Ulrich Berger

Learning to cooperate via indirect reciprocity

Article (Accepted for Publication)
(Refereed)

Learning to cooperate via indirect reciprocity

Ulrich Berger

WU Vienna, Department of Economics, Augasse 2-6, 1090 Wien, Austria

Abstract

Cooperating in the Prisoner's Dilemma is irrational and some supporting mechanism is needed to stabilize cooperation. Indirect reciprocity based on reputation is one such mechanism. Assessing an individual's reputation requires first-order information, i.e. knowledge about its previous behavior, as it is utilized under image scoring. But there seems to be an agreement that in order to successfully stabilize cooperation, higher-order information is necessary, i.e. knowledge of others' previous reputations. We show here that such a conclusion might have been premature. *Tolerant scoring*, a first-order assessment rule with built-in tolerance against single defections, can lead a society to stable cooperation.

Key words: Evolution; Cooperation; Prisoner's Dilemma; Indirect reciprocity; Scoring rule; First-order information

1 Introduction

1.1 Indirect reciprocity

Cooperation is ubiquitous in economic life, but in its purest form it is neither rational nor evolutionarily stable. Cooperating by acting altruistically and helping those in need reduces the actor's payoff while it increases the recipient's payoff. Such a behavior, while socially desirable, is dominated by defection, as the paradigm of the Prisoner's Dilemma game teaches us. Sustainable cooperation requires a supporting mechanism, i.e. a framework which

Email address: ulrich.berger@wu-wien.ac.at (Ulrich Berger).

[1] I want to thank two anonymous referees and participants of the Biomathematics seminar and the Miniworkshop on Evolutionary Game Theory (University of Vienna), the JERS seminar (Max Planck Institute of Economics, Jena), and the TECT conference (IIASA, Laxenburg) for helpful comments.

transforms the basic game into a situation where the dilemma vanishes. Several such mechanisms are known—see Nowak (2006) for a recent review—and among those, *indirect reciprocity* has recently received increasing attention.

Like direct reciprocity, indirect reciprocity (Trivers, 1971, Sugden, 1986, Alexander, 1987) relies on the idea that the probability of receiving help is higher for those who helped in the past. However, under indirect reciprocity helpful acts are not returned by the receivers of these acts but by third parties.[2] This requires that individuals carry an observable reputation which is informative about their past behavior. From Kandori (1992) and Okuno-Fujiwara and Postlewaite (1995) to Takahashi (2010), work on repeated games with random matching has demonstrated that under these assumptions, *community enforcement* has the potential to uphold equilibrium cooperation as a social norm in a group of forward-looking rational agents. Likewise, evolutionary approaches have shown that *discriminators* who base their decision whether or not to help on the potential recipient's reputation have the chance to survive or even spread in populations of unconditional cooperators and defectors. Since Nowak and Sigmund (1998a, 1998b) first presented a formal evolutionary model, a variety of specific rules for indirect reciprocity have been studied in detail. Our own approach starts with a critical evaluation of this very first model.

1.2 Image scoring

Under the *image scoring* mechanism (Nowak and Sigmund, 1998a, 1998b, Lotem et al, 1999), each individual is equipped with a numerical score which measures its past cooperativeness by counting how often it helped on its last interactions. In the simplest case, scoring is *binary* and discriminators assess other individuals as either *Good* or *Bad*, depending on whether or not they helped on their last interaction. Upon meeting an individual, discriminators then help those and only those which they assessed as Good. Scoring is called a *first-order assessment rule*, since in assessing an individual, i.e. in updating this individual's reputation, it relies only on the focal individual's behavior towards its partner, but neither on this partner's reputation nor on the focal individual's previous reputation.

Under binary scoring, discriminators cannot be exploited by unconditional defectors, since they will never give help to those. While this sounds promising for the sustainability of cooperation in a population made up of cooperators, defectors, and discriminators, there are two subtle problems. These are associated with two terms Axelrod (1984) used to characterize successful strategies in the Iterated Prisoner's Dilemma.

[2] Binmore (1992) called this the *I-won't-scratch-your-back-if-you-won't-scratch-their-backs principle* and traced it to David Hume.

- The *"niceness problem"*: Discriminators are "nice" in the sense that they never exploit unconditional cooperators, since the latter are assessed as Good. But this means that in the absence of defectors, discriminators and cooperators are behaviorally indistinguishable. Both cooperate on each interaction and receive the same payoff. As a consequence, discrimination is not evolutionarily stable and neutral drift or random fluctuations might drive such a population to a state where unconditional cooperators abound. Thereafter, as soon as experiments or mutations reintroduce defectors into the population, these can exploit the large number of cooperators, thereby collect high payoffs and finally overtake the population.
- The *"provocability problem"*: Discriminators are "provocable" since they react to an observed defection by defecting themselves. While this behavior seems to be necessary to avoid exploitation by defectors, it has a consequence which is detrimental for the success of discriminators. A discriminator who punishes a defector by withholding help is itself assessed as Bad and hence being punished by defection if he meets another discriminator thereafter. The reason is that under image scoring, discriminators do not distinguish between "justified" and "unjustified" defections. As a result, the presence of a few defectors can trigger a wave of cannibalism among discriminators, which lowers their payoffs and makes them increasingly vulnerable.

