Choosing Opponents in Prisoners' Dilemma: An Evolutionary Analysis

Bergh, Andreas

2005

Choosing Opponents in Prisoners' Dilemma: An Evolutionary Analysis*

Peter Engseld† and Andreas Bergh‡

Dept of Economics, Lund University, P.O. Box 7082, 220 07 Lund, Sweden

November 29, 2005

Abstract

We analyze a cooperation game in an evolutionary environment. Agents make noisy observations of opponents' propensity to cooperate, called reputation, and form preferences over opponents based on their reputation. A game takes place when two agents agree to play. Pareto optimal cooperation is evolutionarily stable when reputation perfectly reflects propensity to cooperate. With some reputation noise, there will be at least some cooperation. Individual concern for reputation results in a seemingly altruistic behavior. The degree of cooperation is decreasing in anonymity. If reputation is noisy enough, there is no cooperation in equilibrium.

JEL classification: C70; C72

Keywords: Cooperation; Conditioned Strategies; Prisoners Dilemma; Signaling; Reputation; Altruism; Evolutionary Equilibrium

1 Introduction

1.1 Background

The literature on the problem of cooperation is huge and spans several disciplines, see e.g. Gintis et al. (2005) and Hammerstein (2003). There is, however, still no consensus on how to explain both the emergence and deterioration of cooperation with unrelated strangers in finite interactions. A classical example of a cooperation game is the Prisoners' Dilemma, in which playing defect is a strictly dominant strategy. Nevertheless, both agents would be better off if

*We are grateful to Sergiu Hart, Frank Thuijsman, Hans Carlsson, and Håkan J. Holm for helpful comments.

†Corresponding author: peter.engseld@nek.lu.se.

‡The Ratio institute: andreas.bergh@ratio.se.

they could somehow commit themselves to play cooperate. Note that feasible commitments requires that agents are able to observe, directly or indirectly, the actions of the opponent, otherwise defection can not be retaliated.

The tit-for-tat strategy was described in Axelrod (1984) as retaliation mechanism against defections, and became widely known as the best strategy in repeated games of Prisoners' Dilemma. However, as pointed out in Boyd and Lorberbaum (1987), a population of cooperating tit-for-tats can be invaded by nice but less retaliatory strategies, resulting in a population vulnerable to invasion by defecting strategies. Thus, tit-for-tat is not an evolutionarily stable strategy. The main mechanism behind this result is that in standard game theoretical models, the agents are unable to choose with whom they are matched up. Instead, it is typically assumed that they are matched up with the same opponent or through random/tournament matching, see e.g. Kandori, Mailath, and Rob (1993). This convention stems not from descriptive accuracy, but rather from methodological considerations: Allowing other matching procedures would open countless possibilities.

Nevertheless, many interactions in real life are the results of individual choices, and not randomly imposed. Moreover, it is implausible to assume that individuals, given a choice, continue to interact with those who treat them unfavorably, see e.g. Tullock (1985). For this reason, it makes sense to analyze the Prisoners' Dilemma when agents have both some ability to observe the actions of others, and some possibilities to choose between potential opponents. Since the agents' payoff is strictly increasing in the opponents probability to play cooperate, all agents will seek to be matched with opponents who are more likely to play cooperate. This imposes a restriction on the matching possibilities: If you want to play with a cooperative agent, you have to play cooperative too.

1.2 Outline of the Model

In this paper, we analyze the Prisoners' Dilemma in an evolutionary environment, see Maynard Smith (1982), using a theoretical framework similar to that of Kandori, Mailath, and Rob (1993) and Young (1993). The main difference is that the probability of mutations in our model is given and assumed to be close to zero. We instead assume that agents make imperfect observations of opponents' propensity to play cooperate, interpreted as reputation, which enables them to form preferences over opponents based on their reputation. We introduce preference based matching which pairs agents with their most preferred feasible opponent.

The driving mechanism for our main results is the combination of observational skills and the ability to choose opponent. Under our assumptions, we show that if observational skills are perfect, i.e. reputation perfectly reflects each agent's propensity to cooperate, the payoff maximizing strategy in evolutionarily stable populations, is to cooperate and prefer to play with cooperative agents.

When observational skills are imperfect, so that reputation only imperfectly reflects past actions, any population can always be invaded by strategies with marginally higher degree of defection. This decreases the degree of cooperation in the population over time. However, if observational skills are accurate enough, a population with a sufficient degree of defection can be invaded by pure cooperators. In this case, the behavior in the population will change in a cyclical pattern. To capture this dynamic, we introduce a new equilibrium selection model which basically is a slightly modified *absorbing set*, and less restrictive than the conventional *evolutionarily stable strategies ESS*, see Maynard Smith (1982).

When observational skills are sufficiently inaccurate, the model yields the same equilibrium as standard models: there will be no cooperation in equilibrium. However, if agents are able to evolve such that observational skills can improve, the observational skills will endogenously, due to evolutionary pressure, improve over time.

1.3 Related Literature

The idea of conditioned actions in the Prisoners' Dilemma, is not new. Dawkins (1982) observed that if cooperative agents has an observable characteristic, such as a "green beard", agents with green beards will cooperate with each other and play defect with others. A similar idea, with a secret handshake was later formally modelled by Robson (1990). Frank (1988) considers the case when agents send different signals regarding whether they play cooperative or defect; cooperation is driven by an outside option which enables agents not to play. Grégoire and Robson (2003) show that when the population is divided into at least three subpopulations and imitation across subpopulations of the best strategy occurs, all equilibria involve cooperation. Using the Prisoners' Dilemma, Rob and Yang (2005) show that the ability to leave a defecting partner can induce long term cooperative relationships. Jackson and Watts (2005) introduce the term social games for games where agents chose not only strategies, but also with whom they play. In a non-evolutionary environment, they show that the threat of rematching can sustain new equilibria.

2 The model

Consider a population I with a large even number N agents who are repeatedly matched to play a symmetric 2×2 game below.

