EQUILIBRIUM VENGEANCE

DANIEL FRIEDMAN
NIRVIKAR SINGH

CESIFO WORKING PAPER NO. 766
CATEGORY 10: EMPIRICAL AND THEORETICAL METHODS
AUGUST 2002

CESifo *Working Paper No. 766*

EQUILIBRIUM VENGEANCE

Abstract

This paper introduces two ideas, emotional state dependent utility components (ESDUCs), and evolutionary perfect Bayesian equilibrium (EPBE). Using a simple extensive form game, we illustrate the efficiency-enhancing role of a powerful ESDUC, the vengeance motive. Incorporating behavioral noise and observational noise leads to a range of (short run) Perfect Bayesian equilibria (PBE) involving both vengeful and non-vengeful types. We then derive two (long run) EPBE, one where both types survive and reap mutual gains, and a second where only the non-vengeful type survives and there are no mutual gains.

JEL Classification: C73, D64, Z13, B52.

Keywords: negative reciprocity, perfect Bayesian equilibrium, evolutionary perfect Bayesian equilibrium, emotional state dependent utility.

Daniel Friedman
Department of Economics
University of California
Santa Cruz CA 95064
U.S.A.
dan@cats.ucsc.edu

Nirvikar Singh
Department of Economics
University of California
Santa Cruz CA 95064
U.S.A.
boxjenk@cats.ucsc.edu

We are grateful for the helpful comments of seminar participants at UC Berkeley, UC Riverside and UC Santa Cruz, and to readers of an earlier version of the paper, particularly Joshua Aizenman, Steve Goldman, Steffen Huck, Matt Rabin, Donald Wittman, Huibin Yan and Daniel Zizzo. We are also grateful to Matt McGinty for patient research assistance.

1 Introduction

Craving vengeance is a powerful human motive: when some culprit harms you or your loved ones, you may choose incur a substantial personal cost to harm him in return. There can be major economic and social consequences, positive and negative.

Economic theory has not yet fully come to grips with such motives. In this paper we propose a general approach to modelling other-regarding preferences that we call emotional state dependent utility components (ESDUCs), and use it to investigate vengeance.

A taste for vengeance, the desire to "get even," is so much a part of daily life (and the evening news) that it is easy to miss the evolutionary puzzle. We shall argue that indulging your taste for vengeance in general reduces your material payoff or fitness. Absent countervailing forces, the meek (less vengeful people) should have inherited the earth long ago, because they had higher fitness. Why then does vengeance persist?

The other title word of our paper is equilibrium. We also propose an apparently new equilibrium concept, evolutionary perfect Bayes equilibrium (EPBE), that seems germane in a wide variety of applications. It defines long run equilibrium (generalizing the equal profit condition of competitive markets) for games of incomplete information with possible entry, exit and/or switching among multiple player types. In this paper we use EPBE to show how vengeance can persist despite its apparent fitness handicap.

Vengeance is closely tied to several vexing issues, methodological and substantive. To clear the underbrush, we begin with preliminary discussions on the nature of social dilemmas, the meaning of positive and negative reciprocity, evidence on why both are important to economists, and various modelling approaches employed so far. Section 3 presents the basic social dilemma as a simple extensive form game, and shows how vengeful preferences can dramatically improve equilibrium efficiency. It also spotlights the evolutionary problem when an individual's vengefulness cannot be perfectly known in advance and when behavioral errors are possible. Section 4 derives three families of perfect Bayes equilibria (PBE), two pooling equilibria and one separating equilibrium. The PBE are short-run in that the nature and proportions of different types are fixed. Section 5 examines the long-run in which the nature and proportions of types can evolve. We define EPBE and derive a unique EPBE that supports social gains as well as a trivial, inefficient EPBE. Following a concluding discussion, Appendix A collects the

mathematical details.

2 Preliminaries

An action has has a social dimension when it affects non-actors as well as the actor. Figure 1 lays out the possibilities in terms of the net material benefit ($x > 0$) or cost ($x < 0$) to the actor, denoted "Self," and the net material benefit ($y > 0$) or cost ($y < 0$) to counterparties, denoted "Other". Economists think most often about the mutual gains quadrant I, where actions simultaneously benefit Self and Other. Such symbiotic actions increase social efficiency.

Quadrant IV is the well-studied opportunistic region, where Self benefits at Other's expense; the biological terms are parasitism and predation. The flip side is the altruism quadrant II, where Self bears a personal cost in order to benefit Other. Quadrant III is especially interesting to us. Carlo Cipolla (1976) refers to actions producing such outcomes as stupidity, but we shall interpret them as vengeance.