Note that in the presence of errors in implementing cooperation, provocability even exacerbates the niceness problem. If a population of discriminators is subjected to a small rate of involuntary defections, these defections are punished by withholding help, which increases the rate of defection even further. A single cooperator entering such a population then has a higher probability of receiving help than a discriminator, which results in a payoff advantage enabling cooperators to invade the discriminator population even without relying on neutral drift.

These problems with the image scoring mechanism and their implications have been laid out in the literature during the last couple of years. It is now generally accepted that under image scoring, the discriminator strategy can survive in the population for some time but is bound to get extinct in the long run, see Panchanathan and Boyd (2003), Ohtsuki (2004), or Brandt and Sigmund (2006). While image scoring behavior in humans is supported by experimental research,[3] it remained unclear how it might be able to evolve or stabilize.[4]

1.3 Higher-order information

In overcoming the problems associated with image scoring, researchers have turned to more sophisticated, *second-order* and *third-order assessment rules*

[3] E.g. Wedekind and Milinski (2000), Milinski et al (2001), Bolton et al (2005), or Seinen and Schram (2006).

[4] For a review see Nowak and Sigmund (2005).

for binary reputations.[5] The primary focus of this literature has been to study assessment rules which are able to distinguish between "justified" and "unjustified" defections, thereby alleviating the provocability problem noted above.

For example, when assessing an individual's behavior based on its last action chosen, Sugden's (1986) *standing rule* takes into account the reputation of this individual's opponent. Under this rule, cooperation makes you Good, and defection against a Good individual makes you Bad, but defecting against a Bad opponent is viewed as justified and does not change your current reputation.

While the standing rule as well as several other higher-order assessment rules have been found to render discrimination evolutionarily stable,[6] those rules suffer from being informationally and cognitively extremely demanding as well as having weak experimental support (Milinski et al, 2001). This makes it somewhat questionable if they can provide a rationale for the evolution of indirect reciprocity.

1.4 Tolerant scoring

In this paper we choose a different route. Instead of trying to completely avoid the provocability problem by introducing higher-order information, we stick to a simple assessment rule of the scoring type requiring only first-order information. However, we reduce the intensity of the provocability problem by increasing the amount of information provided on individuals' past behavior. To be precise, we assume that in assessing his opponent a discriminator samples *two actions* from its past. Then the discriminator views the opponent as Good (and therefore cooperates) if and only if the opponent has helped *at least once* in these two actions. We call this rule *tolerant scoring*, since it tolerates a single defection, and distinguish it from Nowak and Sigmund's (1998b) binary scoring model by referring to the latter as *simple scoring*. The relationship between simple scoring and tolerant scoring may be viewed as analogous to the relationship between Tit-for-Tat and the Tit-for-Two-Tats strategy Axelrod (1984) devised for the second of his computer tournaments.

Note that sampling two instead of only one action from the opponent's past increases the amount of information available, but it still remains first-order information. No knowledge about reputation or the actions of the focal opponent's previous opponents is involved. Sampling two actions from an individual's past behavior may be viewed as remembering having directly watched these past interactions or, alternatively, as having gathered this information

[5] E.g. Leimar and Hammerstein (2001), Ohtsuki (2004), Ohtsuki and Iwasa (2004), Brandt and Sigmund (2004), Chalub et al (2006), and Pacheco et al (2006).
[6] See Ohtsuki and Iwasa (2006, 2007), Ohtsuki et al (2009), Uchida and Sigmund (2010).

indirectly by asking around.

Obviously, under tolerant scoring discriminators are considerably less provocable than under simple scoring. By tolerating a single defection, they are less likely to punish other discriminators who defected previously. To illustrate this, consider an overall defection rate of, say, 0.1 among discriminators. Under simple scoring this induces a ten percent probability for a discriminator of being punished when meeting its own kind. Under tolerant scoring, however, this probability is squared to one percent. At the same time, discriminators remain immune to exploitation by defectors, since defectors never help and are therefore always punished by discriminators. So discriminators using tolerant scoring have at least the potential to survive in a world of unconditional cooperators and defectors. Indeed we show here that under tolerant scoring discrimination is a strict Nash equilibrium and hence evolutionarily stable for small positive error rates. Even more, it turns out that under the best-response dynamics almost fully cooperative discrimination can become established from arbitrary inital conditions.

