ON THE EVOLUTIONARY EMERGENCY OF OPTIMISM

Aviad Heifetz

Yossi Spiegel
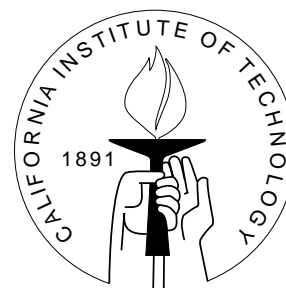
# ON THE EVOLUTIONARY EMERGENCE OF OPTIMISM

Aviad Heifetz[*] and Yossi Spiegel[†]

First version: December 1999
This version: October 2000

## Abstract

Successful individuals were frequently found to be overly optimistic. These finding are puzzling, as one could expect that realists would perform best in the long run. We show, however, that in a large class of strategic interactions of either cooperation or competition, the equilibrium payoffs of optimists may be higher than those of realists. This is because the very fact of being optimistic changes the game, and drives the adversary to change her equilibrium behavior, possibly to the benefit of the optimist. Suppose, then, that a population consists initially of individuals with various perceptional tendencies – pessimists and optimists to various extents, as well as of realists. Individuals meet in pairs to interact, and more successful tendencies proliferate faster. We show that as time goes by, some moderate degree of optimism will take over, and outnumber all other tendencies.

---

[*]The Division of Humanities and Social Sciences, California Institiute of Technology, mail code 228-77, Pasadena, CA 91125, and The Eitan Berglass School of Economics, Tel Aviv University 69978, Israel. email: heifetz@post.tau.ac.il

[†]Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208, and the Eitan Berglass School of Economics, Tel Aviv University, Tel Aviv 69978, Israel. email: spiegel@post.tau.ac.il.

# 1 Introduction

There is by now a considerable body of evidence, that in many kinds of circumstances, successful individuals are overly optimistic regarding the return to their own investment or effort. Taylor and Brown (1988) found that most mentally healthy people have somewhat unrealistically positive self-views, while the less mentally healthy perceive themselves more accurately. In peer reviews, for example, non-depressed individuals' self-ratings were considerably more favorable than those given to them by others (Lewinsohn, Mischel, Chaplin and Barton 1980). Non-depressed individuals exhibited an illusion of control in a dice-throwing experiment (Fleming and Darley 1986). Most individuals believe that their driving ability is above average (Svenson 1981), and most young Americans know that half of U.S. marriages end in divorce, but they are confident that theirs will not (Lehman and Nisbett, 1985).[1] In the Economic arena, excess entry of new businesses that fail within several years is common in the US, and recent experimental work (Camerer and Lovallo, 1999) suggests that this phenomenon may be due to entrepreneurs being overconfident regarding their own ability in comparison with other entrepreneurs. Babcock and Loewenstein (1997) review experimental work that suggests that parties to legal disputes are reluctant to settle out of court because they hold overly optimistic beliefs about the merits of their case.

These observations are puzzling because it might be thought that optimistic individuals, who consistently overestimate their eventual payoffs, will not do as well as realistic individuals who assess the situation more accurately, and hence will not survive evolutionary pressures. If success pays off in wealth, which translates to more supportable descendants and more imitators, one might have guessed that individuals whose estimations are not biased would perform best on average, and thus would outnumber the biased types in the long run.

In this paper we show that this intuition need not be right and in fact, there is a wide range of circumstances where optimistic individuals not only survive, but also prosper and take over the entire population. Our model is based on the consequences of biased perceptions in strategic interactions between individuals in a large population, who are continuously matched at random to interact. The individuals differ from one another in the way they perceive their payoffs from interacting with their rivals: optimists overes-

---

[1]In a similar vein, most cigarette smokers smoke high-tar brands, but only 17 percent believe their brand to have a more hazardous tar level than most others (Segerstrom et al, 1993), and sexually active undergraduate women, especially those who do not consistently use effective contraception, perceive their vulnerability to unwanted pregnancy as smaller than the other women at their university perceive this risk (Burger and Burns, 1988). In a study of optimism and behavior, Seligman and Schulman (1986) compared the sales made by new life insurance agents and found that those who put an optimistic spin on setbacks by seeing them as flukes or as suggesting a new approach rather than viewing them as signs of incompetence sold more policies during their first year and were half as likely to quit.

timate the positive impact of their own actions, pessimists underestimate it, and only realists assess it correctly. Biased individuals do not react optimally to their environment as shaped by the objective circumstances and the actions of others. However, it turns out that in many strategic settings, being recognized as optimistic may give individuals a strategic advantage, by inducing opponents to alter their behavior in a favorable way. When individuals misperceive the impact of their own actions on their payoffs, the resulting equilibrium behavior and true payoffs are different than those that would appear if the interacting parties were both realistic. Given a specific way the other party perceives her utility, it may very well "pay" to misperceive one's own utility, so that the resulting equilibrium would be better according to the true payoffs. For moderate levels of optimism, this beneficial effect outweighs the losses due to the biases in judgement, in a large class of interactions.

Interestingly, cautious optimism may pay off both in games in which the actions are strategic substitutes in the sense of Bulow, Geanakoplos and Klemperer (1985) as well as games in which the actions are strategic complements. The reason for this is that an optimist who overestimates the return to his actions, behaves more "aggressively" than a realist and chooses a higher level of action.[2] When actions are strategic substitutes and the actions of one individual impose a negative externality on the payoffs of rivals, this aggressive behavior triggers a favorable soft behavior from the rival. Examples for such interactions include the tragedy of the commons (the joint use of congested common resources) and some tournaments.[3] When the actions are strategic complements and the actions of one individual impose a positive externality on the payoff of rivals, the aggressive behavior of the optimist triggers a favorable aggressive behavior from rivals. Examples for this kind of interaction include the Bertrand duopoly model with differentiated products and the Cournot duopoly model with complementary products. Hence, in both cases, cautiously optimistic types fare better on average than realists. The reason why overly optimistic types do not do as well is that the strategic advantage from being "too" optimistic is outweighed by the associated loss from having a biased perception and hence from taking suboptimal actions.[4]

The idea that a biased objective function that does not coincide with the true payoff

---

[2]In this paper, we use the term aggressiveness to mean a higher level of action, not a tendency to hurt the other player.

[3]For instance, consider the Tournament model of Lazear and Rosen (1981), where two individuals 1 and 2, compete for a prize, $w$. Each of the individuals expands effort, $\mu_i$ $(i = 1, 2)$, to produce an output $q_i = \mu_i + \epsilon_i$, where $\epsilon_i$ is a random luck component. The individual with the higher output wins the prize. If the random luck components are independently drawn from the same exponential distribution, $F(\epsilon_i) = 1 - e^{-\lambda \epsilon_i}$, then the expected payoff of individual $i$ is $\pi_i = \frac{w e^{-\lambda(\mu_j - \mu_i)}}{2} - C(\mu_i)$, where $C(\mu_i)$ is the disutility of effort which is increasing and convex. It can be verified that this model, the efforts are strategic substitutes and the effort of one individual lowers the expected payoff of the other individual.

[4]For an alternative exploration of optimism and self-confidence see Benabou and Tirole (1999a,b) and Brocas and Carrillo (1999).

may confer a strategic advantage is well established and goes back at least to Schelling (1960). It was used extensively in many areas in economics including, Macroeconomics (Rogoff, 1985), International Trade (e.g. Brander and Spencer 1985, Eaton and Grossman 1986), Industrial Organization, (e.g., Brander and Lewis 1986, Fershtman and Judd 1987), and delegation (e.g. Green 1992, Fershtman, Judd and Kalai 1991, Fershtman and Kalai 1997, Katz 1991). In a similar vein, the literature on the evolution of preferences has shown that a population of "irrational" types who care about fairness (Güth and Yaari 1992, Huck and Oechssler 1998), are socially minded (Fershtman and Weiss 1997, 1998), altruistic (Bester and Güth 1998), spiteful (Possajennikov 2000, Bolle 2000), concerned with relative success (Koçkesen, Ok and Sethi 1998), or overconfident in financial investments (Kyle and Wang 1997, Benos 1998), may be evolutionary stable, i.e. immune to the appearance of few selfish or rational "mutants".[5]