	D	C
D	$0,\ 0$	$1,\ \beta$
C	$\beta,\ 1$	$\alpha,\ \alpha$

$\Gamma(\alpha, \beta)$

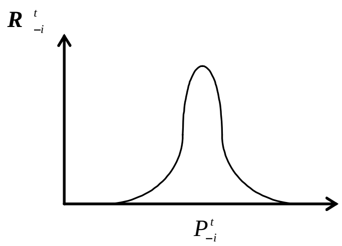
Let $\alpha \in (0,1)$ and $\beta \in (-\infty, 0)$. Prisoners' Dilemma is the special case when cooperation is socially optimal $(2\alpha > 1+\beta)$. The action set is $A \equiv \{D, C\}$, where $a \in A$. Actions are taken in discrete time, $t \in \{1, 2, 3, ...\}$. The actions D and C can be thought of as *Defect* and *Cooperate*.

The game $\Gamma(\alpha, \beta)$ is played repeatedly by the agents in the population. We use the term propensity as a measure of action history, formally defined as follows:

Definition 1 *The propensity* $P_i^t, \forall i \in I$ *at time* t *is a recursive function, where* $P_i^t \equiv \rho \Pr(a_i^t = C \mid \cdot) + (1-\rho) P_i^{t-1}, P_i^0 \in [0,1]$, *and* $\rho \in (0,1)$.

The propensity in period t is defined as a weighted average of the probability to play C in period t and the propensity in the previous period. This implies that agents with identical action history will have identical propensity.

Every agent $i \in I$ observes the *reputation* r_{-i}^t of an opponent $-i$, which is a realization of the stochastic variable $\mathbf{R}_{-i}^t$. $\mathbf{R}_{-i}^t$ is symmetrically, unimodally and smoothly distributed around P_{-i}^t, as depicted below. As a measure of the observational skills denoted $O_i^t \in \mathbb{R}_+$ we use the inverted standard deviation of $\mathbf{R}_{-i}^t$. The value of $r_{-i}^t \in \mathbb{R}$ is private information for i. Note that reputation is not limited to the unit interval. Henceforth we omit the time index when there is no risk for confusion.



Property 1 $\lim_{O_i \to \infty} r_{-i} = P_{-i}, \forall i, -i \in I.$

By being able to observe the reputation of opponents, agents can form preferences over all possible opponents. Let $\mathbf{\Psi}$ denote the set of all complete and transitive preference relations defined on $\mathbb{R}$. Let $\succsim \in \mathbf{\Psi}$ and let $\succsim_i$ denote the preferences of agent i.

Let $\mathbf{S}$ denote the set of all pure strategies. A pure strategy $s \in \mathbf{S}$ is a mapping from own propensity, and opponents' reputation onto A and $\mathbf{\Psi}$. More formally, we have $s : [0,1] \times \mathbb{R} \mapsto A \times \mathbf{\Psi}$. Less formally, a strategy assigns an action and a preference order over every feasible population. Note that this mapping allows agents to condition their actions on the reputation of their opponents.

A mixed strategy is denoted σ and is defined as a probability distribution over $\mathbf{S}$. Formally, $\sigma \equiv (\sigma_s)_{s \in \mathbf{S}}$, $\sigma_s \in [0,1]\,, \forall s \in \mathbf{S}$ and $\int_{s \in \mathbf{S}} \sigma_s = 1$. Any mixed strategy can consequently be seen as a vector $\sigma \in \mathbb{R}_+^\infty$, that belongs to the unit simplex Δ_σ, where

$$\Delta_\sigma \equiv \left\{ \sigma \in \mathbb{R}_+^\infty \mid \int_{s \in \mathbf{S}} \sigma_s = 1 \right\}.$$

A combination of strategies in the population is denoted Q_I and defined as a probability distribution over Δ_σ. Let q_σ denote the fraction of agents in I with strategy σ. Formally, $Q_I = (q_\sigma)_{\sigma \in \Delta_\sigma}$, $q_\sigma \in [0,1]\,, \forall \sigma \in \Delta_\sigma$ and $\int_{\sigma \in \Delta_\sigma} q_\sigma = 1$. Any combination of mixed strategies in the population can be seen as a vector $Q_I \in \mathbb{R}_+^\infty$, that belongs to the unit simplex Δ_Q, where

$$\Delta_Q \equiv \left\{ Q_I \in \mathbb{R}_+^\infty \mid \int_{\sigma \in \Delta_\sigma} q_\sigma = 1 \right\}.$$

Note that Q_I both can be viewed as a point in $\mathbb{R}_+^\infty$ and as a set of strategies. We use $Q_{I \setminus i}$ to denote the strategy mix in the population $I \setminus i$, and we let $O_{I \setminus i}$ denote the observational skills of all agents $I \setminus i$. The expected payoff for agent i with strategy σ_i at period t will be $\pi^t \left(\sigma_i, O_i; Q_{I \setminus i}, O_{I \setminus i} \right)$. When there is no risk of confusion, we use Q to denote Q_I.

2.1 Evolutionary Stability

To analyze how the population evolves we apply an evolutionary setting similar to e.g. Kandori, Mailath, and Rob (1993) and Young (1993). We impose *perturbations* such that every agent in the population in each period with a small given probability will "mutate", meaning that they change strategy. Just as in the papers cited above, one or more agents can change strategy in each time period and all mutations have equal probability.

The perturbations can be divided into three subgroups: successful mutations, unsuccessful mutations and evolutionary drift. A successful mutation implies that the change of strategy yields a strictly higher payoff, whereas an unsuccessful mutation yields a strictly lower payoff. Evolutionary drift occurs when the change of strategy yields the same payoff, see e.g. Binmore and Samuelson (1999).

The *growth* in the population is such that strategies with higher payoffs will have a higher representation in the population in the next period:[1]

$$\text{sign}\left(\frac{q_{\sigma}^{t+1}}{q_{\sigma}^{t}} - \frac{q_{\sigma'}^{t+1}}{q_{\sigma'}^{t}}\right) = \text{sign}\left(\pi^{t}\left(\sigma, \cdot; \cdot\right) - \pi^{t}\left(\sigma', \cdot; \cdot\right)\right). \tag{1}$$

Offsprings are assumed to inherit both strategy, propensity and observational skill from the parent. The question whether a mutant strategy could invade the current (incumbent) strategy distribution is not as straightforward as in standard models. As usual, the payoff of an agent i depends both on the agent's strategy σ_i and on the opponent's strategy σ_{-i}. However, the opponent's actions can also depend on the agent's reputation, just as the agent's action can depend on the opponent's reputation. This implies that if an agent changes strategy, her actions and thus her propensity can change, which could trigger different actions from other agents and thereby change their propensity, which in turn might lead to other agents changing their actions ad infinitum.