2 Model

2.1 The donation game, errors, and reputations

Consider a large population of infinitely-lived individuals. Time τ is continuous and individuals are repeatedly and randomly matched in pairs to interact in the donation game. During each interaction, one individual is randomly chosen to be the donor and the other to be the receiver. Donors can either give help (cooperate, C) or not (defect, D) to the receiver. Helping decreases the donor's payoff by an amount c and increases the receiver's payoff by b, where $b > c > 0$. For convenience we will make the usual assumption that actually each individual plays in both roles at the same time during an interaction. With a small probability $\alpha > 0$ a donor who intends to cooperate is not able to do so (e.g. due to lack of resources) and instead defects. With probability $\bar{\alpha} = 1 - \alpha$ an intended cooperation is implemented correctly. No implementation errors are assumed if a donor intends to defect.

Before the donor implements his action, he is informed of ("recalls") the outcome of two past interactions where the current receiver was in the role of a donor. Those two interactions are sampled independently from the receiver's past behavior.[7] The reputation the donor assigns to the receiver depends on the number of helping acts (C's), and the receiver is assessed by the donor as *Good* if he helped on one or both occasions and as *Bad* if he defected in both

[7] For technical simplicity we assume sampling with replacement. While this makes it possible that the two past actions sampled are one and the same, it doesn't change the results.

interactions.

2.2 Strategies and population dynamics

We study the following three pure strategies:

- *Cooperators* always (intend to) cooperate. This strategy is called *AllC* as usual. Its frequency is denoted by x.
- *Defectors* always defect. This is the *AllD* strategy with frequency y.
- *Discriminators* (intend to) cooperate if and only if the opponent's reputation is assessed as Good. This strategy is denoted by *Disc* and has frequency z.

Since indirect reciprocity has predominantly been studied in the biological literature, changes in strategy frequencies have usually been interpreted as resulting from evolutionary forces and consequently been modeled by replicator dynamics. However, in the context of human behavior we find it at least equally reasonable to posit a basic version of bounded rationality and therefore assume that strategy updating is guided by the learning dynamics known as the *best-response dynamics* (Gilboa and Matsui, 1991, Matsui, 1992, see also Hofbauer, 2000 and Hofbauer and Sigmund, 1998).[8] Under the best-response dynamics, individuals are chosen every now and then to update their strategy choice. Updating individuals choose a myopic pure best response to the current population state. This results in the population state $(x(\tau), y(\tau), z(\tau))$ moving along (possibly non-unique) solutions of the differential inclusion

$$(\dot{x}, \dot{y}, \dot{z}) \in B(x, y, z) - (x, y, z), \tag{1}$$

where $B(x, y, z)$ is the set of (pure or mixed) best responses to the strategy profile (x, y, z). As long as the best responses are unique, the population state moves along a straight line pointing to the current pure best response.

2.3 Reputation dynamics

Changes of strategy frequencies are due to differential payoffs, which depend on individuals' reputations. But reputations influence behavior and are therefore themselves subject to change. We assume here the existence of two different time-scales. Learning via the best-response dynamics occurs on a slow time scale τ, while reputation dynamics occur on a discrete and fast time-scale t. We show below that reputations always converge to an equilibrium and it is therefore justified to treat reputation as instantly equilibrated when deriving the payoffs which determine the learning dynamics on the slow time-scale.

So let us assume for a moment that the population state (x, y, z) is fixed and

[8] The choice of dynamics does not change our results qualitatively, however.

interactions occur at discrete time steps $t = 0, 1, 2, \ldots$. Let the initial two moves at $t = 0$ and $t = 1$ for a cooperator be C, C, for a defector D, D, and for a discriminator two arbitrary moves. For any time $t_0 \geq 2$ the *past* of an individual at time t_0 is defined as its sequence of chosen actions during $0 \leq t \leq t_0 - 1$. We assume that when sampling two past moves of an individual, errors in perception do not occur. However, assessments of the same individual may of course differ across assessors, since sampling makes an individual's reputation a random variable—reputation is in the eye of the beholder. [9]

The helping probability of defectors is always zero. The helping probability of cooperators depends only on the error rate and is given by $\bar{\alpha}$, independently of the population state. For discriminators the situation is more difficult. Their helping probability is subject to dynamic change. Discriminators help if they assess their opponent as Good, the probability of which depends on the average cooperation rate during the opponent's past. This average past helping rate can in turn be derived from the population state (x, y, z), the error rate α and the sequence of discriminators' *current helping probabilities*. At time t the current helping probability of discriminators, denoted by p_t, is defined as the probability that a randomly chosen discriminator plays a C at time t. We show in the appendix that p_t always converges to a limit value p, which determines the *reputation equilibrium*.