In the current work, we take these ideas one step further. We show under what conditions optimism would evolve in *a full-fledged, dynamic evolutionary context.* That is, instead of just showing that optimism is evolutionary stable, we show under what conditions the distribution of individual types will converge over time to a unit mass on some level of optimism. Specifically, we consider a large population of individuals who are continuously matched in pairs at random and interact with one another. Individuals differ in the way they perceive the impact of their own actions on their payoffs, with optimists overestimating the impact and pessimists underestimating it. Over time, more successful types of individuals proliferate. This is either because they are more fit in the biological sense and therefore have higher reproduction rates, or because they are imitated by others. For a large class of interactions, we show that some degree of cautious optimism would reign the population in the long run and wipe out all other types.[6]

To prove our result, we posit an artificial, preliminary game, in which two players commit simultaneously to their degree of optimism or pessimism, knowing that consequently they will be playing the equilibrium of the game defined by the types to which they committed, but get the true payoffs that result from their behavior. When this artificial game is dominance solvable, a regular payoff-monotonic dynamics will wipe out all serially dominated types, and the population will converge in distribution to a unit mass at some level of biased perception. This result is known for a distribution of finitely many strategies or types (Samuelson and Zhang 1992), and is proved in the appendix

---

[5]Related ideas appear already in the works of Frank (1987, 1988). The indirect evolutionary approach, where the preferences rather than the strategies are the subject of evolutionary pressures, is employed also by Dekel and Scotchmer (1999), Dufwenberg and Güth (1999), Rogers (1994), Robson (1996a,b), Waldman (1994) Vega-Redondo (1997) and Bergman and Bergman (2000). See also further references in the sequel.

[6]In a similar dynamic setting, Huck, Kirchsteiger and Oechssler (1997) examine the emergence of an endowment effect – an excess valuation of one's own endowment in bargaining. They show that the proportion of realists with no such effect will shrink to zero with time, as will types with a very high endowment effect. However, unlike in our case, the dynamics is not shown to converge in the long run.

for general distributions when the payoff function is continuous and the set of types is a compact interval.

The paper is organized as follows. Section 2 brings a simple example that exhibits most of the properties of the model. Section 3 discusses several interpretational issues. Section 4 explores general conditions on the payoff functions which are sufficient for our results to hold, and elaborates on the dynamic selection process and its properties. Section 5 concludes.

# 2   A Simple Example

Consider a large population of individuals who are continuously and randomly matched in pairs to interact with one another. In this interaction, the matched individuals $i = 1, 2$ choose actions $x^i \in \mathfrak{R}$. These actions are to be interpreted as the degree of effort or the level of investment the individuals put into the interaction.[7] Given a pair of actions $x^1, x^2$, the payoffs of the individuals are

$$\Pi^i(x^i, x^j) = (\alpha - bx^j - x^i)x^i \quad \text{for } i = 1, 2 \ \text{ and } \ j \neq i, \tag{2.1}$$

where $\alpha > 0$ and $-1 < b < 1$. When $b$ is positive, the individuals are competing with each other: The larger the action of the other party, the lower is the return to every unit of one's own action. Moreover, since $\Pi^i_{ij} = b > 0$ (subscripts are used to denote partial derivatives), the best-response functions are decreasing in the $(x^1, x^2)$ space, so the actions are strategic substitutes in the sense of Bulow, Geanakoplos and Klemperer (1985). With a negative $b$, the individuals cooperate with one another since now there is a positive, linear correlation between the return to one's action and the action of the other individual. And, since $\Pi^i_{ij} = b < 0$, the best-response functions are increasing in the $(x^1, x^2)$ space, so the actions are strategic complements.

Although the payoffs of all individuals are symmetric, individuals differ from one another in the way they perceive the interaction between them: Pessimistic types underestimate the value of $\alpha$, optimistic types overestimate it, and only realistic types assess it correctly. Specifically, player $i$ conceives the value of $\alpha$ to be

$$\alpha^i = \alpha + \tau^i, \tag{2.2}$$

where $\tau^i$ is positive (zero, negative) when the individual is optimistic (realistic, pessimistic) regarding the return to her action for any given action of the other individual. Thus, the individuals conceive their utility functions to be

---

[7]For some interpretations, it may be suitable to consider only non-negative actions. Our arguments continue to hold with such a restriction, though the analysis gets more involved.

$$U^i(x^i, x^j) = (\alpha^i - bx^j - x^i)x^i \quad \text{for } i = 1, 2 \text{ and } j \neq i \tag{2.3}$$

Let us assume that the players play the unique Nash equilibrium of the game with these utility functions:

$$\widehat{x}^i = \frac{(2 - b)\,\alpha + 2\tau^i - b\tau^j}{4 - b^2} \quad \text{for } i = 1, 2 \text{ and } j \neq i. \tag{2.4}$$

An important assumption here is that in every pairwise meeting, the players know each other's type. This is either because the types $\tau^1$, $\tau^2$ are "written on the players' foreheads" and are thus immediately recognized, or alternatively because each interaction consists of several rounds, in which the players' actions converge relatively quickly to the equilibrium behavior (e.g. because in those rounds both players play their best response to the other's previous action or some average of the other's previous actions).[8]

Substituting $\widehat{x}^1$ and $\widehat{x}^2$ into (2.1), the true equilibrium payoffs of the players are given by

$$f^i(\tau^i, \tau^j) \equiv \Pi^i(\widehat{x}^i, \widehat{x}^j) = \frac{((2 - b)\,\alpha + 2\tau^i - b\tau^j)\,((2 - b)\,\alpha - (2 - b^2)\tau^i - b\tau^j)}{(4 - b^2)^2}. \tag{2.5}$$

Now imagine that these true payoffs translate into fitness terms, so that the instantaneous growth rates of the types are monotonic in their average true payoffs when interacting with randomly matched individuals from the current population of types. That is, the proportion of types with high current average payoffs tends to increase, at the expense of types with low current average payoffs. The evolution of types therefore follows a regular, payoff monotonic dynamics, as shall be formally defined in section 5. The mechanism by which the frequency of types evolves over time can be seen as either purely biological (types with higher payoffs have a higher ability to reproduce), as a process by which parents transmit their attitudes to life (i.e. their degree of "optimism") to their children via education or parental influence, or as a process by which more successful attitudes to life are imitated more often and increase in popularity.

Given this evolution of attitudes, which levels of optimism will perform best and survive in the long run? To provide an answer, we consider a preliminary, artificial two-player "types game". In this game, each of the players $i = 1, 2$ chooses a type $\tau^i \geq -\alpha$,

---

[8]If types were not mutually observed, individuals would not be able to gain a strategic advantage from being recognized as optimistic or aggressive. Individuals would then choose their actions to maximize their expected payoffs in a random match with an individual from the population, and realistic types would fare best. Consequently, as shown by Dekel, Ely and Yilankaya (1998) and Ok and Vega-Redondo (1999) (see also an example in Possajenikov 1999), in an evolutionary stable distribution of the population, all the individuals play a (true) Nash equilibrium strategy.

or, equivalently, an assessment $\alpha^i = \alpha + \tau^i$ which is non-negative. The payoffs $f^i(\tau^i, \tau^j)$ as a function of these types are given by (2.5) above.

The best-response functions in this game are

$$BR^i(\tau^j) = \frac{b^2\left((2-b)\alpha - b\tau^j\right)}{4(2-b^2)} \quad \text{for } i = 1, 2 \text{ and } j \neq i. \tag{2.6}$$

These functions are downward sloping in the $(\tau^1, \tau^2)$ space when $b > 0$, and upward sloping when $b < 0$. In other words, the types are strategic substitutes (complements) in this artificial game whenever the actions are strategic substitutes (complements) in the original game itself. Moreover, since $|b| < 1$, the slope of the best-response functions $BR^i$ is less than $\frac{1}{4}$ in absolute value (each as a function of its variable). Therefore, they have a unique intersection in the $(\tau^1, \tau^2)$ space at $(\hat{\tau}^1, \hat{\tau}^2)$, where

$$\hat{\tau} \equiv \hat{\tau}^1 = \hat{\tau}^2 = \frac{b^2}{4 + 2b - b^2}\alpha. \tag{2.7}$$

are the unique Nash equilibrium assessments. The following graph depicts $\frac{\hat{\tau}}{\alpha} = \frac{b^2}{4+2b-b^2}$. As can be seen, the players will stick with an optimistic assessment ($\frac{\hat{\tau}}{\alpha} > 0$) for any $b \neq 0$. When $b = 1$ (the case of competition and strategic substitutes), the degree of optimism will be 20%. When $b = -1$ (the case of cooperation and strategic complements), the degree of optimism will reach 100%. Only when $b = 0$, where the individuals do not face a genuine strategic interaction but rather a one-person decision problem, no optimism will appear.
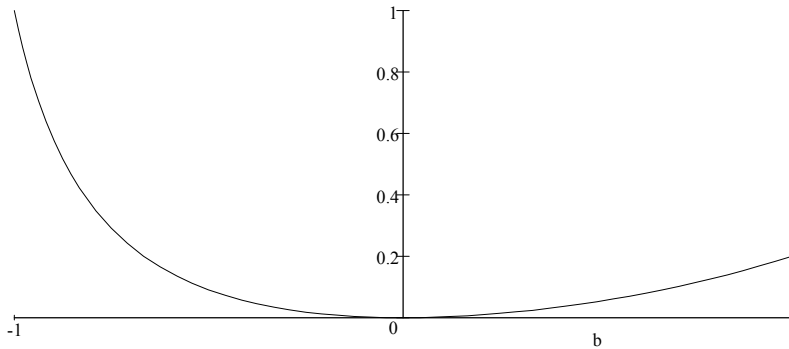