An adiabatic relationship between the processes help us avoid such cumbersome dynamic:

Assumption 1 *The distribution of propensity in the population converges to a limit state before the growth begins, and the growth converges to a limit state before new perturbations.*

[1] The growth in this model is identical to that in Kandori, Mailath, and Rob (1993) and Young (1993).

Consequently, the adjustment process of the propensity is much faster than the growth process, which in turn is much faster than the process of perturbations. This implies that the population on average can be considered stationary in so far as the pair σ_i, P_i is fixed $\forall i \in I$. This renders the index for time redundant in most cases. Another consequence of Assumption 1 is that $\forall Q \in \Delta_Q$, there exists a corresponding propensity distribution.

A population can evolve from Q to Q' either through growth, successful mutations or evolutionary drift.

Definition 2 $\overrightarrow{QQ'}$ *denotes a path connected curve in* Δ_Q *between* Q *and* Q'*, implying that* Q *can evolve to* Q' *through growth or through perturbations.*

Due to the potential existence of oscillating strategy mixes, standard equilibrium concepts, such as *ESS*, are too restrictive. Let us therefore define a *mutation proof attraction set* (*MAS*), which basically is a slightly modified *absorbing set*, see e.g. Samuelson (1998), where the set is closed under the growth mechanism and mutations, whereas absorbing sets are closed only under the growth mechanism.

Definition 3 (*MAS*) $\mathbf{Q}^{MAS}(\Gamma)$ *is a set of strategy mixes* $Q \in \mathbf{Q}^{MAS}(\Gamma)$ *where*

- $\exists\ \overrightarrow{QQ'},\ \forall Q, Q' \in \mathbf{Q}^{MAS}(\Gamma)$, *and*
- $\nexists\ \overrightarrow{QQ''}$ *for any* $Q'' \notin \mathbf{Q}^{MAS}(\Gamma)$.

Let $\Delta^{MAS}(\Gamma) \equiv \bigcup \mathbf{Q}^{MAS}(\Gamma)$.

Property 2 $\Delta^{MAS}(\Gamma) \neq \emptyset, \forall \Gamma$.

A population I belongs to a *MAS*, precisely if the strategy mix Q in the population belongs to an attraction set $\mathbf{Q}^{MAS}(\Gamma)$ such that $\exists\ \overrightarrow{QQ'},\ \forall Q, Q' \in \mathbf{Q}^{MAS}(\Gamma)$. That is, each combination of strategies in the population that belongs to the attraction set $\mathbf{Q}^{MAS}(\Gamma)$ must be able to evolve to any other point in the attraction set, either through growth or through evolutionary drift. Moreover, there must not exist any feasible path such that the population could evolve to a point $Q'' \notin \mathbf{Q}^{MAS}(\Gamma)$. Note that *MAS* yields identical equilibria on unconditioned strategies as neutrally stable strategies *NSS*, see Maynard Smith (1982).

2.2 Matching of the Agents

The individual preference ordering $\succsim_i \in \mathbf{\Psi}$ enables each agent to make pairwise comparisons of all other agents in the population, such that $k \succsim_i j$ implies that agent i weakly prefers agent k over agent j, whereas $k \succ_i j$ implies that agent i has a strict preference for agent k over agent j.

Let $\mathbb{I}$ denote the set of matched pairs. Formally, *preference based matching* is described as follows:

Definition 4 (*Preference based matching*)

$$\nexists (i,j), (k,l) \in \mathbb{I} \text{ such that } k \succ_i j, \text{ and } i \succ_k l.$$

Preference based matching implies that agents are matched up with their most preferred feasible opponent. Many matching procedures may satisfy the conditions above, for an example see Appendix B. Preference based matching procedures do not generate a deterministic set of matched agents. In order to make the matching procedure path independent, the games are evaluated through the expected payoffs given a fixed set of preferences and observational skills in the population. For technical reasons, we assume that each strategy present in the population is utilized by an even number of agents. By this assumption we avoid the pathological case when a non-preferred opponent imposed on an odd agent with a given strategy, possibly decreases the expected payoff for agents this strategy.

Note that when observational skills are non-existent, preference based matching is equivalent to random matching.

3 Evaluating the Game

Denote the strategy mixes where all choose action C and D respectively:

$$Q^C \equiv \{Q \mid \Pr(a_i = C) = 1, \forall i \in I\}, \text{ and } Q^D \equiv \{Q \mid \Pr(a_i = C) = 0, \forall i \in I\}.$$

Regardless of an agent's strategy, the payoff is always higher if the opponent is more likely to play C. Since changes in strategies are assumed to be rare, the adiabatic relationship propensity and actions (Assumption 1) implies that the expected payoff π is strictly increasing in the opponent's propensity.

Property 3 $\frac{\partial \pi_i}{\partial P_{-i}} > 0$.

From the definition of reputation we know that the expected value of the reputation equals the propensity.

Property 4 $\frac{\partial \pi_i}{\partial r_{-i}} > 0$.

Denote by $\succsim^{\mathcal{C}}$ the set of preferences such that the agent prefers opponents with higher reputation:

Definition 5 $\succsim^{\mathcal{C}} \equiv \{\succsim \mid r_j \geq r_k \Leftrightarrow j \succsim k, \forall r \in \mathbb{R}\}$.

3.1 Perfect Observational Skills

Let us begin with the special case when reputation is identical to the propensity, i.e. $O_i = \infty, \forall i \in I$. From property 3 we know that the payoff is strictly increasing in the opponent's propensity. For this reason, assume for now that all agents have preferences $\succsim^{\mathcal{C}}$.

Since all agents will be able to avoid being matched up with opponents of lower propensity, agents will only be matched up with opponents of identical propensity.