Social dilemmas arise from the fact that evolution directly supports behavior that benefits Self, i.e., outcomes $x > 0$ in quadrants IV (or I) but not $x < 0$ in II (or III), while in contrast, efficiency requires outcomes above the diagonal $[x + y = 0]$.[1] Social creatures (such as humans) thrive on devices that support outcomes in the half-quadrant II+ and discourage outcomes in IV-. Such devices somehow internalize Other's costs and benefits.

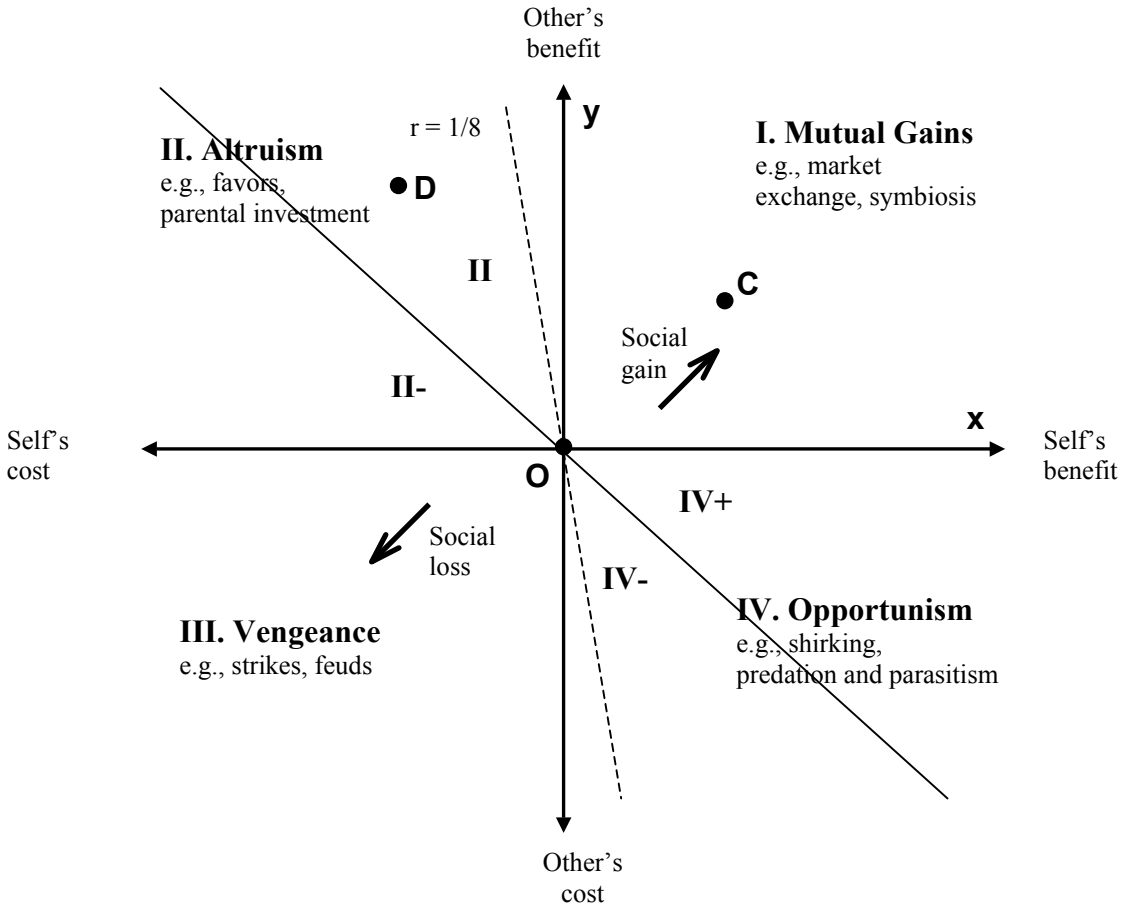
————fig 1 about here————

2.1 Efficiency-enhancing devices

Biologists emphasize the device of genetic relatedness. If Other is related to Self to degree $r > 0$, then a positive fraction of Other's payoffs are internalized via "inclusive fitness" (William Hamilton, 1964) and evolution favors outcomes above the line $[x + ry = 0]$. For example, the unusual genetics of insect order hymenoptera produce r up to 3/4 between sisters, so it is no surprise that most social insects (including ants and bees) belong to this order and that the workers are sisters. For humans and most other species, r

[1] More precisely, Self's iso-fitness curves are the vertical lines $x = C$ while iso-efficiency curves are diagonal lines $x + y = C$. The status quo point (0, 0) ensures that $C \geq 0$ is feasible.