In reputation equilibrium, the discriminators' helping probability p turns out to be implicitly given by

$$p = \bar{\alpha}[(1 - \alpha^2)x + p(2 - p)z]. \tag{2}$$

To see this, note that a discriminator intends to help if he meets a cooperator (probability x) who has not erroneously defected twice (probability $1 - \alpha^2$) or if he meets another discriminator (probability z) who has not defected in both instances (probability $1 - (1 - p^2) = p(2 - p)$). Finally, intended cooperation succeeds with probability $\bar{\alpha}$.

Thus, a discriminator's helping probability can be calculated as

$$p = \begin{cases} \bar{\alpha}(1 - \alpha^2)x & \ldots \quad z = 0 \\ \\ 1 - \frac{1}{2\bar{\alpha}z} + \sqrt{\left(1 - \frac{1}{2\bar{\alpha}z}\right)^2 + (1 - \alpha^2)\frac{x}{z}} & \ldots \quad z > 0 \end{cases} \tag{3}$$

Note that in the absence of cooperators, i.e. for $x = 0$, the expression for p reduces to

[9] With the obvious exception of defectors, who will always be assessed as Bad.

$$x = 0 \Rightarrow \quad p = \begin{cases} 0 & \dots \quad z \leq \frac{1}{2\bar{\alpha}} \\ \\ 2 - \frac{1}{\bar{\alpha} z} & \dots \quad z > \frac{1}{2\bar{\alpha}} \end{cases} \tag{4}$$

2.4 Payoffs

Given a reputation equilibrium determined by p, the cooperation and defection probabilities of the three strategies as well as their probability of being assessed as Good by a Discriminator are given by

strategy	coop. prob.	defection prob.	prob. of Good assessment
$AllC$	$\bar{\alpha}$	α	$1 - \alpha^2$
$AllD$	0	1	0
$Disc$	p	$1 - p$	$1 - (1-p)^2 = p(2-p)$

This allows us to write down the *probability-matrix* of an S-strategist's probability $q(S, T)$ of *intending* to help a T-strategist.

$q(S,T)$	$AllC$	$AllD$	$Disc$
$AllC$	1	1	1
$AllD$	0	0	0
$Disc$	$1 - \alpha^2$	0	$p(2-p)$

From this probability-matrix we can finally derive the payoff-matrix. To do this, note that due to the possibility of errors, an S-strategist's probability of *actually* helping a T-strategist is $\bar{\alpha} q(S, T)$ and the S-strategist's payoff therefore is

$$\bar{\Pi}(S, T) = \bar{\alpha}[bq(T, S) - cq(S, T)]. \tag{5}$$

Dividing by the strictly positive factor $\bar{\alpha}$ and inserting from the probability-matrix leads to the payoff-matrix $\Pi(S, T) = \frac{1}{\bar{\alpha}} \bar{\Pi}(S, T)$, given by

$$\Pi = \begin{array}{c|ccc} & AllC & AllD & Disc \\ \hline AllC & b-c & -c & (1-\alpha^2)b-c \\ AllD & b & 0 & 0 \\ Disc & b-(1-\alpha^2)c & 0 & p(2-p)(b-c) \end{array} \tag{6}$$

Note that this payoff-matrix is state-dependent, since p depends on the population state (x, y, z) via (3).

2.5 Best response regions

2.5.1 Vanishing error rate

Let us start with the case $\alpha = 0$. In this case the payoffs reduce to

$$\Pi_0 = \begin{array}{c|ccc} & AllC & AllD & Disc \\ \hline AllC & b-c & -c & b-c \\ AllD & b & 0 & 0 \\ Disc & b-c & 0 & p(2-p)(b-c) \end{array} \tag{7}$$

where $p = 1 - \frac{1}{2z} + \sqrt{\left(1-\frac{1}{2z}\right)^2 + \frac{x}{z}}$ for $z > 0$ and $p = x$ for $z = 0$. From equations (3) and (4), on the *AllC*-*Disc* edge $\{y = 0\}$ we have $p = 1$ and on the *AllD*-*Disc* edge $\{x = 0\}$ we have $p = 0$ for $z \leq \frac{1}{2}$ and $p = 2 - \frac{1}{z}$ for $z > \frac{1}{2}$.

Substituting for p in the payoff matrix, we can determine the best response to any population state mixture (x, y, z) with $x > 0$ by computing the vector of expected payoffs $\Pi_0 \cdot (x, y, z)^T$ and comparing the components. This results in the following best response regions for $x > 0$:

- *AllD* is a best response if and only if $z \leq \frac{c(b-c)}{2b(b-c)-b^2x}$.
- *AllC* is a best response if and only if ($y = 0$ and $z \geq \frac{c}{b}$) or $x \leq \frac{\left(z-\frac{c}{b}\right)\left(\frac{b-c}{b}-z\right)}{z}$.
- *Disc* is a best response if and only if $x \geq \frac{\left(z-\frac{c}{b}\right)\left(\frac{b-c}{b}-z\right)}{z}$ and $z \geq \frac{c(b-c)}{2b(b-c)-b^2x}$.