Fig.1: The optimism factor $\frac{\hat{\tau}}{\alpha}$ at equilibrium as a function of $b$

Since the slopes of the best-response functions $BR^i$ are smaller than 1 in absolute value, this artificial "types game" can be also solved by a process of iterative elimination of strictly dominated strategies, in which every type but $\hat{\tau}$ will eventually be eliminated.[9] By theorem 1 below (which is a generalization of a result of Samuelson and Zhang (1992)),

---

[9]For instance, suppose that $b = 1$, so that the payoff functions (2.5) in the "types game" are

the serially dominated types (those that do not survive iterative elimination of strictly dominated strategies) are wiped out in a regular, payoff monotonic dynamics. Precisely, for any initial distribution of types whose support is an interval that contains $\widehat{\tau}$, the distribution of types will converge in distribution to the unit mass at $\widehat{\tau}$, so that the density of all other types will converge to zero. Figure 1 above depicts, therefore, the relative optimism of the surviving type as a function of $b$.

The intuition underlying the evolution of optimism is as follows. Individuals with optimistic types play more aggressively than realists and choose larger actions. When $b$ is positive, the actions are strategic substitutes, so the aggressive behavior of optimists invites a softer behavior from opponents. Since the interactions are competitive when $b$ is positive, this soft behavior from opponents is beneficial, so optimists end up with a higher payoff than realists or pessimists, and hence end up with a higher payoff. Of course, the return to aggression has its limits, as overly aggressive strategies harm not only the opponent but also the aggressor. Hence, wildly optimistic types do not do as well as more moderately optimistic types. When $b$ is negative, actions are strategic complements, so typically there is a free-rider problem: The players fail to take into account the benefit of their actions for their opponents, and therefore the equilibrium actions tend to be "too low". Since optimism induces players to be more aggressive than they would be otherwise, it invites a similar reaction from the opponents. And since when $b$ is negative the interactions between individuals are cooperative, this eventually benefits both players. Once again, wildly optimistic types end up doing "too much", while realistic or pessimistic types do "too little". Hence, cautious optimists fare better on average and gradually take over the population.

## 3   Discussion

Before we continue with more general results, we sidestep to discuss several interpretational issues of the model.

---

$f^i(\tau^i, \tau^j) = \frac{(\alpha + 2\tau^i - \tau^j)(\alpha - \tau^i - \tau^j)}{9}$, and the best-response functions (2.6) are $BR^i(\tau^j) = \frac{\alpha - \tau^j}{4}$. Since player 1's assessment of $\alpha$ is non-negative, i.e. $\tau^1 \geq -\alpha$, it is better for player 2 to be of type $\tau^2 = \frac{\alpha}{2}$ and assess $\alpha^2 = \frac{3}{2}\alpha$ than to have any higher assessment. Understanding this, having type $\tau^1 = \frac{\alpha}{8}$ and assessing $\alpha^1 = \frac{9}{8}\alpha$ is better for player 1 than having any lower assessment. But with this in mind, player 2 is better off having the type $\tau^2 = \frac{7}{32}\alpha$ and the assessment $\alpha^2 = \frac{39}{32}\alpha$ than any higher value, and so forth. The only type that survives this elimination process is $\frac{\alpha}{5}$, the Nash equilibrium type $\widehat{\tau}$.

## 3.1 Learning about $\alpha$ over time

The above example immediately raises the following question: How come that optimists do not come to realize that they overestimate $\alpha$ once they observe their true payoffs (2.5)? A persistent overestimation of $\alpha$ may be compatible with what is called in the psychology literature the *belief perseverance phenomenon,* which is the tendency to cling to one's beliefs in the face of contrary evidence.[10]

Another possible interpretation is that individuals face uncertainty regarding the true current value of $\alpha$, which is drawn at random from some distribution. To illustrate, suppose that each period, $\alpha$ could either assume a high value, $\alpha_h$, with probability $p$, or a low value, $\alpha_\ell$, with probability $1 - p$, independently across periods. Individuals do not know the value of $p$, but can observe past realizations of $\alpha$ and use them to update their beliefs about $p$. Using $p'$ to denote the posterior estimate of $p$, and noting that the payoff functions (2.1) are linear in $\alpha$, substituting $\alpha' = p'\alpha_h + (1 - p')\alpha_\ell$ in place of $\alpha$ in (2.1) gives the expected payoff functions.

Initially, individuals have in mind a prior distribution over values of $p$, that contains the true value $p$ in its support. Realists update their beliefs about $p$ using Bayes rule, so by the strong law of large numbers, their posterior distribution will converge in distribution to a point mass on $p$ almost surely. As a result, their $\alpha'$ will converge almost surely to the true mean $p\alpha_h + (1 - p)\alpha_\ell$, and they will get closer and closer to maximizing their true expected payoff.

In contrast, optimists (pessimists) do not use all the relevant information they have to update their beliefs about $p$. Rather, they discard some low (high) realizations of $\alpha$, attributing them the exceptional, non-systematic bad (good) circumstances, which render them irrelevant for updating. In other words, the behavior of optimists and pessimists exhibits what is called in the psychology literature a *confirmation bias* – the tendency to seek information that confirms one's own views and overlook evidence that may disconfirm these views. For instance, reflecting on many experiments, Wason (1981) reports that once people have a wrong idea they "...evade facts, become inconsistent, or systematically defend themselves against the threat of new information relevant to the

---

[10]For instance, Lord, Ross and Lepper (1979) found that people who held opposing views on capital punishments and were shown supposedly new research findings, one supporting and the other refuting the claim that death penalty deters crime, were more impressed by the study that supported their beliefs, and readily disputed the other study. Anderson, Lepper and Ross (1980) found that people who were told about a risk-taker who was an excellent fire fighter and a cautious person who was a poor fire fighter, surmised that risk-takers tend to be better fire fighters, while other people who were told the opposite concluded that cautious people are better fire fighters. These beliefs did not change by much even after the researchers revealed that the cases were made up for the experiment, arguably because the participants held on to their explanations for why the their beliefs made sense, even though the evidence was gone.

issue." For experiments that demonstrate the confirmation bias phenomenon, see e.g. Klayman and Ha (1987) and Skov and Sherman (1986).

Assuming that optimistic (pessimistic) discard each some constant proportion of low (high) realizations of $\alpha$, their posterior estimation of $p$ will converge almost surely to some point mass above (below) $p$. As a result, $\alpha'$ in the perceived expected payoffs of optimistic types will converge almost surely to a value above (below) the true average of $\alpha$. The higher the percentage of discarded realizations, the further $\alpha'$ will get from the true average.

Our results show that the confirmation bias of optimistic types can not only survive evolutionary pressures but also take over the entire population. Using a learning interpretation of selection dynamics, this suggests in turn that there is no reason to believe that over time, individuals will learn to update their beliefs in a Bayesian fashion.

## 3.2   Larger families of perceived utility functions

We have thus far considered a one-dimensional family of distortions in the perception of one's utility function - those that result from different evaluations of the impact of one's own actions on payoffs. What would happen if we were to consider *every* possible distortion of the utility function?