Figure 1: Payoffs to Self and Other



is only $\frac{1}{2}$ for full siblings and for parent and child, is 1/8 for first cousins, and goes to zero exponentially for more distant relations. On average r is rather small in human interactions, as in the steep dashed line in Figure 1, since we typically have only a few children but work and live in groups with dozens of individuals. Clearly non-genetic devices are needed to support human social behavior.

Economists emphasize devices based on repeated interaction, as in the "folk theorem" (e.g., Drew Fudenberg and Eric Maskin, 1986). Suppose that Other returns the benefit ("positive reciprocity") with probability and delay together summarized in discount factor $\delta \in [0, 1)$. Then that fraction of other's payoffs are internalized (Robert Trivers, 1971) and evolution favors behavior producing outcomes above the line $[x + \delta y = 0]$. This device can support a large portion of socially efficient behavior when δ is close to 1, i.e., when interactions between two individuals are symmetric, predictable, frequent and ongoing. But humans specialize in exploiting once-off opportunities with a variety of different partners, and here δ is small, as in the same steep dashed line. Other devices are needed to explain such behavior.

Here we will emphasize devices based on other-regarding preferences. For example, suppose Self gets a utility increment of ry from his or her action,[2] in addition to the material benefit x. Hence Self partially internalizes the material externality, and undertakes behavior that is above the line $[x + ry = 0]$. Friendly preferences, $r \in [0, 1]$, thus can explain the same range of behavior as genetic relatedness and repeated interaction. However, by itself the friendly preference device is evolutionarily unstable: those with lower positive r will tend to make more personally advantageous choices, gain higher material payoff (or fitness), and displace the more friendly types. Friendly preferences therefore require the support of other devices.

Vengeful preferences rescue friendly preferences. Self's material incentive to reduce r disappears when others base their values of r on Self's previous behavior and employ $r < 0$ if Self is insufficiently friendly. Such visits to quadrant III will reduce the fitness of less friendly behavior and thus boost friendly behavior. But visits to quadrant III are also costly to the avenger, so less vengeful preferences seem fitter. What then supports vengeful preferences: who guards the guardians? This is the central question in the present paper.

[2]Rilling et al (2002) present recent physiological evidence for such increments, based on fMRI brain scans of subjects playing prisoner's dilemma.

2.2 Modelling other regarding preferences

Two main approaches can be distinguished in the recent literature. The distributional preferences approach is exemplified in the Ernst Fehr and Klaus Schmidt (1999) inequality aversion model, the Gary Bolton and Axel Ockenfels (2000) mean preferring model, and the Gary Charness and Matthew Rabin (2001) social maximin model. These models begin with a standard selfish utility function and add additional terms capturing self's response to how own payoff compares to other's payoffs. In Fehr-Schmidt, for example, my utility decreases (increases) linearly in your payoff when your payoff is above (below) my own.

The other main approach is to model reciprocal preferences directly. Building on the John Geanakoplos, David Pearce and Ennio Stacchetti (1989) model of psychological games, Rabin (1993) constructs a model of reciprocation for two player normal form games, extended by Martin Dufwenberg and Georg Kirchsteiger (1998) and Armin Falk and Urs Fischbacher (1998) to somewhat more general settings. The basic idea is that my preferences regarding your payoff depends on my beliefs about your intentions, e.g., if I believe you want to increase my payoff then I want to increase yours. Such models are intractable except in the simplest settings. David Levine (1998) improves tractability by replacing beliefs about others' intentions by estimates of others' type.

We favor a further simplification. Model reciprocal preferences as state dependent: my attitude towards your payoffs depends on my emotional state, e.g., friendly or vengeful, and your behavior systematically alters my emotional state. This emotional state dependent other-regarding utility component (ESDUC) approach is consistent with the discussion in Joel Sobel (2000) and is hinted at in some other papers including Charness and Rabin. The approach is quite flexible and tractable, but in general requires a psychological theory of how emotional states change (van Winden, 2001). Fortunately a very simple rule will suffice for present purposes: you become vengeful towards those who betray your trust, and otherwise have standard selfish preferences.

Empirical evidence is now accumulating that compares the various approaches. James C. Cox and Daniel Friedman (2002), for example, review about two dozen very recent papers. Some authors find evidence favoring the distributional models, but most authors find evidence mainly favoring state dependent or reciprocal models. Our own reading convinces us to focus

on state dependent preferences, while noting that distributional preferences may also play a role.