It follows that on the edge $\{y = 0\}$, both *AllC* and *Disc* are best responses for $z \geq \frac{c}{b}$. This implies that the edge-segment $N_{xz} = \{y = 0, z \geq \frac{c}{b}\}$ is a component of Nash equilibria for $\alpha = 0$.

Remember from equation (4) that in the absence of cooperators the cooperation rate vanishes once more than one half of the population is comprised of defectors. On the edge $\{x = 0\}$, therefore, both *AllD* and *Disc* are best responses whenever $z \leq \frac{1}{2}$ and $z \leq \frac{c}{b}$ and $z \leq 1 - \frac{c}{b}$, i.e. whenever $z \leq \min\{1-\frac{c}{b}, \frac{c}{b}\}$. This edge-segment therefore constitutes a second component of Nash equilibria we

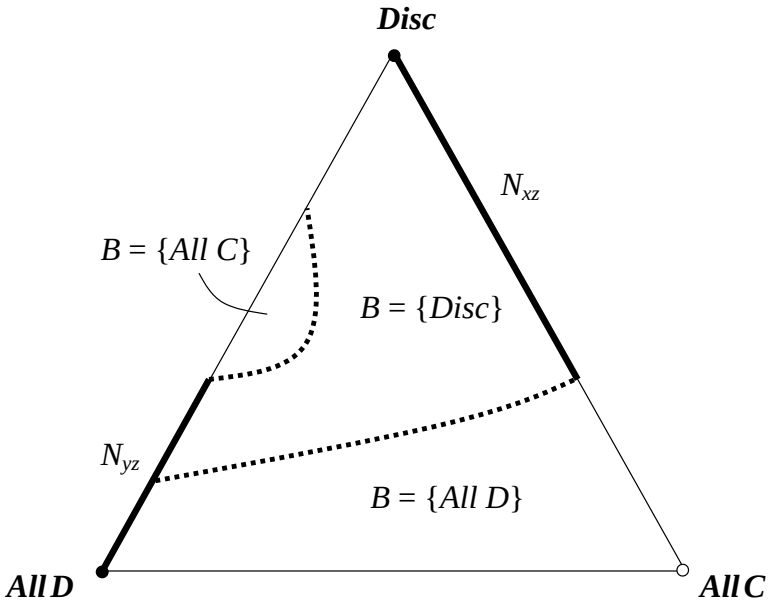

Fig. 1. Interior best-response regions and Nash equilibrium components for $\alpha = 0$ and $b = 3c$.

denote by N_{yz}. *Disc* is also a best response on the segment $z \geq \max\{\frac{c}{b}, 1 - \frac{c}{b}\}$. If $b > 2c$, then there is an intermediate region $\frac{c}{b} \leq z \leq 1 - \frac{c}{b}$, where *AllC* has the highest payoff. The best-response regions and Nash equilibrium components are shown in Figure 1.

2.5.2 Positive error rate

If we go from $\alpha = 0$ to a small positive error rate $\alpha > 0$, then continuity of payoffs in α implies that the best response regions in the interior of the state space change only little. This holds also for the best response regions on the *AllD-Disc* edge $\{x = 0\}$, where the Nash equilibrium component N_{yz} now extends over the range $z \leq \min\{1 - \frac{c}{(1-\alpha^2)b}, \frac{c}{(1-\alpha^2)b}\}$.

Writing the payoff difference between discriminators and cooperators at the population state (x, y, z) as $\delta(x, y, z) = \Pi(Disc) - \Pi(AllC)$, we can calculate the payoff difference Δ at the *AllC-Disc* edge as $\Delta := \delta(1 - z, 0, z) = \alpha^2[zb + (1 - z)c] - z(b - c)(1 - p)^2$.

Denoting by Δ' and p' the derivatives with respect to α, from equation (2) it follows that $p' = -[(1-\alpha^2)x + p(2-p)z] + 2\bar{\alpha}[p'(1-p)z - \alpha x]$ and since $\alpha = 0$ implies $p = 1$ at the *AllC-Disc* edge, $p'|_{\alpha=0} = -1$ for $x = 1 - z$. We can also calculate $\Delta' = 2\alpha[zb + (1-z)c] + 2z(b-c)(1-p)p'$ and $\Delta'' = 2[zb + (1-z)c] + 2z(b-c)[(1-p)p'' - p'^2]$. Substituting $\alpha = 0$, $p = 1$, and $p'|_{\alpha=0} = -1$ yields $\Delta|_{\alpha=0} = \Delta'|_{\alpha=0} = 0$ and $\Delta''|_{\alpha=0} = 2c > 0$. A second-order Taylor expansion of Δ in the error rate α therefore shows that $\Delta > 0$ for small $\alpha > 0$, implying that on the *AllC-Disc* edge discriminators outcompete cooperators for small positive error rates. In this case the Nash equilibrium component N_{xz} collapses to a single and strict Nash equilibrium at the *Disc*-vertex. This proves that the *Disc* strategy is evolutionarily stable. The equilibrium cooperation rate is given by $p = 2 - \frac{1}{\bar{\alpha}} = \frac{1-2\alpha}{1-\alpha}$, which is close to full cooperation for a small error rate.