Dekel, Ely and Yilankaya (1998) show that if the distribution of types (i.e. the perceived payoffs in our case) ever reaches a stable state in which all the individuals take the same action, then this action must be efficient. In the example of the previous section, this action is the joint payoff maximizing action $\frac{\alpha}{2(1+b)}$. A population of types who all take a different action is not immune to an invasion of mutants who perceive their payoffs differently. To see why, assume for simplicity that all the individuals perceive the payoffs in the same way, and suppose, by way of contradiction, that when they meet each other they play some inefficient $x \neq \frac{\alpha}{2(1+b)}$. Consider mutants that differ from the rest of the population only in that they are extremely satisfied (over and beyond the true payoff) if both players in the interaction play $\frac{\alpha}{2(1+b)}$. Thus, when a mutant meets an incumbent, she will mimic the incumbent and play $x$, so that both the incumbent and the mutant will get the same true payoff. But when a mutant meets another mutant they will recognize each other, play $\frac{\alpha}{2(1+b)}$ and fare better. Hence, such mutants would not tend to disappear.

This result, however, is static in nature. In particular, it does not predict if and when a payoff monotonic dynamics would ever converge. Moreover, it does not preclude stable states with a polymorphic distribution of types. But in any case, it does imply that our result need not hold with an extended family of utility distortions, and it leads to ask

10

what biases are *relevant* for consideration and in which contexts. Thus, it will certainly be of interest to model more families of biases, and confront the theoretical predictions with empirical findings. The results of Dekel et al. (1998) show that the triumph of biased types is not a trivial result in such an exercise, and thus open the door for a challenging process of modeling relevant biases.

# 4  A General Analysis

In section 2 we analyzed an example of a symmetric two-player game, in which individual $i$ of type $\tau^i$ misperceived her true payoff function $\Pi^i(x^i, x^j)$ to be

$$U^i(x^i, x^j) = \Pi^i(x^i, x^j) + \tau^i x^i, \quad i = 1, 2, \tag{4.1}$$

where $x^i, x^j$ are the actions of individual $i$ and her opponent $j$, respectively. Optimistic types had a positive $\tau^i$, and thus overestimated the return to each unit of their action, while the converse was the case for pessimistic types. In this section we shall present a much broader class of payoff functions $\Pi^i$, for which some degree of optimism emerges in a dynamic selection process that favors successful biases. The properties of the payoff functions $\Pi^i$ in this class and the induced properties of the "types game" will be detailed in sections 4.1 and 4.2, respectively. In Section 4.3 we bring the precise definition of the selection dynamics and analyse conditions for its convergence to a unique selected type.

## 4.1  The pairwise interactions

We begin by specifying properties of the true payoff functions $\Pi^i \equiv \Pi(x^i, x^j)$ of individual $i$, which we assume to be thrice continuously differentiable. For convenience, we shall often denote its partial derivatives by $\Pi^i_i \equiv \frac{\partial \Pi(x^i, x^j)}{\partial x^i}$, $\Pi^i_j \equiv \frac{\partial \Pi(x^i, x^j)}{\partial x^j}$, $\Pi^i_{ii} \equiv \frac{\partial^2 \Pi^i(x^i, x^j)}{(\partial x^i)^2}$, and $\Pi^i_{ij} \equiv \frac{\partial^2 \Pi^i(x^i, x^j)}{\partial x^i \partial x^j}$.

**Property 1:**  The true payoff function of individual $i$ is strictly concave in $x^i$, i.e., $\Pi^i_{ii} < 0$.

The concavity of $\Pi^i$ in $x^i$ ensures that the best-response of individual $i$ against individual $j$'s action, $BR^i(x^j)$, is implicitly defined by the first order condition

$$U^i_i(x^i, x^j) = \Pi^i_i + \tau^i = 0. \tag{4.2}$$

**Property 2:**  The true payoff function of individual $i$ is such that $\left| \frac{\Pi^i_{ij}}{\Pi^i_{ii}} \right| < 1 - \varepsilon$, for some $\varepsilon > 0$.

11

Property 2 ensures that the slope $\frac{\mathrm{d}BR^i(x^j)}{\mathrm{d}x^j}$ of each individual's best-response function is uniformly smaller than 1 in absolute value. This implies in turn that every pair of best-response functions intersect exactly once in the actions space, so that every pairwise interaction has a unique Nash equilibrium.[11]

Let $(\widehat{x}^i(\tau^i, \tau^j), \widehat{x}^j(\tau^i, \tau^j))$ be the Nash equilibrium in the interaction between individuals $i$ and $j$ given their types. The equilibrium strategies are implicitly defined by the following system of equations:

$$\begin{aligned} \Pi_i(\widehat{x}^i(\tau^i, \tau^j), \widehat{x}^j(\tau^i, \tau^j)) + \tau^i &= 0, \\ \Pi_j(\widehat{x}^i(\tau^i, \tau^j), \widehat{x}^j(\tau^i, \tau^j)) + \tau^j &= 0. \end{aligned} \tag{4.3}$$

## 4.2  The artificial types game

To study the evolution of optimism we once again consider an artificial two-player "types game" in which each player $i$ selects a type $\tau^i$. We assume that the set of possible types for each individual is a (compact) interval $T = [\underline{\tau}, \overline{\tau}]$, where $\underline{\tau} < 0 < \overline{\tau}$, so that both pessimists and optimists are represented in the population. The payoff of player $i$ in the "types game" is given by

$$f^i(\tau^i, \tau^j) \equiv \Pi(\widehat{x}^i(\tau^i, \tau^j), \widehat{x}^j(\tau^i, \tau^j)), \tag{4.4}$$

which is the true equilibrium payoff in the interaction between $i$ and $j$. The uniqueness of the Nash equilibrium in every pairwise interaction implies that the (reduced-form) payoff function of each individual in the types game, $f^i(\tau^i, \tau^j)$, is well-defined. Moreover, since all individuals have the same true payoff function, the payoffs in the types game are symmetric in the sense that $f^i(\tau^i, \tau^j) = f^j(\tau^j, \tau^i)$. The needed properties of these payoff functions are:

**Property 3:** The payoff function $f^i$ of each player $i$ in the artificial "types game" is twice differentiable and strictly concave in $\tau^i$

**Property 4:** The payoff functions of individuals $i$ and $j$ in the artificial "types game" are such that $\left| f^i_{ii} \right| < \left| f^i_{ij} \right|$.

Property 3 ensures that the types game is well-behaved. Property 4 ensures that the types game has a unique Nash equilibrium because it implies that the slope of each

---

[11]The uniformity requirement is needed to guarantee that the best-respone functions of the individuals are not mutually asymptotic without intersecting each other. Actually, property 2 ensures that a myopic best-reply process in a repetition of the game (in which each individual plays a best reply to the previous action of the opponent) converges to the unique Nash equilibrium. This may justify the assumption that the individuals essentially play this equilibrium even if initially they do not recognize each other's type, but each encounter between them consists of several rounds.

player's best-response function in the types game is less than 1 in absolute value. Both properties are satisfied in the example studied in Section 2. Since the payoff functions in the types game are symmetric, the Nash equilibrium in the types game is also symmetric. We denote the equilibrium types by $\widehat{\tau}$, so that the Nash equilibrium is $(\widehat{\tau}, \widehat{\tau})$.

To study the properties of the "types game" further, we return for a moment to the underlying pairwise interactions and differentiate (4.3) to obtain the following comparative statics conditions:

$$J \times \left( \begin{array}{c} \widehat{x}_i^i \\ \widehat{x}_i^j \end{array} \right) = \left( \begin{array}{c} -1 \\ 0 \end{array} \right), \quad J \times \left( \begin{array}{c} \widehat{x}_j^i \\ \widehat{x}_j^j \end{array} \right) = \left( \begin{array}{c} 0 \\ -1 \end{array} \right), \tag{4.5}$$

where

$$J = \left( \begin{array}{cc} \Pi_{ii}^i & \Pi_{ij}^i \\ \Pi_{ji}^j & \Pi_{jj}^j \end{array} \right). \tag{4.6}$$

Using Cramer's rule, this yield

$$\widehat{x}_i^i = -\frac{\Pi_{jj}^j}{\det J}, \quad \widehat{x}_j^i = \frac{\Pi_{ji}^i}{\det J}. \tag{4.7}$$

Property 2 implies $\det J > 0$. Hence, From (4.7) and property 1 it is easy to see that as $\tau^i$ increases (individual $i$ becomes more optimistic), the equilibrium action of individual $i$, $\widehat{x}^i$, increases (individual $i$ becomes more aggressive), while the equilibrium action of her rival, $\widehat{x}^j$, increases if $\Pi_{ij}^j > 0$ (actions are strategic complements) and decreases if $\Pi_{ij}^j < 0$ (actions are strategic substitutes).