2.3 Indirect Evolution

Many economists concede that empirical evidence inconsistent with selfish rationality is very strong, but nevertheless, on theretical grounds, resist models with other regarding preferences. The problem is that arbitrary behavior can be rationalized by putting in arbitrary preferences for such behavior, but such models have no predictive power.

The theoretical justification for selfish rationality is evolutionary, as exemplified in Armin Alchian (1950), Milton Friedman (1953) and Gary Becker (1962). We believe that our ESDUC model, or any other preference model, requires the same justification. The model must account for the empirical data but also must pass the following theoretical test: people with the hypothesized preferences receive at least as much material payoff (or evolutionary fitness) as people with alternative preferences. Otherwise, the hypothesized preferences would disappear over time, or never appear in the first place.[3]

This test is sometimes referred to as indirect evolution (Werner Guth and Menachem Yaari, 1992) because evolution operates on preference parameters that determine behavior rather than operating directly on behavior. The idea goes back at least to Gary Becker (1976) and Paul Rubin and Charles Paul (1979), and can be seen a many recent papers such as Steffen Huck and Jorg Oechssler (1999), Jeffrey Ely and Okan Yilankaya (2001), and Larry Samuelson and Jeroen Swinkels (2001). Most of these papers focus on positive reciprocity rather than negative reciprocity, or vengeance.[4]

[3]Some behavioral economists disagree, and believe that it is sufficient to write a parsimonious model consistent with the data. Sendhil Mullainathan and Richard Thaler (2000), for example, argue that evolutionary forces are not swift enough to wipe out irrational behavior in complex, rapidly changing economies. Our response is that theoretical discipline is necessary, because many different parsimonious models can be developed to account for any given set of data. Even though it is relatively slow, genetic evolution operating in the social environments of our hominid ancestors surely helped shape our emotional capacities. We would emphasize individual learning and cultural transmission as rapid evolutionary forces that operate in modern economies.

[4]Huck and Oechssler is an exception. They study negative reciprocity in a small group environment where a vengeful person can impair others' fitness more than his own. The model presented below considers only large groups where no single person can affect the

Our task is to show that people whose utility functions contain a vengeful component will achieve in social interactions at least as much material payoff as other people whose utility functions contain only their own material payoff. Equally important, and neglected in most of the indirect evolution work so far, we want to show that a greater or lesser degree of vengefulness will not lead to higher material payoffs.[5]

3 The Underlying Game

The first step in analyzing social preferences is to model explicitly the underlying social dilemma. We use a simple extensive form version of the prisoner's dilemma, sometimes known as the Trust game (e.g., Guth and Hartmut Kliemt, 1994), shown in Panel A of Figure 2. Player 1 (Self) can opt out (N) and ensure zero payoffs to both players. Alternatively Self can trust (T) player 2 (Other) to cooperate (C), giving both a unit payoff and a social gain of 2. However, Other's payoff is maximized by defecting (D), increasing his payoff to 2 but reducing Self's payoff to -1 and the social gain to 1. (These payoffs are labelled **O, C** and **D** in Figure 1.) The basic game has a unique Nash equilibrium found by backward induction: Self chooses N because Other would choose D if given the opportunity, and social gains are zero.

To this underlying game we add a punishment technology and a punishment motive as shown in Panel B. Self now has the last move and can inflict harm (payoff loss) h on Other at personal cost ch. The marginal cost parameter c captures the technological opportunities for punishing others.

————fig 2 about here————

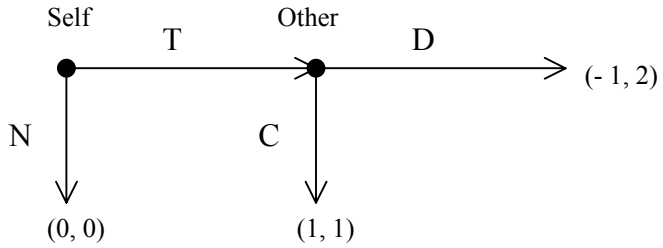
Self's punishment motive is given by state dependent preferences. If Other chooses D then Self receives a utility bonus of $v \ln h$ (but no fitness bonus) from Other's harm h. In other states utility is equal to own payoff. The motivational parameter v is subject to evolutionary forces and is intended to capture an individual's temperament, e.g., his susceptibility to

average fitness.