The best response regions for $\alpha > 0$ are depicted in Figure 2 for the case $b = 3c$. From this figure we can also derive the dynamic properties of population-level learning. In the region where discriminators earn the highest payoff, solutions of the best-response dynamics head straight towards the *Disc*-vertex. For a set of initial points near the Nash equilibrium component, solutions run into the boundary of the hump-shaped best response region of *AllC*. Those solutions then wander upwards along the boundary until they can continue to move along a straight line and converge to the *Disc*-vertex.

Any solution starting in the *AllD* best response region in the interior of the state space converges along a straight line towards the *AllD*-vertex. Note, however, that the *AllD*-vertex itself lies within a component of Nash equilibria and is therefore not evolutionarily stable. Indeed, starting from the *AllD*-vertex, there exist solutions which travel along the Nash equilibrium component until they enter the best response region of *AllC*, following the hump-shaped boundary of this region upwards and finally heading off towards the discriminator vertex.

3 Discussion

In the presence of implementation errors tolerant scoring makes both the "niceness problem" and the "provocability problem" disappear. A small level of noise leads to slightly different cooperation rates among discriminators and cooperators, even in the absence of defectors. Discrimination then results in a higher payoff and cooperators cannot invade. At the same time, tolerance against single defections greatly reduces provocability of discriminators among its own kind, but does not diminish its strength against unconditional defectors.

We have seen that a low initial level of discriminators may lead to the disappearance of both cooperators and discriminators, but as soon as uncondi-

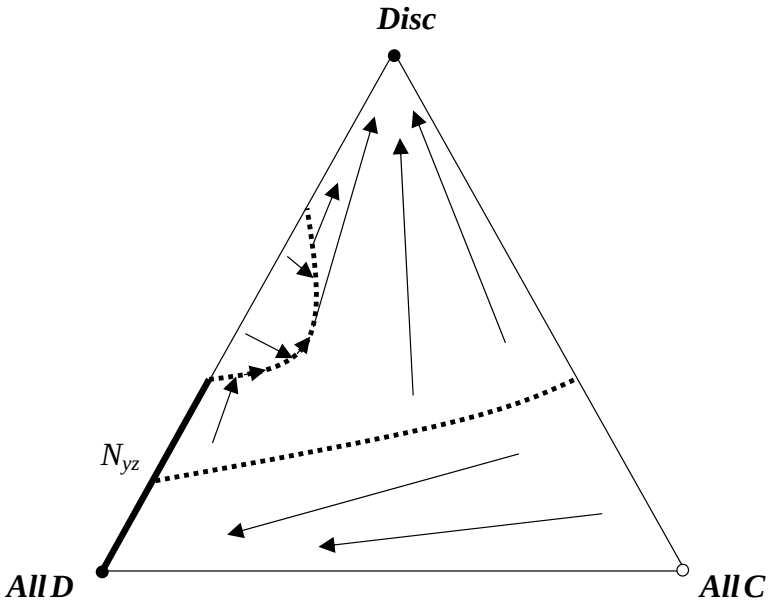

Fig. 2. Best-response dynamics for tolerant scoring. Here a small positive error rate is assumed and the case $b = 3c$ is shown. Within the Nash equilibrium component N_{yz} the direction of movement is indetermined.

tional cooperation vanishes and everyone defects, the discriminating strategy no longer suffers from a payoff disadvantage and individuals may (but need not) return to discrimination, offering the possibility that in the long run the whole society learns to effectively discriminate. So discrimination is not only a strict and therefore locally attracting Nash equilibrium, but may finally be learned from arbitrary initial conditions.

It seems plausible that the discriminating strategy should carry additional cognitive costs as compared to the two unconditional strategies. Clearly, the introduction of such a cost would result in defection being evolutionarily stable. But since discrimination is a strict Nash equilibrium, it would still remain strict as long as the costs are small enough. Thus, evolutionary stability of discrimination is robust to the introduction of small cognitive costs.