Let $z_i \equiv \Pr(a_i = C)$. Since agents matched with each other will have identical propensity, let $z \equiv z_i$. Focus now on how the payoff depends on the propensity. The payoff for an arbitrary agent i is given by $\pi_i(\cdot) = \alpha z^2 + z(1-z)(1+\beta)$, which is maximized for

$$z = \frac{1}{2} + \frac{1}{2}\frac{\alpha}{1+\beta-\alpha}.$$

When cooperation is socially optimal, $2\alpha \geq 1+\beta$, this implies $z = 1$. If $2\alpha < 1+\beta$, the payoff is maximized for mixed strategies where the probability of playing C is equal to $\frac{1}{2}+\frac{1}{2}\frac{\alpha}{1+\beta-\alpha} > \frac{1}{2}$. Thus, the probability of playing action C is strictly increasing in α, and always higher than 50 percent. However, when all agents have identical propensity, the population will be vulnerable to a neutral invasion of agents with the same propensity as the incumbents, but with $\succsim \neq \succsim^{\mathcal{C}}$.

Lemma 1 $P_i = P_j, \forall i, j \in I \Rightarrow \pi\left(\sigma \mid \succsim^{\mathcal{C}}\right) = \pi\left(\sigma \mid \succsim \neq \succsim^{\mathcal{C}}\right), \forall \sigma \in \Delta_\sigma$.

In other words, the population will through mutations drift away from all agents preferring opponents with higher reputation.

Proposition 1 *If* $Q \in \Delta^{MAS}(\Gamma)$ *and* $O_i = \infty, \forall i \in I$ *then*

- $2\alpha \geq 1+\beta \Rightarrow \lim_{N\to\infty} \Pr(a_i = C) = 1$, *and*
- $2\alpha < 1+\beta \Rightarrow \lim_{N\to\infty} \Pr(a_i = C) = \frac{1}{2} + \frac{1}{2}\frac{\alpha}{1+\beta-\alpha}, \forall i \in I.$

The intuition is as follows: Since agents with $\succsim \neq \succsim^{\mathcal{C}}$ can only make a neutral invasion, they will, for a given growth and perturbation speed, represent a small subset of I of fixed size. This fraction can be exploited by strategies more inclined to play D, with preferences $\succsim^{\mathcal{C}}$. These strategies will initially yield more than all other strategies in the population. However, strategies with $\succsim \neq \succsim^{\mathcal{C}}$ will yield less than all other strategies and therefore grow slower. This implies that agents with low propensity strategies to a higher degree will become matched up themselves, and thus earn a lower payoff. This process will eventually stabilize when the expected payoff for agents with $\succsim \neq \succsim^{\mathcal{C}}$ equals that of agents with low propensity strategies.

Note that when the exploiting agents, due to their lower payoff, eventually disappear from the strategy mix, the population will again drift away from all agents having preferences $\succsim^{\mathcal{C}}$, and the process described above will start over, causing a rare reoccurring limited cyclical movement in Q.

3.2 Imperfect Observational Skills

Assume now that reputation is noisy and only imperfectly reflects propensity.

Lemma 2 *$O_i < \infty, \forall i \in I$, and $\exists P_i \neq P_j$, for some $i, j \in I \Rightarrow \pi\left(\sigma \mid \succsim^{\mathcal{C}}\right) > \pi\left(\sigma \mid \succsim \neq \succsim^{\mathcal{C}}\right), \forall \sigma \in \Delta_\sigma$.*

When the population contains agents with different propensity, preferences $\succsim^{\mathcal{C}}$ will yield higher payoff. Moreover, if the population contains agents with different propensity, the payoff is strictly increasing in observational skills for agents with $\succsim^{\mathcal{C}}$.

Lemma 3 *$O_i < \infty, \forall i \in I$, and $\exists P_i \neq P_j$, for some $i, j \in I \Rightarrow \frac{\partial \pi_i\left(\sigma \mid \succsim^{\mathcal{C}}\right)}{\partial O_i} > 0, \forall i \in I$ and $\forall \sigma \in \Delta_\sigma$.*

Corollary 1 *$\exists P_i \neq P_j$, for some $i, j \in I$ and $Q \in \Delta^{MAS}(\Gamma) \Rightarrow O_i = O, \forall i \in I$.*

If agents are able to evolve such that observational skills can improve, the observational skills will improve over time.

Corollary 2 *$\exists P_i \neq P_j$, for some $i, j \in I$ and $Q \in \Delta^{MAS}(\Gamma) \Rightarrow \lim_{t \to \infty} O^t = \infty$.*

As a consequence of Corollary 1, we henceforth analyze the game under the assumption that all agents have the same observational skills: $O_i = O, \forall i \in I$.

Now consider two types of agents with strategies σ_1 and σ_2, and corresponding propensities $P_1 > P_2$. Define $\pi_1 \equiv \pi(\sigma_1)$ and $\pi_2 \equiv \pi(\sigma_2)$. Let π_{11} denote the payoff for type 1 agent when matched against another type 1 agent. Let π_{12} denote the payoff for a type 1 agent when matched against type 2. Analogously, π_{22} denotes the payoff when two type 2 agents meet, and π_{21} is the payoff for type 2 agents when matched against type 1.

ρ_1 denotes the fraction of type 1 agents who meet type 2, i.e. matching failures for type 1. Analogously, ρ_2 denotes the fraction of type 2 agents who meet type 1. Let N_1 and N_2 be the number of agents of type 1 and 2, and note that $\rho_1 N_1 = \rho_2 N_2$.