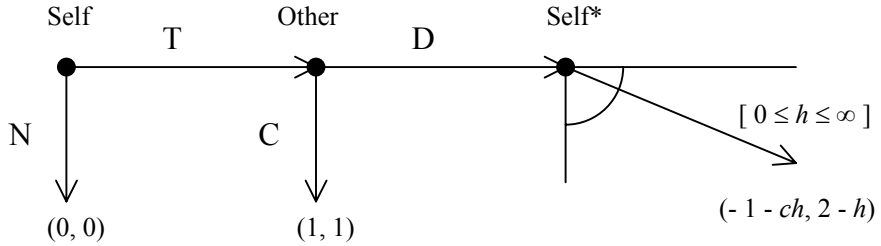
[5]Thus we respond to the first challenge raised by Samuelson (2001) in his introduction to a recent symposium on the evolution of preferences. The point is important here because most previous models of negative reciprocity are susceptible to unravelling: slightly lesser degrees of vengefulness have higher fitness. Our PBE and EPBE models also respond to Samuelson's other challenge, to consider issues of preference observability.

Figure 2: Fitness Payoffs

A. Basic Trust Game

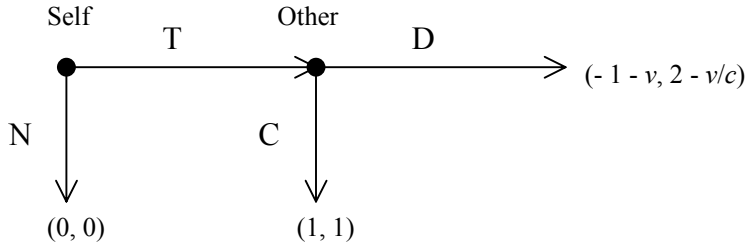


B. Extended Trust Game



*Utility payoff to Self is *lnh* - 1 - *ch*

C. Reduced Trust with a vengeance



anger. See Robert Frank (1988) for an extended discussion of such traits. The functional forms for punishment technology and motivation are convenient (we will see shortly that v parameterizes the incurred cost) but are not necessary for the main results. The results require only that the chosen harm and incurred cost are increasing in v and have adequate range.

Using the notation I_D to indicate the event "Other chooses D," we write Self's utility function as $U = x + vI_D \ln h$, that is, own material payoff x plus the relevant ESDUC. When facing a "culprit" (I_D =1), Self chooses the reduction h in Other's payoff so as to maximize $U = -1 - ch + v \ln h$. The unique solution of the first order condition is $h^* = v/c$ and the incurred cost is indeed $ch^* = v$. For the moment assume that Other correctly anticipates this choice. Then we obtain the reduced game in Panel C. For selfish preferences ($v = 0$) it coincides with the original version in Panel A with unique Nash equilibrium (N, D) yielding the inefficient outcome (0, 0). For $v > c$, however, the transformed game has a unique Nash equilibrium (T, C) yielding the efficient outcome (1, 1). The threat of vengeance rationalizes Other's cooperation and Self's trust.

3.1 Can vengeful preferences evolve?

Vengeance thus may have a pro-social role, but is it viable? Existing literature on preference evolution suggests an equivocal answer: Yes if Self's vengefulness is observable by Others, but No if it is not. To answer the question properly (Samuelson, 2001), we must consider intermediate cases, which we refer to as observational noise. To begin, assume Other perceives Self's vengeance level as $u = v + y$ when the true vengeance level is v. The observational error y has scale (e.g., standard deviation) $\sigma \geq 0$. We will see (perhaps surprisingly) that moderate positive σ can help stabilize a high level of vengeance.

Behavioral noise is also crucial for the viability of vengeance. Self may intend to choose N but may twist an ankle and find himself depending on Other's cooperative behavior, and Other may intend to choose C but oversleeps or gets tied up in traffic. Such considerations can be summarized in a behavioral noise amplitude or 'tremble rate' $e \geq 0$. Larger values of e would seem to raise Self's cost of vengefulness and reduce fitness.

A preliminary viability analysis proceeds as follows. Fix the noise levels $e \geq 0$ and $\sigma \geq 0$ as well as the marginal punishment cost c, and assume that for a given distribution of v within the population, the choices of Self and

Other adjust rapidly towards (short run) Nash equilibrium. The task is to compute Self's expected fitness or material payoff $W(v;\sigma,e)$ for each value of v at the relevant short run equilibrium.

First consider the case $\sigma = e = 0$, where v is perfectly observed and behavior is noiseless. Recall that in this case the short run equilibrium (N, D) with payoff $W = 0$ prevails for $v < c$, and (T, C) with $W = 1$ prevails for $v > c$. Thus $W(v;0,0)$ is the unit step function at $v = c$.