Given Property 3 and using (4.7), an interior Nash equilibrium in the "types game" is defined implicitly by the following system of equations:

$$f_i^i(\tau^i, \tau^j) = \frac{1}{\det J} \left( -\Pi_i^i \Pi_{jj}^j + \Pi_j^i \Pi_{ji}^j \right) = 0, \tag{4.8}$$

$$f_j^j(\tau^i, \tau^j) = \frac{1}{\det J} \left( -\Pi_j^j \Pi_{ii}^i + \Pi_i^j \Pi_{ij}^i \right) = 0,$$

where the partial derivative of the actions game are evaluated at the Nash equilibrium actions $(\widehat{x}^i(\tau^i, \tau^j), \widehat{x}^j(\tau^i, \tau^j))$. To interpret the equilibrium conditions in the types game, note that (4.8) implies that at an interior Nash equilibrium of the "types" game we have

$$-\frac{\Pi_i^i}{\Pi_j^i} = -\frac{\Pi_{ji}^j}{\Pi_{jj}^j}. \tag{4.9}$$

The left side of (4.9) represents the slope of an iso-payoff curve of individual $i$ in the underlying actions game in the $(x^i, x^j)$ space, and the right side of (4.9) represents the

slope of the best response function of individual $j$ in the $(x^i, x^j)$ space. Thus, equation (4.9) says that individual $i$ chooses $\tau^i$ optimally in the types game by selecting to be on the "highest" true iso-payoff curve, taking as given the best response function of the rival in the action game. The first order condition for this constrained maximization problem requires that, holding $\tau^j$ constant, the iso-payoff curve of individual $i$ will be tangent to the best response function of individual $j$.[12]

In order to express Properties 3 and 4 directly in terms of the true payoff function $\Pi$, one has to differentiate further (4.8), using (4.7). The resulting complicated expressions are strict inequalities which involve third-order derivatives of $\Pi$, and they do not have a direct economic interpretation.

Next, we show when it is the case that at the symmetric Nash equilibrium of the types game, $\widehat{\tau} > 0$. That is, in the equilibrium of the types game, individuals "choose" to become optimistic.

**Lemma 1** *The Nash equilibrium in the artificial "types game" is such that $\widehat{\tau} > 0$ if and only if $\Pi_j^i$ and $\Pi_{ij}^i$ have the same sign, i.e. $\Pi_j^i \Pi_{ij}^i > 0$.*

**Proof.** Substituting from (4.9) into (4.3) and rearranging terms yields

$$\widehat{\tau} = -\frac{\Pi_j^i \Pi_{ji}^j}{\Pi_{jj}^j}, \tag{4.10}$$

where the right hand side is evaluated at the symmetric Nash equilibrium $(\widehat{x}(\widehat{\tau}, \widehat{\tau}), \widehat{x}(\widehat{\tau}, \widehat{\tau}))$ in the interaction between two individuals who both have the type $\widehat{\tau}$. By the symmetry of the actions game, it is also the case that

$$\widehat{\tau} = -\frac{\Pi_j^i \Pi_{ij}^i}{\Pi_{jj}^j}, \tag{4.11}$$

at $(\widehat{x}(\widehat{\tau}, \widehat{\tau}), \widehat{x}(\widehat{\tau}, \widehat{\tau}))$. The denominator of (4.16) is negative by Property 1. Therefore, $\widehat{\tau} > 0$ if and only if $\Pi_j^i \Pi_{ij}^i > 0$. ∎

To interpret Lemma 1, note that $\Pi_j^i$ captures the externality that individual $j$'s action imposes on individual $i$'s payoff in a pairwise interaction between them (the externality is positive if $\Pi_j^i > 0$, and negative if $\Pi_j^i < 0$), and the sign of $\Pi_{ij}^i$ determines whether the actions are strategic substitutes ($\Pi_{ij}^i < 0$) or strategic of complements ($\Pi_{ij}^i > 0$). Hence

---

[12]This is exactly like the behavior of a leader in a Stackelberg duopoly model who chooses the level of output at which its iso-profit curve is tangent to the best response function of the follower.

Lemma 1 shows that in a Nash equilibrium of the types game, players will choose to become optimistic either if they impose negative externalities on one another and their actions are strategic substitutes or if they impose positive externalities on one another and their actions are strategic complements. The feature that $\Pi^i_j$ and $\Pi^i_{ij}$ have the same sign is common in many types of games, including the Cournot duopoly with homogenous or differentiated products, the tragedy of the commons, and tournaments (see footnote 3), where $\Pi^i_j$ and $\Pi^i_{ij}$ are both negative, and the Bertrand duopoly model with differentiated products or the Cournot duopoly model with complementary products, where $\Pi^i_j$ and $\Pi^i_{ij}$ are both positive.[13] As we explained above, in either case, the aggressive behavior associated with being an optimist gives individuals a strategic advantage; not surprisingly then, in a Nash equilibrium of the "types game", both players will choose to become optimistic.

**Remark:** In what follows, we will show that properties 1-4 and the property $\Pi^i_j \Pi^i_{ij} > 0$ in Lemma 1 guarantee the dynamic evolutionary emergence of optimism, in a sense that we shall make precise. The payoff functions $\Pi$ which satisfy these properties constitute an open set in the space $C^3(R^2_+)$ of thrice continuously differentiable functions[14], because they are defined using finitely many strict inequalities which involve continuous functions of up to third-order derivatives of $\Pi$.[15] Thus, the family of quadratic payoff functions explored in section 2, which clearly obey all these properties, is not exceptional: Other payoff functions whose derivatives up to the third-order are not too far from one of those quadratic functions admit the same kind of evolution of preferences towards some level of optimism.

The following lemma will be useful for the sequel:

**Lemma 2** *The artificial "types game" is strictly dominance solvable. The unique outcome that survives the process of iterated elimination of strictly dominated strategies is the symmetric Nash equilibrium $(\widehat{\tau}, \widehat{\tau})$, where $\widehat{\tau}$ is defined implicitly by the first order condition, $f^i_i(\widehat{\tau}, \widehat{\tau}) \equiv 0$.*

---

[13]Fershtman and Weiss (1998) show that the same condition implies that social mindedness (enjoying doing what is socialy highly considered) is evolutionary stable. It should be noted however that there are important classes of games in which $\Pi^i_j$ and $\Pi^i_{ij}$ do not have the same sign. For instance, in arms races, where $\Pi^i = V(x^i - x^j) - C(x^i)$, with $V'(.) > 0 > V''(.)$, $C'(.) > 0$, and $C''(.) \geq 0$, we have $\Pi^i_j < 0 < \Pi^i_{ij}$. Likewise, in voluntary contribution to public goods games, where $\Pi^i = V(x^i + x^j) - C(x^i)$, with $V'(.) > 0 > V''(.)$, and $C'(.) > 0$ and $C''(.) \geq 0$, we have $\Pi^i_j > 0 > \Pi^i_{ij}$.

[14]$C^3(R^2_+)$ is the space of thrice continuously differentiable functions $\Pi : R^2_+ \to R$, with the minimal topology in which $\Pi_n$ converges to $\Pi$ iff $(\Pi_n - \Pi)$ and each of its first, second and third derivatives converge to zero uniformly on compact sets on $R^2_+$.

[15]More precisely, for every fixed $\varepsilon > 0$ in property 2 there are finitely many strict inequalities involved in the definition of this set of payoff functions – denote it $V_\varepsilon$ – which is therefore open. The actual set of payoff functions $\Pi$ which obey all the properties is the union of these open $V_\varepsilon$ over all the positive $\varepsilon$, which is open as the union of open sets.