We stress that tolerance is the key to successful cooperation in our model. Previous cooperation is rewarded by discriminators through helping. A single defection is tolerated, but not two defections. This allows discriminators to

strike a balance between being too lenient, which leads to exploitation by defectors, and being too rigid, which results in mutual punishment within its own kind. This middle course requires three distinct types of observable behavior. Checking only a single instance of an individual's past behavior seems to provide too coarse a measure to build a working reputation system on.

Double-checking her deeds before judging a person's character is not an unrealistic requirement for a basic reputation system. But the prime advantage of tolerant scoring is that the cognitive and informational load it puts on its subjects is considerably lower than the one of higher-order assessment rules. For the evolution of cooperation via indirect reciprocity this makes tolerant scoring a promising alternative rule.

Appendix

By definition, $p_t = \bar{\alpha} G_t$, where G_t is the probability of assessing an opponent as Good at time t. Let $g_z(t)$ denote the probability that a randomly chosen discriminator is assessed as Good at time t. Then $G_t = (1-\alpha^2)x + 0y + g_z(t)z$ and substituting yields

$$p_t = \bar{\alpha}[x(1-\alpha^2) + z g_z(t)].$$

Next, let a_t denote the *average past helping frequency* of a tolerant discriminator at time t, i.e. the probability that a randomly chosen past move of a randomly chosen discriminator is a C. If the expected number of C's in the past is c_t, then $a_t = c_t/t$. Hence $g_z(t) = 1 - (1-a_t)^2$, or $g_z(t) = a_t(2-a_t)$, and we can write

$$p_t = \bar{\alpha}[(1-\alpha^2)x + a_t(2-a_t)z].$$

In round $t+1$, let the discriminator's move be $m \in \{C, D\}$. Then $m = D \Rightarrow a_{t+1} = \frac{c_{t+1}}{t+1} = \frac{c_t}{t+1} = \frac{t}{t+1} a_t$ and $m = C \Rightarrow a_{t+1} = \frac{c_{t+1}}{t+1} = \frac{c_t+1}{t+1} = \frac{t}{t+1} a_t + \frac{1}{t+1}$. The probability for a cooperative move at time t is p_t, hence $a_{t+1} = \frac{t}{t+1} a_t + \frac{1}{t+1} p_t$ and inserting from above yields

$$a_{t+1} = \frac{t}{t+1} a_t + \frac{1}{t+1} \bar{\alpha}[(1-\alpha^2)x + a_t(2-a_t)z].$$

We show now that the sequence (a_t) of average past helping frequencies always converges. To see this, note that in the absence of discriminators, i.e. for $z = 0$, the sequence converges to $\bar{\alpha}(1-\alpha^2)x$. If $z > 0$ and $x > 0$, then there is a unique fixed point a, which is implicitly given by $a = \bar{\alpha}[(1-\alpha^2)x + a(2-a)z]$ and can be calculated as

$$a = 1 - \frac{1}{2\bar{\alpha}z} + \sqrt{\left(1 - \frac{1}{2\bar{\alpha}z}\right)^2 + (1-\alpha^2)\frac{x}{z}}.$$

The sequence (a_t) gives rise to the sequence (p_t) of current helping probabilities with a unique fixed point p. Note that at the fixed point the current helping probability equals the average past helping frequency, i.e. $p = a$.

For fixed t we can consider p_t as a function of a_t. It is easy to see that for $z > 0$ this function is strictly increasing, rising from $\bar{\alpha}(1-\alpha^2)x$ for $a_t = 0$ to $\bar{\alpha}(1-\alpha^2)x + \bar{\alpha}z$ for $a_t = 1$, and crossing the diagonal at $p_t = a_t = a$. It follows that for $x > 0$,

$$a_t < a \Rightarrow a_t < p_t < a \Rightarrow a_t < a_{t+1} < a$$

and

$$a_t > a \Rightarrow a_t > p_t > a \Rightarrow a_t > a_{t+1} > a.$$

Therefore, if $x > 0$, a_t converges monotonically to a from any initial value. The same holds for the sequence p_t, so for $x > 0$ and $z > 0$ the tolerant discriminators' helping probability converges to

$$p = 1 - \frac{1}{2\bar{\alpha}z} + \sqrt{\left(1 - \frac{1}{2\bar{\alpha}z}\right)^2 + (1-\alpha^2)\frac{x}{z}}.$$

If there are no cooperators, i.e. if $x = 0$, then $\hat{a} = 0$ is a fixed point of the sequence (a_t). If $z > \frac{1}{2\bar{\alpha}}$, then this fixed point is a repellor and a_t converges to $a = 2(1 - (2\bar{\alpha}z)^{-1}) > 0$. However, if $z \leq \frac{1}{2\bar{\alpha}}$, then $a = \hat{a} = 0$. This unique fixed point is attracting and cooperation disappears completely. For the *AllD-Disc* edge this means that *Disc* has a payoff advantage whenever it is already sufficiently frequent, viz. whenever $z > \frac{1}{2\bar{\alpha}}$. If the frequency of discriminators is below this threshold, then both strategies earn zero payoff in reputation-equilibrium. Along the edge-segment $x = 0$ and $0 \leq z \leq \frac{1}{2\bar{\alpha}}$ therefore defectors and discriminators are indistinguishable.