With behavioral but no observational noise, $e > 0 = \sigma$, more vengeful types incur a greater cost when punishment is called for. Figure 3 shows that now Self's fitness function slopes downward at approximate rate $-e$, the punishment probability. Finally, with observational noise also present, $\sigma > 0$, the sharp step at $v = c$ is smeared out. The underlying calculations are collected in the Appendix.

————fig 3 about here————

4 Perfect Bayesian Equilibrium

Figure 3 shows two local fitness maxima for Self, one at $v = 0$ and the other at $v = v_H > c$,when σ and e are both small and positive. The function W defines a fitness landscape in which evolution pushes the evolving trait v uphill (Sewall Wright, 1949; Ilan Eshel, 1983; Stuart Kauffman, 1993) along the fitness gradient. The figure therefore suggests that we will end up with some fraction x of the Self population with vengeance near v_H and the rest, $(1 - x)$, with vengeance near $v = 0$. The fractions represent the arbitrary portions of the population initially above and below the fitness minimum (near $c - \sigma$).

But do we have an equilibrium? Figure 3 assumes that Other always attends to his perception of Self's vengeance. However, if the error amplitude σ is sufficiently large relative to the fraction $(1 - x)$ of non-vengeful types, then Other might be better off ignoring the perception and always playing C. (See Donald Wittman, 1989, for an analogous situation in arms control.) Likewise, if vengeful types are sufficiently rare, Other might be better off playing D regardless of his perception. These possibilities are formalized below as pooling equilibria.

To investigate, we write out the game of incomplete information. In doing so, we replace the error amplitude σ with a probability a that there is an observational error. This latter formulation is more suitable for the case

Figure 3: Self's Fitness, *W*, as a Function of Vengefulness *v*

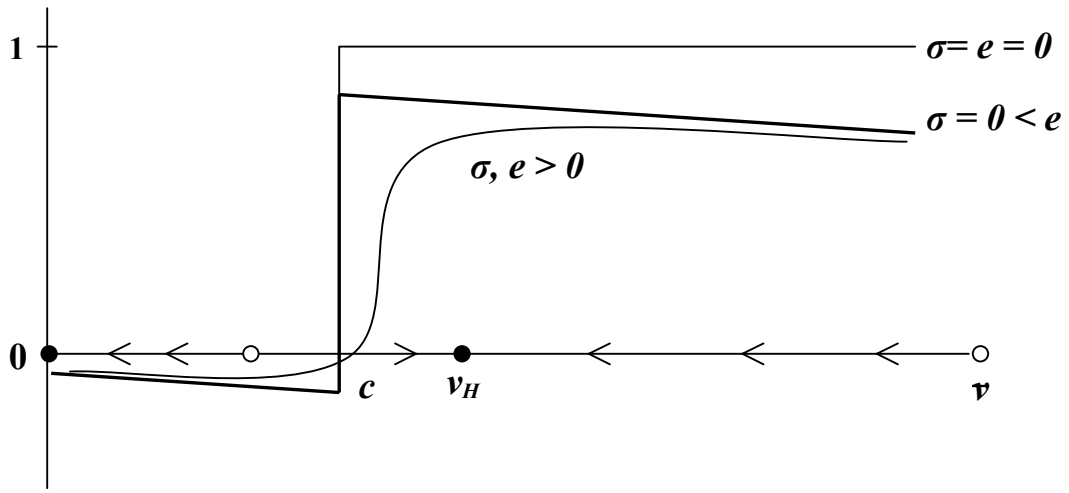
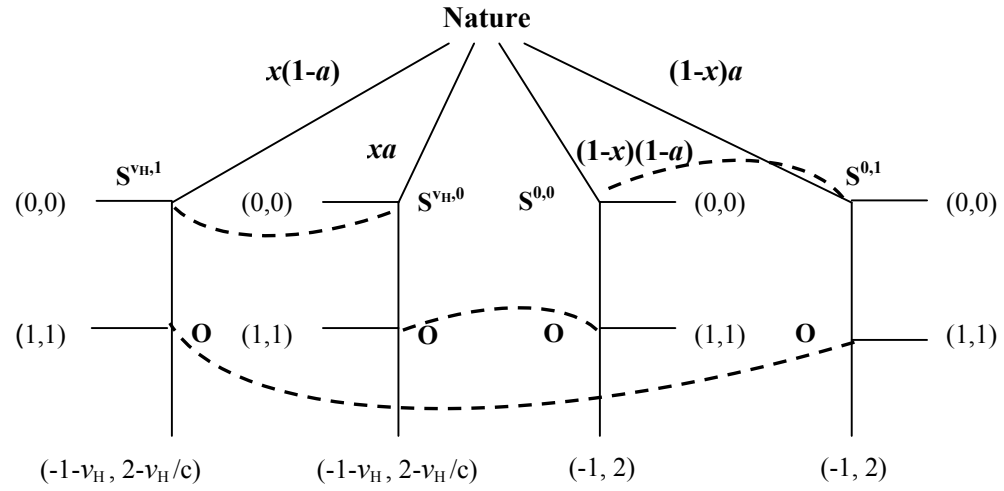