**Proof.** To prove the lemma we invoke Theorem 4 in Moulin (1984) that provides sufficient conditions for normal form games to be dominance solvable. In the present context, these conditions are:

(i) The strategy set of each player is a one-dimensional compact interval

(ii) The payoff function of each player is continuous over the space of outcomes, twice differentiable, and strictly concave with respect to the player's strategy

(iii) The slope of each players' best-response functions is less than 1 in absolute value

Condition (i) is satisfied in the types game because of the assumption that the set of possible types for each player is a (compact) interval $T = [\underline{\tau}, \overline{\tau}]$. Properties 3 and 4 ensure that conditions (ii) and (iii) are satisfied. Hence, the types game is dominance solvable. ∎

## 4.3   The selection dynamics

In this subsection we turn to the way the population of types evolves over time. To this end, let $G_t$ be the distribution of types in the population at time $t \geq 0$ on the support $T = [\underline{\tau}, \overline{\tau}]$. We assume that $G_t$ evolves according to a payoff monotonic selection dynamics, where types with higher average payoffs have higher growth rates. Specifically, we shall define the growth rate of types as follows:

**Definition.** A continuous growth-rate function $g : T \times \Delta(T) \to R$ is payoff monotonic and regular if for every $G \in \Delta(T)$, the following holds:

(i) Higher payoffs are associated with higher growth rates:

$$g(\tau^i, G) > g(\tilde{\tau}^i, G) \quad \Longleftrightarrow \quad f(\tau^i, G) > f(\tilde{\tau}^i, G). \quad (4.12)$$

(ii) The total size of the population is preserved:

$$\int_T g(\cdot, G) \mathrm{d}G(\cdot) = 0. \quad (4.13)$$

16

(iii) $g$ can be extended to the domain $T \times X$, where $X$ is the set of signed measures $G$ with variational norm smaller than 2, and on this extended domain, $g$ is bounded and Lipschitz continuous:

$$\sup_{\tau^i \in T, G_t \in X} \left| g(\tau^i, G) \right| < \infty, \qquad (4.14)$$

$$\sup_{\tau^i \in T} \left| g(\tau^i, G) - g(\tau^i, \widetilde{G}) \right| < K \left\| G_t - \widetilde{G} \right\|, \qquad G, \widetilde{G} \in X,$$

for some constant $K$, where $\|G\| = \sup_{|h| \leq 1} \left| \int_T h \, dG \right|$ is the variational norm on signed measures.

Oechssler and Riedel (1999, Lemma 3) proved that property (iii) guarantees that the mapping $G \to \int_T g(\cdot, G) dG$ is bounded and Lipschitz continuous in the variational norm, which implies that the differential equation in the space of distributions $\Delta(T)$ defined by

$$\dot{G}_t(S) = \int_S g(\cdot, G_t) dG_t(\cdot), \qquad S \subseteq T, \qquad (4.15)$$

has a unique solution for any initial distribution $G_0$. A special case of the growth rate that we consider is the familiar replicator dynamics that was introduced by Taylor and Jonker (1978) for distributions with a finite support, and by Oechssler and Riedel (1999) for general distributions. In the case of the replicator dynamics, the distribution of types at time $t$, $G_t$, evolves according to the differential equation

$$\dot{G}_t(S) = \int_S \left[ f(\tau^i, G_t) - f(G_t, G_t) \right] dG_t(\tau^i), \qquad S \subseteq T, \qquad (4.16)$$

where

$$f(\tau^i, G_t) \equiv \int_T f^i(\tau^i, \tau^j) dG_t(\tau^j) \qquad (4.17)$$

is the expected true equilibrium payoff of an individual of type $\tau^i$ at time $t$ from an interaction with individual $j$ drawn at random from the population, and

$$f(G_t, G_t) \equiv \int_T \int_T f^i(\tau^i, \tau^j) dG_t(\tau^j) dG_t(\tau^i) \qquad (4.18)$$

is the expected true equilibrium payoffs when both individuals are drawn at random from the population at time $t$. That is, if the average performance of a subset of types $S \subseteq T$ is better than the average performance in the population, the relative weight in the population of the types in $S$ increases, at the expense of other sets of types whose performance is below the average. The more general selection dynamics that we consider

17

may be appropriate when the reproduction process of types is not purely biological, but rather relies on education or imitation (see e.g., Weibull 1995, Section 4.4).

Having defined the selection dynamics, we are now interested in the following question: starting from some initial distribution, $G_0$, how will the distribution of types, $G_t$, evolve over time with a regular, payoff monotonic dynamics? To provide an answer, we now establish the following theorem (the proof of the theorem appears in the Appendix). The theorem, which is of independent interest, generalizes Theorem 1 in Samuelson and Zhang (1992) to the case of games with infinitely many strategies (Samuelson and Zhang (1992) prove their result for the case of games with finitely many strategies). For preserving the coherence with our setting, we state the theorem for symmetric two-players games with a compact one-dimensional strategy space; the method of proof works however just as well for more general compact strategy spaces and for asymmetric games.

**Theorem 1** *Let $T = [\underline{\tau}, \overline{\tau}] \subseteq R$ be a space of strategies, $f : T \times T \to R$ a continuous payoff function of a symmetric two-player game, and $g : T \times T \to R$ a regular, payoff monotonic growth-rate function. Let $G_t$ be the population dynamics defined by the differential equation (4.15) with an initial distribution of strategies $G_0$ with support $T$. Suppose that $D \subseteq [\underline{\tau}, \overline{\tau}]$ is the subset of serially dominated strategies (those that do not survive the process of iterated elimination of strictly dominated strategies). Then the strategies in $D$ are asymptotically eliminated from the population: Every iteratively dominated strategy $d \in D$ has an open neighborhood $W_d$ for which $\lim_{t \to \infty} G_t(W_d) = 0$. In particular, if there is only one non-eliminated strategy $u \in T \setminus D$, then $G_t$ converges in distribution to the unit mass probability at $u$.*

We are now ready to state our main result:

**Theorem 2** *Suppose that the payoffs in the pairwise interactions have properties 1-4 and are such that $\Pi_j^i$ and $\Pi_{ij}^i$ have the same sign. Then given any initial distribution of types with support $T$, the population of types will converge in distribution to a unit mass on the optimistic type $\widehat{\tau}$ under any regular, payoff-monotonic selection dynamics.*

**Proof.** ¿From Lemma 2 we know that given properties 1-4, the artificial "types game" is strictly dominance solvable, with the solution being $(\widehat{\tau}, \widehat{\tau})$. Using Theorem 1 it therefore follows that under a regular, payoff monotonic growth-rate function, the population of types will converge in distribution to a unit mass on $\widehat{\tau}$. Finally, Lemma 1 ensures that if $\Pi_j^i$ and $\Pi_{ij}^i$ have the same sign, then $\widehat{\tau} > 0$. ∎

18

# 5    Conclusion

We have shown how the pressures of explicit, dynamic evolutionary processes select for moderate optimism rather than for realism, when fitness is gained through interactions of either a competitive nature and strategic substitutes or cooperative nature and strategic complements. According to this insight, the well-documented phenomenon of overconfidence and unrealistic high self-esteem of individuals may be due to a bias that "pays" well in many kinds of *strategic* settings.

Clearly, the way humans evaluate their environment has evolved along the generations via conflicts with both natural hazards and strategic social interactions with other individuals or groups of individuals. The premises of our model are therefore far from being all-encompassing. And in practice, society is composed of heterogeneous individuals who may differ from one another in their degree of optimism/pessimism, unlike the long-run equilibrium in our model where all individuals share the same attitude. Thus, our modest aim was to point at one possible source for the optimism that is so frequently observed in the process of decision making. Searching for competing and complementing evolutionary insights for this and similar behavioral puzzles is a challenge for future research.

# 6    Appendix

**Proof of Theorem 1:** Let $D_n$ be the set of strategies that do not survive $n$ or less rounds of iterated elimination of strictly dominated strategies, so $D = \cup_{n=0}^{\infty} D_n$. Denote also by $U_n = T \setminus D_n$ the set of strategies that do survive $n$ rounds of iterated elimination of strictly dominated strategies. We prove by induction on $n$ that $U_n$ is compact, and every eliminated strategy $d \in D_n$ has an open neighborhood $W_d$ for which $\lim_{t \to \infty} G_t(W_d) = 0$.

Since $D_0 = \emptyset$ and $U_0 = T$, the claim holds for $n = 0$. If $D_1$ is empty as well, i.e. no strategies are strictly dominated, then the claim holds vacuously. So from now on assume that $D_1 \neq \emptyset$. Suppose, by induction, that the claim holds for $n < k$.

We first prove that $U_k$ is compact. Indeed, let $d \in D_k$ be round-$k$ dominated by the strategy $x \in T$, that is for every $y \in U_{k-1}$

$$f(x,y) - f(d,y) > 0. \tag{A.1}$$

Since $f$ is continuous, $[f(x,y) - f(d,y)]$ is continuous in $y$, and therefore attains its

minimum on $U_{k-1}$, as this set is compact by the induction hypothesis. Hence by (5.12)

$$\rho(x, d) \equiv \min_{y \in U_{k-1}} [f(x, y) - f(d, y)] > 0. \tag{A.2}$$

Furthermore, the function $\rho(x, d)$ is continuous since $[f(x, y) - f(d, y)]$ is. Therefore, for every $\varepsilon > 0$, the set of strategies which are dominated by $x$ by a payoff difference of at least $\varepsilon$

$$\{d : \rho(x, d) > \varepsilon\} \tag{A.3}$$

is open. Consequently, the set of strategies dominated up to round $k$

$$D_k = D_{k-1} \cup \bigcup_{x \in U_{k-1}} \bigcup_{\varepsilon > 0} \{d : \rho(x, d) > \varepsilon\}, \tag{A.4}$$

is open as a union of open sets, and $U_k = T \setminus D_k$ is compact, as required.