The payoffs can be described as follows:

$$\pi_1 = (1-\rho_1)\,\pi_{11} + \rho_1 \pi_{12} \quad \text{and} \quad \pi_2 = (1-\rho_2)\,\pi_{22} + \rho_2 \pi_{21},$$

As before, z_1 denotes the probability to play C for a type 1 agent with propensity P_1. Let $z_2 = z_1 - x$ denote the corresponding probability for a type 2 agent, where $x \in (0, z_1]$. Hence, the relevant payoffs can be written:

$$\pi_{11} = z_1^2\alpha + z_1(1-z_1)(1+\beta), \tag{2}$$

$$\pi_{12} = z_1(z_1-x)\,\alpha + (z_1-x)(1-z_1) + z_1(1-z_1+x)\,\beta, \tag{3}$$

$$\pi_{22} = (z_1-x)^2\,\alpha + (z_1-x)(1-z_1+x)(1+\beta), \tag{4}$$

$$\pi_{21} = z_1(z_1-x)\,\alpha + (z_1-x)(1-z_1)\,\beta + z_1(1-z_1+x). \tag{5}$$

Since $\pi_{11} > \pi_{12}$ and $\pi_{22} < \pi_{21}$ it follows that if matching failures for type 1 agents are sufficiently common, the less cooperative type 2 agents will earn more. Let $\rho_1^{\max}$ be the value of ρ_1 for which $\pi_1 = \pi_2$. In other words, $\rho_1^{\max}$ is the maximum fraction of matching failures allowed for type 1 agents in order to prevent type 2 agents from earning a higher payoff and thereby successfully invade the population.

Lemma 4 $\exists P_2 < P_1$ *such that* $\pi_2 > \pi_1, \forall P_1 \in (0,1]$ *when* $O < \infty$.

The maximum allowed fraction of matching failures for agents of type 1 in a close proximity of P_2 corresponds to less mistakes than random matching, i.e. $\rho_1^{\max} < \frac{N_2}{N_2+N_1}$. From the assumption about the noise it follows that the ability to identify whether an agents is type 1 or 2 when $|P_1 - P_2| \approx 0$ is close to non-existent.

Consequently, unless observational skills are perfect, any population can always be invaded by agents less prone to play cooperative. The intuition behind this result is that the observational skills needed to prevent invasion by more defecting agents requires fewer errors than random matching, which is impossible when the invasion occurs arbitrarily close to the propensity of the incumbents.

Lemma 4 suggests a dynamic that eventually will drive the population towards a state where all agents play defect. Nevertheless, from the definition of noise it follows that $\succsim_i = \succsim^{\mathcal{C}}, \forall i \in I \Rightarrow \frac{\partial \rho_1}{\partial x} < 0$. That is, matching failures for type 1 agents are decreasing in x, and thus also in the propensity distance between type 1 and 2. Less formally, a slightly less cooperative opponent is harder to recognize than an opponent with much lower propensity, and thereby also harder to avoid being matched up with.

Lemma 5 *$\exists O < \infty, z_1 \in [0,1]$ and $x \in (0, z_1]$ such that $\rho_1^{\max} > \rho_1 > 0$.*

Corollary 3 *$\exists O < \infty$ and $x \in (0, z_1]$ such that $\pi_1 > \pi_2$.*

That is, there exists an imperfect observational skill which enables a more cooperative strategy to invade a population of less cooperative players. Let O^* denote the minimum observational skill with which Lemma 5 is satisfied.

Corollary 4 *$O \geq O^* \Rightarrow \exists x \in (0, z_1]$ such that $\pi_1 > \pi_2$.*

If the observational skill in the population is better than O^*, then type 1 agents can successfully invade a population with much less cooperative type 2 agents. But as shown in Lemma 4, this more cooperative population can in turn be invaded by slightly less cooperative strategies. The process described above will start over, causing a cyclical movement in Q such that the population will oscillate between different degrees of cooperation.

Unless the observational skill in the population is higher than O^*, complete defection is a unique *MAS*.

Proposition 2 *$O < O^* \Rightarrow Q^{\mathrm{D}} = \Delta^{MAS}(\Gamma)$.*

Nevertheless, given that observational skills can evolve, note that the population will always be subjected to perturbations. Hence, agents with $P > 0$ will rarely, but repeatedly emerge in the population, i.e. $\exists P_i \neq P_j$, for some $i, j \in I$. From Lemma 3 it follows that $\frac{\partial \pi_i(\sigma|\succsim^{\mathcal{C}})}{\partial O_i} > 0$. Hence, eventually will $O \geq O^*$. Then from Lemma 4 and 5 it follows that $\exists P_i \neq P_j$, for some $i, j \in I$.

Lemma 6 $\succsim_i = \succsim^{\mathcal{C}}, \forall i \in I \Rightarrow \frac{\partial \rho_1}{\partial O} < 0$.

The fraction of matching failures for type 1 agents is decreasing in observational skills. Hence, the possibility of successful invasion by type 1 agents will increase as observational skills improve.

Proposition 3 *The degree of cooperation in the population is strictly increasing in observational skills when $O \geq O^*$.*

From Proposition 2 and 3 it follows:

Corollary 5 *The degree of cooperation in the population is weakly increasing in observational skills.*

3.3 Summary

The results can now be summarized as follows:

1. If observational skills are perfect, i.e. $O_i = \infty, \forall i \in I$, we have two cases:
   (a) When cooperation is socially optimal, $2\alpha \geq 1 + \beta$, almost total cooperation is a unique *MAS*.
   (b) When cooperation is inefficient, $2\alpha < 1 + \beta$, more than half of the actions in the *MAS* will be cooperative.
2. If observational skills are imperfect, but sufficiently good, i.e. $O \geq O^*$, the strategy mix in the population will oscillate between different degrees of cooperation. The degree of cooperation in the population is strictly increasing in observational skills.
3. If observational skills are poor enough, i.e. $O < O^*$, complete defection is a unique *MAS*, i.e. $Q^D = \Delta^{MAS}(\Gamma)$.

4. If agents are able to evolve such that observational skills can improve, the observational skills will improve over time.

4 Conclusions and Remarks

In our model individual concern for reputation results in a seemingly altruistic behavior. We have thus shown that prosocial behavior, such as cooperation in Prisoners' Dilemma can be explained without resorting to models with altruism or inequity aversion, see e.g. Fehr and Fischbacher (2003) and Fehr and Schmidt (1999). Reputation based choice can also potentially explain the big impact of the degree of anonymity on behavior. When reputation does not perfectly reflect behavior, there are situations where the payoff associated with defections will outweigh the reputational costs. For experimental evidence, see e.g. Hoffman, McCabe, and Smith (1996).

Regarding the effect of reputation based choice of opponent, there is less experimental evidence available. McCabe, Rigdon, and Smith (2003) pair participants in a trust game based on their degree of trust and trustworthiness, which allows cooperation to emerge and protects cooperation from being invaded by defecting players.