Figure 4: Game Tree



where v takes on just two values, and we shall use it in the remainder of the paper.

————fig 4 about here————

Figure 4 shows the game tree when there is observational noise. Nature chooses Self's true preference parameter as $v = 0$ with probability $1 - x$, or as $v = v_H > c$ with probability x. Nature also independently chooses Other's perception as correct ($s = 0$ for $v = 0$ and $s = 1$ for $v = 1$) with probability $1 - a$, or incorrect with probability $a = \Pr[s = 0|v = v_H] = \Pr[s = 1|v = 0]$. Self knows own preference but not the realized perception, and Other knows the perception but not the true preference.

Self's strategy (given her realized preference parameter v) is just the mixture probability for choosing T; the unconstrained stategy set is $[0, 1]$ but with behavioral noise $e \in [0, 1/2)$, Self's strategy set is $[e, 1 - e]$. Other's strategy is a pair of mixture probabilities for choosing C after observing respectively $s = 0$ and $s = 1$. The behavioral noise constrained strategy set is $[e, 1-e] \times [e, 1-e]$. The payoffs are as in the reduced Trust game of Figure 2C.

The relevant equilibrium concept is perfect Bayesian equilibrium, PBE (see e.g., Fudenberg and Tirole, 1991, chapter 8), suitably rephrased to deal with large populations and explicit behavioral trembles. PBE requires all players to optimize given beliefs, and requires that beliefs are Bayesian posterior probabilities obtained from observed actions and signals and from prior information on the type proportions.

We seek a separating PBE in which the v_H type of Self tries to Trust (i.e., plays T with maximal probability $1 - e$) and the $v = 0$ type tries to play N (i.e., plays T with minimal probability e), while all Others try to play C (and do so with probability $1 - e$) when the perception is $s = 1$ and try to play D when the perception is $s = 0$. Table 1 writes out the resulting fitness outcomes and probabilities.

The key conditions for the separating PBE arise from Other's decision problem after a noisy perception. Other compares the expectation of the D payoff $2 - v/c$ to the C payoff 1. This comparison immediately leads to the rule: try to choose D if $E(v|s = 0) \leq c$ and C if $E(v|s = 1) \geq c$. An $s = 0$ perception will arise in the separating PBE from a v_H type only if she plays T and Other has an erroneous perception, which happens with probability $x(1 - e)a$. The same perception will arise from from a $v = 0$ type only if she erroneously plays T and is correctly perceived, which happens with

probability $(1-x)e(1-a)$. It is straightforward to show (see Appendix) that the critical posterior expectation $E(v|s=0)=c$ corresponds to prior probability (or population fraction) $x^s = 1/(1+(\frac{a}{1-a})(\frac{1-e}{e})(\frac{v_H-c}{c}))$. Hence a necessary condition is $x \leq x^s$. Using the log odds function $L(y) = \ln(\frac{1-y}{y})$, the condition can be written $L(x) \geq L(x^s) = -L(a) + L(e) + L(c/v_H)$.

————table 1 about here————

Similar reasoning regarding the $s=1$ perception gives a lower bound on x (or an upper bound on $L(x)$.) Slightly simpler reasoning gives bounds for the pooling equilibria in which Other disregards the signal. The general result for separating and pooling PBE is as follows.

Proposition 1. Given perceptions with error rate a and choices with tremble rate e, and given types $v=0$ and $v=v_H>c$ constituting Self population fractions $(1-x)$ and $x \in (0,1)$, assume that $0<a,e<1/2$ and $\alpha = a+e-2ae \leq 1/(2+v_H)$. Then

- the separating PBE given in Table 1 exists iff $L(c/v_H)+L(e)-L(a) \leq L(x) \leq L(c/v_H)+L(e)+L(a)$;
- the Good Pooling equilibrium exists iff $L(x) \leq L(c/v_H)-L(a)$, and
- the Bad Pooling equilibrium exists iff $L(x) \geq L(c/v_H)+L(a)$.