We now turn to complete the inductive step, and prove that every eliminated strategy $d \in D_k$ has an open neighborhood $W_d$ for which $\lim_{t \to \infty} G_t(W_d) = 0$. Indeed, let $d$ be dominated by $x$ in one of the rounds up to $k$. Then if $k > 1$, not only does $x$ perform better than $d$ against strategies in $U_{k-1}$, but it also does so against some strategies in an open subset of $D_{k-1}$: Since $f$ is continuous, the set

$$B = \{y \in T : f(x, y) - f(d, y) \leq 0\} \tag{A.5}$$

is a compact subset of the open set $D_{k-1}$. Hence $B$ is a *proper* subset of $D_{k-1}$, as $D_{k-1}$ is open by the induction hypothesis (except when $k = 1$, in which case $D_{k-1} = \emptyset$, and $B = \emptyset$).

This implies that for some positive number $s$, we have

$$\lim_{t \to \infty} G_t(C) = 0, \tag{A.6}$$

where

$$C = \{y \in D_{k-1} : f(x, y) - f(d, y) \leq s\}. \tag{A.7}$$

Indeed, for $k = 1$, $D_{k-1} = D_0 = \emptyset$, so for whatever positive $s$ chosen we have $G_t(C) = 0$, and (5.16) holds. For $k > 1$, let

$$s = \frac{1}{2} \sup_{y \in D_{k-1}} [f(x, y) - f(d, y)], \tag{A.8}$$

which is positive, by the fact that $B \subsetneq D_{k-1}$, as explained above. With this $s$, the set $C$ is a closed subset of $D_{k-1}$ (because $f(x, y) - f(d, y)$ is continuous in $y$), and hence compact.

20

By the induction hypothesis, every $y \in D_{k-1}$ has an open neighborhood $W_y \ni y$ such that $\lim_{t\to\infty} G_t(W_y) = 0$. Since $C \subseteq \bigcup_{y\in C} W_y$, the compactness of $C$ implies that there exist $y_1, \ldots y_m \in C$ such that $C \subseteq \bigcup_{i=1}^m W_{y_i}$. Therefore, $G_t(C) \leq \sum_{i=1}^m G_t(W_{y_i})$ and hence $\lim_{t\to\infty} G_t(C) = 0$, as claimed.

Observe now the following considerations. The continuity of $f$ and $\rho$ imply that there are open neighborhoods $V_x \ni x$ and $W_d \ni d$ such that for every $x' \in \overline{V}_x$, $d' \in \overline{W}_d$

$$\inf_{y \in D_{k-1}\backslash C} [f(x',y) - f(d',y)] \geq \frac{s}{2} > 0, \tag{A.9}$$

and

$$\min_{y \in U_{k-1}} [f(x',y) - f(d',y)] = \rho(x',d') \geq \frac{\rho(x,d)}{2} > 0. \tag{A.10}$$

Thus, against any strategy $y \notin C$, every strategy $x' \in \overline{V}_x$ outperforms every $d' \in \overline{W}_d$ by at least

$$\varepsilon = \min \left\{ \frac{s}{2}, \frac{\rho(x,d)}{2}, \frac{1}{2} \right\} > 0, \tag{A.11}$$

i.e.,

$$\inf_{y \in T\backslash C} [f(x',y) - f(d',y)] \geq \varepsilon. \tag{A.12}$$

At the same time, since $f$ is continuous on the compact domain $T$, there exists a bound $M$ such that $|f| \leq M$; and by (A.6), there exists a time $\bar{t}$ such that for $t \geq \bar{t}$ we have $G_t(C) < \frac{\varepsilon}{8M}$ and $G_t(T\backslash C) > 1-\varepsilon$. Altogether this implies that for $x' \in \overline{V}_x, d' \in \overline{W}_d$ and $t \geq \bar{t}$

$$f(x', G_t) - f(d', G_t) = \int_T [f(x', \cdot) - f(d', \cdot)]\, dG_t =$$

$$\int_C [f(x', \cdot) - f(d', \cdot)]\, dG_t + \int_{T\backslash C} [f(x', \cdot) - f(d', \cdot)]\, dG_t > \tag{A.13}$$

$$(-2M)\frac{\varepsilon}{8M} + \varepsilon(1 - \varepsilon) \geq -\frac{\varepsilon}{4} + \varepsilon(1 - \frac{1}{2}) = \frac{\varepsilon}{4}.$$

By the continuity of $f$, (A.13) holds also when $G_t$ is replaced by any probability measure $\mu \in A \equiv \overline{\{G_t\}_{t\geq\bar{t}}}$, the closure of $\{G_t\}_{t\geq\bar{t}}$ in the topology of weak convergence of probability measures.

Now, by the payoff monotonicity of the growth-rate function $g$, for every $\mu \in A$, $x' \in \overline{V}_x$ and $d' \in \overline{W}_d$

$$g(x', \mu) - g(d', \mu) > 0. \tag{A.14}$$

21

The continuous function $[g(x', \mu) - g(d', \mu)]$ attains its minimum on the compact set $\overline{V}_x \times \overline{W}_d \times A$. Therefore, there is in fact a positive $\delta$ for which

$$g(x', G_t) - g(d', G_t) \geq \delta \tag{A.15}$$

for $x' \in \overline{V}_x, d' \in \overline{W}_d$ and $t \geq \overline{t}$. A fortiori, this inequality holds also if we replace $g(x', G_t)$ and $g(d', G_t)$ by their averages in $\overline{V}_x$ and $\overline{W}_d$, respectively. Thus for $t \geq \overline{t}$

$$\frac{\int_{\overline{V}_x} g(\cdot, G_t) dG_t}{G_t(\overline{V}_x)} - \frac{\int_{\overline{W}_d} g(\cdot, G_t) dG_t}{G_t(\overline{W}_d)} \geq \delta. \tag{A.16}$$

Hence, by (4.15), for $t \geq \overline{t}$

$$\frac{\dot{G}_t(\overline{V}_x)}{G_t(\overline{V}_x)} - \frac{\dot{G}_t(\overline{W}_d)}{G_t(\overline{W}_d)} = \frac{d}{dt}\left[\log \frac{G_t(\overline{V}_x)}{G_t(\overline{W}_d)}\right] \geq \delta, \tag{A.17}$$

so that

$$\frac{G_t(\overline{V}_x)}{G_t(\overline{W}_d)} \geq \frac{G_{\overline{t}}(\overline{V}_x)}{G_{\overline{t}}(\overline{W}_d)} \exp[\delta(t - \overline{t})] \rightarrow_{t\to\infty} \infty. \tag{A.18}$$

Therefore, $\lim_{t\to\infty} G_t(W_d) = 0$, as required. ∎

# 7    References

Anderson C. A., M.R. Lepper, and L. Ross (1980), "Perseverance of social theories: The role of explanation in the persistence of discredited information," *Journal of Personality and Social Psychology* **39**, pp. 1037- 1049.

Babcock L. and G. Loewenstein (1997), "Explaining Bragaining Impasse: The Role of Self-Serving Biases," *Journal of Economic Perspectives,"* **11(1)**, pp. 109-126.

Benabou R. and J. Tirole (1999a), "Self-Confidence: Intrapersonal Strategies," mimeo, Princeton University.

Benabou R. and J. Tirole (1999b), "Self-Confidence and Social Interactions," mimeo, Princeton University.

Benos A.V. (1998), "Aggressiveness and Survival of Overconfident Traders," *Journal of Financial Markets* **1**, pp. 353-383.

Bergman N. and Y. Bergman (2000), "Ecologies of Preferences with Envy as an Antidote to Risk-Aversion in Bargaining", mimeo, The Hebrew University of Jerusalem.

Bester H. and W. Güth (1998), "Is Altruism Evolutionary Stable?", *Journal of Economic Behavior and Organization* **34(2),** pp. 211-221.