This supports the idea that an important key to understanding cooperation in repeated games is the matching procedure. Random/tournament matching represents one extreme, whereas reputation based choice as analyzed in this paper represents another. In practice, people encounter some situations where they are able to choose their opponent in strategic interactions and some situations where they are forced to play games of cooperation against random agents in the population. The implications of such mixed matching procedures deserve to be examined closely. The results are likely to be positive for cooperation: As long as there is at least some degree of free opponent choice, agents must take into consideration the reputational consequences of their actions also when they play against randomly assigned opponents.

A Proof

Proof of Lemma 1. $P_i = P_j, \forall i, j \in I \Rightarrow z_i = z_j, \forall i, j \in I$. $z_i = z_j, \forall i, j \in I$ implies that all opponents will yield the same payoff. Consequently, $\pi\left(\sigma \mid \succsim'\right) = \pi\left(\sigma \mid \succsim''\right), \forall \sigma \in \Delta_\sigma$ and $\forall \succsim', \succsim'' \in \mathbf{\Psi}$. ■

Proof of Proposition 1. From Property 3 we know that $\frac{\partial \pi_i}{\partial P_{-i}} > 0$. When observational skills are perfect, we have that $P_{-i} = r_{-i}$ which implies that all agents can avoid being matched up with opponents with lower propensity. We know that each agent i maximizes her payoff when $z_i = 1$ if $2\alpha \geq 1 + \beta$, and $z_i = \frac{1}{2} + \frac{1}{2}\frac{\alpha}{1+\beta-\alpha}$ if $2\alpha < 1 + \beta$. Let Q^* denote the set of strategy mixes where all agents, for given α and β, use payoff maximizing strategies. This results in a constant $P_i = P^*, \forall i \in I$. Denote the incumbent strategy σ^* with preferences $\succsim^{\mathcal{C}}$ yielding the propensity P^*. From Lemma 1 we know that the incumbent population can be neutrally invaded by a strategy σ' with $P' = P^*$ but with $\succsim \neq \succsim^{\mathcal{C}}$. This in turn makes the population vulnerable to invasion by a strategy σ'' with $P'' < P^*$ and with $\succsim^{\mathcal{C}}$. This will according to equation 1 result in

$$\pi\left(\sigma''\right) \geq \pi\left(\sigma^*\right) > \pi\left(\sigma'\right) \Rightarrow \frac{q_{\sigma''}^{t+1}}{q_{\sigma''}^{t}} \geq \frac{q_{\sigma^*}^{t+1}}{q_{\sigma^*}^{t}} > \frac{q_{\sigma'}^{t+1}}{q_{\sigma'}^{t}}$$

However, as the fraction $q_{\sigma'}$ decreases relative to $q_{\sigma''}$, σ'' will gradually to a higher degree become matched up with other σ'', since σ^* will never be matched up with σ''. As a consequence $\pi\left(\sigma''\right)$ will decrease as $q_{\sigma'}$ decreases. $\pi\left(\sigma''\right)$ will continue to decrease until eventually

$$\pi\left(\sigma^*\right) > \pi\left(\sigma''\right) \Rightarrow \frac{q_{\sigma^*}^{t+1}}{q_{\sigma^*}^{t}} > \frac{q_{\sigma''}^{t+1}}{q_{\sigma''}^{t}}, \text{ and } \pi\left(\sigma^*\right) > \pi\left(\sigma'\right) \Rightarrow \frac{q_{\sigma^*}^{t+1}}{q_{\sigma^*}^{t}} > \frac{q_{\sigma'}^{t+1}}{q_{\sigma'}^{t}} \quad .$$

No more invasions are possible as long as $q_{\sigma''} > 0$.

Now consider the neutral invasion by σ'. Since $\pi\left(\sigma^*\right) = \pi\left(\sigma'\right) \Rightarrow \frac{q_{\sigma^*}^{t+1}}{q_{\sigma^*}^{t}} = \frac{q_{\sigma'}^{t+1}}{q_{\sigma'}^{t}}$. Hence, according to equation 1 the fraction $q_{\sigma'}$ is constant. Given Assumption 1, this implies that $q_{\sigma'}$ will be decreasing in population size N. Also note that $q_{\sigma''}$ is bounded by $q_{\sigma'}$. Hence there exist a bounded neighborhood around Q^* which is decreasing in N, such that the population in a *MAS* will converge to:

$$\lim_{N\to\infty} \Pr(a_i = C) = 1, \forall i \in I \text{ if } 2\alpha \geq 1+\beta, \text{ and}$$
$$\lim_{N\to\infty} \Pr(a_i = C) = \frac{1}{2} + \frac{1}{2}\frac{\alpha}{1+\beta-\alpha}, \forall i \in I \text{ if } 2\alpha < 1+\beta.$$

■

Proof of Lemma 2. From Property 3 and 4 we know that $\frac{\partial \pi_i}{\partial P_{-i}} > 0$ and $\frac{\partial \pi_i}{\partial r_{-i}} > 0$. If $\exists P_i \neq P_j$, for some $i, j \in I$, then $\succsim^{\mathcal{C}}$ will result in a lower probability of being matched up with low propensity opponents, and a higher probability of being matched up with high propensity opponents, than any $\succsim \neq \succsim^{\mathcal{C}}$. ■

Proof of Lemma 3. Consider a population with a fixed distribution of observational skills and focus an agent i with observational skill O_i. The probability that agent i mistakenly perceives an opponent to be more cooperative than she really is, is clearly decreasing in observational skill. Analogously, the probability that agent i correctly identifies an opponent as having a higher propensity is increasing in observational skills. Consequently, for agents with $\succsim^{\mathcal{C}}$, a better observational skill leads to a higher probability that low propensity agents are ranked low and high propensity agents are ranked high, which in turn results in a higher probability to become matched up with a high propensity agent. The Lemma follows directly from Property 4. ■