References

[1] Alexander RD (1987) The Biology of Moral Systems. NewYork: Aldine deGruyter.

[2] Axelrod R (1984) The Evolution of Cooperation. Basic Books, New York.

[3] Binmore K (1992) Fun and Games. D. C. Heath, Lexington, MA.

[4] Bolton G, Katok E, Ockenfels A (2005) Cooperation among strangers with limited information about reputation. J Public Econ 89: 1457-1468.

[5] Brandt H, Sigmund K (2004) The logic of reprobation: Assessment and action rules for indirect reciprocation. J Theor Biol 231: 475-486.

[6] Brandt H, Sigmund K (2005) Indirect reciprocity, image scoring, and moral hazard. Proc Natl Acad Sci USA 102: 2666-2670.

[7] Brandt H, Sigmund K (2006) The good, the bad and the discriminator - errors in direct and indirect reciprocity. J Theor Biol 239: 183-194.

[8] Chalub FACC, Santos FC, Pacheco JM (2006) The evolution of norms. J Theor Biol 241: 233-240.

[9] Gilboa I, Matsui A (1991) Social stability and equilibrium. Econometrica 59: 859-867.

[10] Hofbauer J (2000) From Nash and Brown to Maynard Smith: Equilibria, dynamics and ESS. Selection 1: 81-88.

[11] Hofbauer J, Sigmund K (1998) Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge, UK.

[12] Kandori M (1992) Social norms and community enforcement. Rev Econ Stud 59: 63-80.

[13] Leimar O, Hammerstein P (2001) Evolution of cooperation through indirect reciprocity. Proc Biol Sci 268: 745-753.

[14] Lotem A, Fishman AM, Stone L (1999) Evolution of cooperation between individuals. Nature 400: 226-227.

[15] Matsui A (1992) Best response dynamics and socially stable strategies. J Econ Theory 57: 343-362.

[16] Milinski M, Semmann D, Bakker TCM, Krambeck HJ (2001) Cooperation through indirect reciprocity: Image scoring or standing strategy? Proc R Soc Lond B 268: 2495-2501.

[17] Nowak MA (2006) Five Rules for the Evolution of Cooperation. Science 314: 1560-1563.

[18] Nowak MA, Sigmund K (1998a) Evolution of indirect reciprocity by image scoring. Nature 393: 573-577.

[19] Nowak MA, Sigmund K (1998b) The dynamics of indirect reciprocity. J Theor Biol 194: 561-574.

[20] Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437: 1291-1298.

[21] Ohtsuki H (2004) Reactive strategies in indirect reciprocity. J Theor Biol 227: 299-314.

[22] Ohtsuki H, Iwasa Y (2004) How should we define goodness? Reputation dynamics in indirect reciprocity. J Theor Biol 231: 107-120.

[23] Ohtsuki H, Iwasa Y (2006) The leading eight: Social norms that can maintain cooperation by indirect reciprocity. J Theor Biol 239: 435-444.

[24] Ohtsuki H, Iwasa Y (2007) Global analysis of evolutionary dynamics and exhaustive search for social norms that maintain cooperation and reputation. J Theor Biol 244: 518-531.

[25] Ohtsuki H, Iwasa Y, Nowak MA (2009) Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. Nature 457: 79-82.

[26] Okuno-Fujiwara M, Postlewaite A (1995) Social norms and random matching games. Games Econ Behav 9: 79-109.

[27] Pacheco JM, Santos FC, Chalub FACC (2006) Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. PLoS Comput Biol 2: e178.

[28] Panchanathan K, Boyd R (2003) A tale of two defectors: The importance of standing for evolution of indirect reciprocity. J Theor Biol 224: 115-126.

[29] Seinen I, Schram A (2006) Social status and group norms: Indirect reciprocity in a repeated helping experiment. European Econ Rev 50: 581-602.

[30] Sugden, R (1986) The Economics of Rights, Co-operation and Welfare. Basil Blackwell, Oxford.

[31] Takahashi S (2010) Community enforcement when players observe partners' past play. J Econ Theory 145: 42-62.

[32] Trivers RL (1971) Evolution of reciprocal altruism. Quarterly Rev Biol 46: 35-57.

[33] Uchida S, Sigmund K (2010) The competition of assessment rules for indirect reciprocity. J Theor Biol 263: 13-19.

[34] Wedekind C, Milinski M (2000) Cooperation through image scoring in humans. Science 288: 850-852.