Bolle F. (2000), "Is Altruism Evolutionarily Stable? And Envy and Malevolence? - Remarks on Bester and Güth," *Journal of Economic Behavior and Organization* **42(1),** pp. 131-133.

Brander J. and T. Lewis (1986), "Oligopoly and Financial Structure: The Limited Liability Effect," *American Economic Review* **76**, pp. 956- 970.

Brander J. and B.J. Spencer (1985), "Export Subsidies and International Market Share Rivalry," *Journal of International Economics* **18**, pp. 83-100.

Bulow, J., J. Geanakoplos, and P. Klemperer (1985), "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy*; **93(3)**, pp. 488-511.

Burger J. M., and L. Burns L. (1988), "The Illusion of Unique Invulnerability and The Use of Effective Contraception," *Personality and Social Psychology Bulletin*, **14**, pp. 264-270.

Brocas I. and J. Carrillo (1999), "Entry Mistakes, Entrepreneurial Boldness and Optimism," CEPR D.P. 2213.

Camerer C. and D. Lovallo (1999), "Overconfidence and Excess Entry: An Experimental Approach," *American Economic Review* **89(1)**, pp. 306-318.

Dekel E., J. Ely, and O. Yilankaya (1998), "Evolution of Preferences," mimeo, Northwestern University

Dekel E. and S. Scotchmer (1999), "On the Evolution of Attitudes toward Risk in Winner-Take-All Games," *Journal of Economic Theory* **87**, pp. 95-124.

Dufwenberg M. and W. Güth (1999), "Indirect Evolution vs. Strategic Delegation: a Comparison of Two Approaches to Explaining Economic Institutions," *European Journal of Political Economy* **15**, pp. 281-295.

Eaton J. and G.M. Grossman (1986), "Optimal Trade and Industrial Policy under Oligopoly," *Quarterly Journal of Economics*, pp. 383-406.

Fershtman C. and K. Judd (1987), "Incentive Equilibrium in Oligopoly," *American Eco-

*nomic Review* **77(5)**, pp. 927-940

Fershtman C., K. Judd and E. Kalai (1991), "Observable Contracts: Strategic Delegation and Cooperation," *International Economic Review* **32(3)**, pp. 551-559.

Fershtman C. and E. Kalai (1997), "Unobserved Delegation," *International Economic Review* **38(4)**, pp. 763-774.

Fershtman C. and Y. Weiss (1997), "Why do We Care about what Others Think about Us?," in: Ben Ner, A. and L. Putterman (eds.), *Economics, Values and Organization*, Cambridge University Press, Cambridge MA.

Fershtman C. and Y. Weiss (1998), "Social Rewards, Externalities and Stable Preferences," *Journal of Public Economics* **70**, pp. 53-74.

Fleming J. and J.M. Darley (1986), "Perceiving Intension in Constrained Behavior: The Role of Purposeful and Constrained Action Cues in Correspondence Bias Effects," mimeo, Princeton University.

Frank R.H. (1987), "If Homo Economicus Could Choose His Own Utility Function, Would He Choose One With a Conscience?" *American Economic Review* **77(4)**, pp. 593-604.

Frank R.H. (1988), *Passions Within Reason – The Strategic Role of the Emotions,* W.W. Norton & Company, New York.

Güth W. and M. Yaari (1992), "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach," in Witt, U. (ed.), *Explaining Forces and Changes: Approaches to Evolutionary Economics,* University of Michigan Press.

Green J. (1992), "Commitment with Third Parties," *Annales d'Economie et de Statistique*, 25- 26, 101- 121.

Huck S. and J. Oechssler (1998), "The Indirect Evolutionary Approach to Explaining Fair Allocations," *Games and Economic Behavior* **28**, pp. 13-24.

Huck S., G. Kirchsteiger, and J. Oechssler (1997), "Learning To Like What You Have – Explaining the Endowment Effect," mimeo, Humboldt University, Berlin.

Katz M. (1991), "Game-Playing Agents: Unobservable Contracts as Precommitments," *Rand Journal of Economics* **22**, pp. 307- 328.

Klayman J. and Ha Y- W, (1987), Confirmation, Disconfirmation, and Information in Hypothesis Testing, *Psychological Review* **94**, pp. 211- 228.

Koçkesen L., E.A. Ok, and R. Sethi (1998), "Evolution of Interdependent Preferences in Aggregative Games," C.V. Starr RR# 98-19, New York University.

Kyle A.S. and A. Wang (1997), "Speculation Duopoly with Agreement to Disagree: Can Overconfidence Survive the Market Test?" *The Journal of Finance* **LII**, pp. 2073-2090.

Lazear E. and S. Rosen (1981), "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy*, **89(5)**, pp. 841-864.

Lehman D. R., and R.E. Nisbett (1985). "Effects of Higher Education on Inductive Reasoning," Unpublished manuscript, University of Michigan (cited in D. Myers, 1998, *Psychology,* Fifth Edition, Worth Publishers Inc.).

Lewinsohn P.M., W. Mischel W. Chaplin, and R. Barton (1980), "Social Competence and Depression: The Role of Illusory Self-Perceptions," *Journal of Abnormal Psychology* **898**, pp. 203-212.

Lord C. G., Ross L., and Lepper M. (1979), "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology* **37**, pp. 2098- 2109.

Moulin H. (1984), "Dominance Solvability and Cournot Stability," *Mathematical Social Sciences,"* **7(1)**, pp. 83-102.

Oechssler J. and F. Riedel (1999), "Evolutionary Dynamics on Infinite Strategy Spaces," D.P. A-606, Bonn University, forthcoming in *Economic Theory.*

Ok E.A. and F. Vega-Redondo (1999), "On the Evolution of Individualistic Preferences: Complete versus Incomplete Information Scenarios," mimeo, New York University.

Possajennikov A. (2000), "On The Evolutionary Stability of Altruistic and Spiteful Preferences," *Journal of Economic Behavior and Organization* **42(1)** pp. 125-129 (preliminary version appeared as Possajenikov (1999), CentER working paper 9956, Tilburg University).

Robson A.R. (1996a), "A Biological Basis for Expected and Non-Expected Utility," *Journal of Economic Theory* **68**, pp. 397-424.

Robson A.R. (1996b), "The Evolution of Attitudes to Risk: Lottery Tickets and Relative Wealth," *Games and Economic Behavior* **14**, pp. 190-207.

Rogers A.R. (1994), "Evolution of Time Preference by Natural Selection," *American Economic Review* **84**, pp. 460-481.

Rogoff K. (1985), "The Optimal Degree of Commitment to an Intermediate Monetary Target," *Quarterly Journal of Economics*, **100(4)**, pp. 1169-1189.

Samuelson L. and J. Zhang (1992), "Evolutionary Stability in Asymmetric Games," *Journal of Economic Theory* **57**, pp. 363-391.

Schelling T., (1960), *The Strategy of Conflict*, Cambridge MA: Harvard University Press.

Segerstrom S. C., W.J. McCarthy, N.H. Caskey, T.D. Gross, and M.E. Jarvik (1993), "Optimistic Bias Among Cigarette Smokers." *Journal of Applied Social Psychology*, **23**, pp. 1606-1618.

Seligman M. E. P. and , P. Schulman (1986), "Explanatory Style as a Predictor of Productivity and Quitting Among Life Insurance Sales Agents," *Journal of Personality and Social Psychology*, **50**, pp. 832-838.

Skov R. B., and Sherman S. J., (1986), "Information-gathering processes: Diagnosticity, hypothesis- confirmatory strategies, and perceived hypothesis confirmation," *Journal of Experimental Social Psychology* **22**, pp. 93- 121.

Svenson O. (1981), "Are We All Less Risky Drivers and More Skillful than our Fellow Drivers?" *Acta Psychologica* **47(2)**, pp. 143-148.

Taylor S.E. and J.D. Brown (1988), "Illusion and Well-Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin* **103**, pp. 193-210.

Taylor P. and L. Jonker (1978), "Evolutionary Stable Strategies and Game Dynamics," *Mathematical Biosciences* **40**, pp. 145-156.

Vega-Redondo F. (1997), "The Evolution of Walrasian Behavior," *Econometrica* **65**, pp. 375-384.

Waldman M. (1994), "Systematic Errors and the Theory of Natural selection," *American Economic Review* **84(3)**, pp. 482-497.

Wason P. C. (1981), "The importance of cognitive illusions," *The Behavioral and Brain Sciences* **4**, p. 356.

Weibull J. (1995), *Evolutionary Game Theory*, MIT Press, Cambridge MA..