Proof of Lemma 4. Let $O_i = O, \forall i \in I$. Consider the difference $\pi_1 - \pi_2$, and assume that $\pi_1 - \pi_2 = 0$. Hence,

$$\pi_1 - \pi_2 = (1-\rho_1^{\max})\,\pi_{11} + \rho_1^{\max}\pi_{12} - \left(1 - \rho_1^{\max}\frac{N_1}{N_2}\right)\pi_{22} - \rho_1^{\max}\frac{N_1}{N_2}\pi_{21} = 0.$$

Solving for $\rho_1^{\max}$ yields:

$$\rho_1^{\max} = \frac{N_2\,(\pi_{22} - \pi_{11})}{N_2\,(\pi_{12} - \pi_{11}) + N_1\,(\pi_{22} - \pi_{21})}. \tag{6}$$

Substituting the payoffs from equations 2 to 5 into equation 6 yields:

$$\rho_1^{\max} = \frac{N_2\,((2z_1 - x)(\alpha - \beta - 1) + 1 + \beta)}{(N_2 + N_1)(z_1(\alpha - \beta - 1) + 1) - xN_1(\alpha - \beta - 1)}$$

Note that ρ_1 denotes the actual fraction of mistakes for the incumbents. It follows that $\rho_1^{\max} < \rho_1 \Rightarrow \pi_1 < \pi_2$ and $\rho_1^{\max} > \rho_1 \Rightarrow \pi_1 > \pi_2$.

Note that having no observational skills is equivalent to random matching where $\rho_1 = \frac{N_2}{N_2+N_1}$. From the definition of reputation, it follows that the fraction of matching failures converges to random matching as $x \to 0$. Consequently, we have $\lim\limits_{x\to 0} \rho_1 = \frac{N_2}{N_2+N_1}$.

For an invasion arbitrariliy close to the incumbents, we have

$$\lim_{x\to 0} \rho_1^{\max} = \tfrac{N_2(2z_1(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)}.$$

Consider the difference $\lim\limits_{x\to 0} \rho_1^{\max} - \lim\limits_{x\to 0} \rho_1$
$= \frac{N_2(2z_1(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)} - \frac{N_2}{N_2+N_1} = \frac{N_2}{N_2+N_1}\left(\frac{2z_1(\alpha-\beta-1)+1+\beta}{z_1(\alpha-\beta-1)+1} - 1\right)$.
Since $\frac{N_2}{N_2+N_1} > 0$, we have $\lim\limits_{x\to 0} \rho_1^{\max} - \lim\limits_{x\to 0} \rho_1 < 0$
$\Leftrightarrow \frac{2z_1(\alpha-\beta-1)+1+\beta}{z_1(\alpha-\beta-1)+1} - 1 < 0 \Leftrightarrow z_1(1+\beta-\alpha) > \beta$.
Two cases: $\begin{cases} 1+\beta-\alpha < 0 & \Rightarrow \quad z_1 < 1 < \frac{\beta}{1+\beta-\alpha} \\ 1+\beta-\alpha > 0 & \Rightarrow \quad z_1 > 0 > \frac{\beta}{1+\beta-\alpha} \end{cases}$.
Since $z_1 \in [0,1]$, it follows that $\frac{N_2(2z_1(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)} < \frac{N_2}{N_2+N_1}$.
Then $\exists\widetilde{\varepsilon} > 0$ such that:

$$\tfrac{N_2(2z_1(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)} + \widetilde{\varepsilon} < \tfrac{N_2}{N_2+N_1}.$$

Then there also $\exists x > 0$ such that $\frac{N_2((2z_1-x)(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)-xN_1(\alpha-\beta-1)} < \frac{N_2}{N_2+N_1}$.
Consequently, $\exists P_2 < P_1$ such that $\pi_2 > \pi_1$ or more explicitly:

$\pi(\sigma_2, O) > \pi(\sigma_1, O), \forall O < \infty$. ■

Proof of Lemma 5. From the proof of Lemma 4 we know that

$$\rho_1^{\max} = \tfrac{N_2((2z_1-x)(\alpha-\beta-1)+1+\beta)}{(N_2+N_1)(z_1(\alpha-\beta-1)+1)-xN_1(\alpha-\beta-1)}. \tag{7}$$

Remember that: $\rho_1^{\max} > \rho_1 \Rightarrow \pi_1 > \pi_2$. Consequently, $\exists\rho_1^{\max} > 0 \Rightarrow \exists\rho_1 > 0$ such that $\pi_1 - \pi_2 > 0$. Consider the denominator in equation 7:

$$\begin{aligned}(N_2+N_1)(z_1(\alpha-\beta-1)+1) - xN_1(\alpha-\beta-1) &> \\ (z_1-x)(\alpha-\beta-1)+1 &> 0.\end{aligned}$$

Now consider the nominator, for $x = z_1 = 1$:

$$N_2 \left((2z_1 - x) (\alpha - \beta - 1) + 1 + \beta \right) = N_2 \alpha > 0.$$

Consequently, $\exists \rho_1^{\max} > 0$. Then $\exists \rho_1^{\max} > \rho_1 \Rightarrow \pi_1 > \pi_2$. ■

Proof of Proposition 2. Follows from Lemma 5 and Corollary 4. ■

Proof of Lemma 6. Consider a population I with agents of type 1 and 2, and propensities $P_1 > P_2$. Let ρ_1 denote the probability for a type 1 agent to become matched up with a type 2 agent. Let $\rho_{i,j}$ denote the probability that agent i of type 1 is going to be matched with agent j of type 2.

Let $p_{i,j}$ denote the combined probability that an agent i of type 1 mistakenly ranks an agent j of type 2 higher than a type 1 agent, and that agent j correctly ranks agent i higher than a type 2 agent.

Note that $\rho_{i,j}$ is increasing in $p_{i,j}$, $\forall i, j \in I$, i.e. $\frac{\partial \rho_{i,j}}{\partial p_{i,j}} > 0, \forall i, j \in I$.

Let $\widetilde{p}$ denote the probability that an agent i of type 1 perceives an agent of type 2 to be more cooperative than herself, $\widetilde{p} \equiv \Pr(r_2 > P_1)$. Since both types have identical observational skills, the probability that an agent j of type 2 perceives an agent of type 1 to be more cooperative than herself, is $1 - \widetilde{p} = \Pr(r_1 > P_2)$. Note that $\frac{\partial \widetilde{p}}{\partial O} < 0$.

