of convention  $(x + \delta, 1 - x - \delta)$ . If  $\theta \in P_H$  it takes  $p_H^+(x) := \frac{N_A \delta}{1 - x}$  mutations to reach the basin of attraction of  $(x + \delta, 1 - x - \delta)$ .*

**Proof.** Consider an equilibrium  $\theta \in P_L$  corresponding to convention  $(x, 1 - x)$ . Without mistakes, there would be  $N_{AL}$  agents in population  $A$  producing a pie. Suppose there are  $p$  mutations in population  $A$ . These mutations can be of three kinds: (a) low cost  $A$  agents who continue to produce a pie and claim  $x + \delta$ , (b) low cost agents who stop producing a pie and (c) high cost  $A$  agents who start producing a pie and claim  $x + \delta$ . I shall indicate with  $p_1, p_2$  and  $p_3$  the number of  $A$  agents who make these three classes of mistakes ( $p_1 + p_2 + p_3 = p$ ). A  $B$  agent claiming  $(1 - x)$  would get a payoff  $(1 - x)(N_{AL} - p_1 - p_2)$ . A  $B$  agent

claiming  $(1-x-\delta)$  gets a payoff  $(1-x-\delta)(N_{AL}-p_2+p_3)$ . Claiming  $1-x-\delta$  is a best reply provided that  $(1-x-\delta)(N_{AL}-p_2+p_3) > (1-x)(N_{AL}-p_1-p_2)$ , that is

$$p_1(1-x) - p_2\delta + p_3(1-x-\delta) > \delta N_{AL} \quad (1)$$

We must minimize the number of mutations  $p = p_1 + p_2 + p_3$  subject to 1 This obviously requires that  $p_2 = p_3 = 0$ , and  $p = p_1$ . Hence, the minimum number of mutations involve only efficient agents in the  $A$  population who continue to produce a pie and claim  $x + \delta$ . With this proviso, the minimum number of mutations it takes to reach the basin of attraction of  $(x + \delta, 1 - x - \delta)$  is given by  $p(1-x) > \delta N_{AL}$ , that is  $p > \frac{\delta N_{AL}}{1-x} := p_L^+(x)$ .

If  $\theta \in P_H$ , without mistakes all  $A$  agents (low and high cost) will produce a pie. Suppose  $p$  agents in population  $A$  mutate.  $p_1$  of them continue to produce a pie and claim  $x + \delta$ ;  $p_2$  of them stop producing a pie. The total number of mistakes is  $p = p_1 + p_2$ . A  $B$  agent claiming  $(1-x)$  gets a payoff equal to  $(1-x)(N_A - p_1 - p_2)$ . A  $B$  agent claiming  $(1-x-\delta)$  gets a payoff  $(1-x-\delta)(N_A - p_2)$ . Therefore, claiming  $(1-x-\delta)$  is a best reply provided that

$$(1-x-\delta)(N_A - p_2) > (1-x)(N_A - p_1 - p_2)$$

which requires  $p_1(1-x) + p_2\delta > N_A\delta$ . Since  $(1-x) \geq \delta$ , this requires that  $p_2 = 0$ , so that  $p > \frac{N_A\delta}{1-x} := p_H^+(x)$  ■

We are thus ready for the proof of Proposition 3:

**Proof.** The proof is based on an algorithm known as *naive minimization test*. (See the original Binmore, Samuelson and Young [10] paper for the details.) Consider a state  $x \geq x^* = \max(\frac{1}{2}, x_L)$  and suppose first that  $x_L > \frac{1}{2}$  so that  $x^* = x_L$ . All locally stable equilibria are  $(x, 1-x)$  with  $x \geq x_L > \frac{1}{2}$ . Consider now the two outgoing resistances emanating from each locally stable equilibrium. If  $x < x_H$ , these are  $q^+(x)$  and  $p_L^+(x)$ . The minimum of these magnitudes is  $q^+(x)$ . To see this consider  $q^+(x) < p_L^+(x)$  iff  $\frac{\delta N_B}{x} < \frac{\delta N_{AL}}{1-x}$ . Since we assume that  $N_{AL} = N_B$ , this requires that  $1 < \frac{x}{1-x}$ , which holds because  $x \geq x_L > \frac{1}{2}$ . Similarly, if the convention  $(x, 1-x)$  belongs to  $P_H$ , one must have that  $q^+(x) < p_H^+(x)$ , which requires  $\frac{\delta N_B}{x} < \frac{N_A\delta}{1-x} = \frac{2N_B\delta}{1-x}$ . Simple algebraic manipulation suffice to show that this requires  $\frac{1}{6} < x$  which is clearly satisfied because  $x \geq x_H > x_L > \frac{1}{2}$ .

Consider now the locally stable equilibrium  $(x_L, 1-x_L)$ . The lowest outgoing resistance is still  $q^+(x) = \frac{\delta N_B}{x_L}$ , which connects this equilibrium with the basin of attraction of the set of non locally stable states  $\theta \in PN$ . (If  $q^+$   $B$  players mutate from  $1-x_L$  to  $1-x_L-\delta$   $A$  players who receive the learning draw will switch to claiming  $x_L - \delta$  and will stop producing a pie.) However, because of Proposition 2, a single mutation suffices to move the system from any state  $\theta \in PN$  to any locally stable convention  $(x, 1-x)$ . The minimum outgoing resistance towards another locally stable outcome is thus  $q^+ + 1$ .

To produce a minimum resistance tree proceed as follows. Connect each locally stable state corresponding to convention  $(x, 1-x)$  with  $x > x_L$  to  $(x -$

$\delta, 1 - x + \delta$ ). Each of these transitions will have resistance  $q^+(x)$ . Connect the state corresponding to convention  $x_L$  to any locally stable convention. This branch will have resistance  $q^+(x_L) + 1$ . The resulting tree connects all locally stable states and contains a single loop. Since  $q^+(x)$  is decreasing in  $x$ , the loop contains the largest resistance among all the resistances in the tree, namely  $q^+(x_L) + 1$ . Removing that branch creates a tree rooted at  $x_L$  which is (because of the naive minimization test) the required minimum resistance tree. So the stochastically stable state is the one in which only efficient  $A$  agents invest and the selected convention is  $(x_L, 1 - x_L)$ .

The result when  $x_L < \frac{1}{2}$  proceeds along similar lines and will not be reproduced here. ■

## References

- [1] Dawid H. and W. Bentley McLeod (2005) "Hold-Up and the Evolution of Investment and Bargaining Norms", *mimeo*.
- [2] Ellingsen Tore and Jack Robles (2002) "Does Evolution Solve the Hold-Up Problem?", *Games and Economic Behavior*, 39: 28-53.
- [3] Noldeke G. and L. Samuelson (1993) "An Evolutionary Analysis of Backward and Forward Induction", *Games and Economic Behavior*, 5: 425-454.
- [4] Rubinstein A. (1982) "Perfect equilibrium in a bargaining model", *Econometrica*, 50: 97-109.
- [5] Tirole, Jean (1988) *The Theory of Industrial Organization*, MIT Press, Cambridge (Mass.) London.
- [6] Troger T. (2002) "Why sunk costs matter for bargaining outcomes: An evolutionary approach", *Journal of Economic Theory*, 102 (2): 375-402.
- [7] Young, P. H. (1993) "An Evolutionary Model of Bargaining", *Journal of Economic Theory*, 59: 145-168.
- [8] Young, P. H. (1998) *Individual Strategy and Social Structure: An Evolutionary Theory*, Princeton University Press, Princeton.
- [9] Young, P. H. and Mary A. Burke (2000) "Competition and Custom in Economic Contracts: a Case Study of Illinois Agriculture", *mimeo*.
- [10] Binmore Ken, Larry Samuelson and Peyton Young (2003) "Equilibrium Selection in Bargaining Models", *Games and Economic Behavior*, 45: 296-328.