$p_{i,j}$ is increasing in the combined probability $(\widetilde{p})(1 - \widetilde{p}) \equiv \widetilde{p}_{i,j}$, i.e. $\frac{\partial p_{i,j}}{\partial \widetilde{p}_{i,j}} > 0$.
From the definition of reputation it follows that $\widetilde{p} < \frac{1}{2}$, hence $\frac{\partial \widetilde{p}_{i,j}}{\partial \widetilde{p}} > 0$.

Consequently $\frac{\partial \rho_{i,j}}{\partial \widetilde{p}} = \frac{\partial \rho_{i,j}}{\partial p_{i,j}} \frac{\partial p_{i,j}}{\partial \widetilde{p}_{i,j}} \frac{\partial \widetilde{p}_{i,j}}{\partial \widetilde{p}} > 0$. Moreover, $\frac{\partial \rho_{i,j}}{\partial O} = \frac{\partial \rho_{i,j}}{\partial \widetilde{p}} \frac{\partial \widetilde{p}}{\partial O} < 0$. That is, the probability that agent i of type 1 is going to be matched with agent j of type 2 is decreasing in observational skill.

Since the probability that agent i of type 1 is going to be matched with agent j of type 2 is decreasing in observational skill for every pair in the population, we have that $\operatorname{sign}\left(\frac{\partial \rho_{i,j}}{\partial O}\right) = \operatorname{sign}\left(\frac{\partial \rho_1}{\partial O}\right)$. ■

Proof of Proposition 3. From Lemma 6 we know that $\frac{\partial \rho_1}{\partial O} < 0$. Since $\pi_{11} > \pi_{12} \Rightarrow \frac{\partial \pi_1}{\partial \rho_1} < 0$ and $\pi_{22} < \pi_{21} \Rightarrow \frac{\partial \pi_2}{\partial \rho_1} > 0$, it follows that a decrease in ρ_1

benefits more cooperative agents. Hence, fewer mistakes will make it harder for less cooperative agents to exploit cooperative agents. As observational skills improve, the population will converge towards the degree of cooperation described by Proposition 1. ■

B Example of a Matching Procedure

First, for each individual preference ordering $\succsim_i \in \mathbf{\Psi}$ there exists at least one corresponding vector $R_i \equiv \left(R_i^1, R_i^2, ..., R_i^N\right)$ where R_i^k denotes agent i's k-preferred choice. Thus, R_i^1 denotes i's most preferred opponent, R_i^2 her second best, and so on.

This procedure makes use of a randomized choosing order, assumed (without loss of generality) to coincide with the numbers 1 to N. First, agent 1 asks her most preferred opponent, who accepts if agent 1 is her most preferred opponent. Then agent 2 asks her most preferred opponent, and when all agents have proposed to their first best choice, the procedure is repeated for second best choices. The procedure continues until all agents are paired. Formally, the matching procedure can be described by the following algorithm, which pairs all agents in I into $\mathbb{I}$.

Algorithm 1 (Matching procedure) *Let* $\mathbb{I}$ *be the set of matched pairs.*

Step 0. Let $\mathbb{I} = \emptyset$, $i = 1$, *and* $l = 1$.

Step 1. If there exists an $m \in [1, l]$ *such that if* $\left(R_i^l = j\right) \wedge \left(R_j^m = i\right) \wedge (i, j \notin \mathbb{I})$, *then* $(i, j) \in \mathbb{I}$.

Step 2. Increase i *by* 1. *If* $i \leq N$, *go to step 1.*

Step 3. Increase l *by* 1 *and let* $i = 1$. *If* $l \leq N$, *go to step 1.*

To ensure that the realized payoff for every agent at each period is equal to the expected payoff, the matching procedure is assumed to be repeated an infinite number of times within each period.

References

Axelrod, R. (1984): *The Evolution of Cooperation*. Basic Books, New York.

Binmore, K., and L. Samuelson (1999): "Evolutionary Drift and Equilibrium Selection," *Review of Economic Studies*, 66, 363–393.

Boyd, R., and J. Lorberbaum (1987): "No Pure Strategy is Evolutionarily Stable in the Repeated Prisoner's Dilemma Game," *Nature*, 327, 58–59.

Dawkins, R. (1982): *The Extended Phenotype*. Oxford University Press, Oxford.

Fehr, E., and U. Fischbacher (2003): "The Nature of Human Altruism," *Nature*, 425, 785–791.

Fehr, E., and K. M. Schmidt (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 14, 815–868.

Frank, R. H. (1988): *Passions Within Reason*. W. W. Norton, New York, New York.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr (2005): *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life*. MIT Press, Cambridge, Massachusetts.

Grégoire, P., and A. J. Robson (2003): "Imitation, Group Selection and Cooperation," *International Game Theory Review*, 5(3), 229–247.

Hammerstein, P. (ed.) (2003): *Genetic and Cultural Evolution of Cooperation*, Dahlem workshop reports. Dahlem University Press, USA.

Hoffman, E., K. McCabe, and V. Smith (1996): "Social Distance and Other-Regarding Behavior in Dictator Games," *American Economic Review*, 86(3), 653–660.

Jackson, M. O., and A. Watts (2005): "Social Games: Matching and the Play of Finitely Repeated Games," *Mimeo*.

Kandori, M., G. J. Mailath, and R. Rob (1993): "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, 61(1), 29–56.

Maynard Smith, J. (1982): *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, Massachusetts.

McCabe, K., M. L. Rigdon, and V. L. Smith (2003): "Sustaining Cooperation in Trust Games," Mimeo, Harvard University.

Rob, R., and H. Yang (2005): "Long-Term Relationships as Safeguards," *Mimeo*.

Robson, A. J. (1990): "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake," *Journal of Theoretical Biology*, 144(3), 379–396.

Samuelson, L. (1998): *Evolutionary Games and Equilibrium Selection*. MIT Press, Cambridge, Massachusetts.

Tullock, G. (1985): "Adam Smith and the Prisoners' Dilemma," *Quarterly Journal of Economics*, 100, 1073–1081.

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61(1), 57–84.