There is no PBE if $L(c/v_H)-L(a) < L(x) < L(c/v_H)+L(e)-L(a)$.

A proof appears in the Appendix.

The Proposition extends to the limiting cases of vanishing errors. With no observational error, $a=0$, the separating PBE (as in the earlier discussion of Figure 4) exists for all $x \in (0,1)$. The inequalities in the Proposition show that both pooling equilibria disappear as $a \to 0$ or $L(a) \to \infty$. With no behavioral error, $e=0$, the separating equilibrium disappears because $L(c/v_H)+L(e)-L(a) \leq L(x)$ fails as $L(e) \to \infty$. The intuition is that Others will ignore their perceptions when (consistent with separating PBE) unvengeful Selfs never play T; but then unvengeful Selfs' best response is to play T (contrary to separating PBE). The pooling equilibria both exist over ranges that are independent of e.

————fig 5 and table 2 about here——

A numerical example may help fix ideas. Figure 5 (and Table 2) takes the marginal punishment cost to be $c=0.5$ and the vengeful type to have preferred punishment expenditure $v_H = 2.0$. It assumes behavioral noise rate $e=0.05$ and observational noise rate $a=0.10$. Then for sufficiently small

Table 1: PBE Probabilities

		Fitness Payoff	Probability		
	Choice	**Self, Other**	**Separating**	**Good Pooling**	**Bad Pooling**
$v = v_H$	(N, .) (T, C) (T, D)	0, 0 1, 1 -(1+v), 2-v/c	e $(1-e)(1-\alpha)$ $(1-e)\alpha$	e $(1-e)^2$ $(1-e)e$	$1-e$ e^2 $e(1-e)$
$v = 0$	(N, .) (T, C) (T, D)	0, 0 1, 1 -1, 2	$1-e$ $e\alpha$ $e(1-\alpha)$	e $(1-e)^2$ $(1-e)e$	$1-e$ e^2 $e(1-e)$

Notes:
In separating equilibrium, Other tries to play C if s = 1 and D if s = 0.
Other always tries to play C in Good Pooling, and always tries to play D in Bad Pooling.
The signal s = 1 with probability $a \in (0, \frac{1}{2})$ when $v = 0$, and is s = 0 with probability a when $v = v_H$.
The expression $\alpha = a(1-e) + e(1-a) = e + a - 2ae$ represents the probability that Other chooses his less preferred action.

Figure 5: PBE Illustration
Parameter Values: $a = 0.1$, $e = 0.05$, $c = 0.5$, $v_H = 2$

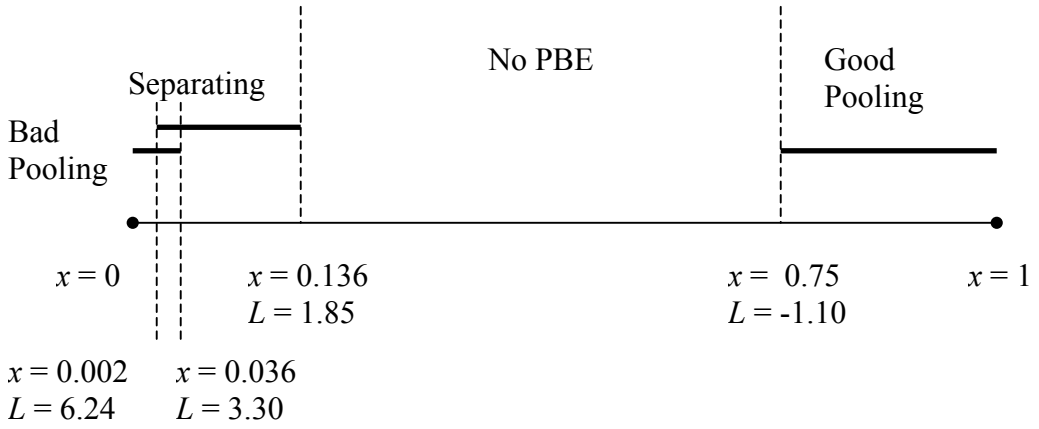


Table 2: PBE Calculations

	Separating Equilibrium		Good Pooling Equilibrium	
	Vengeful type $v = v_H$	Non-vengeful type $v = 0$	Vengeful type $v = v_H$	Non-vengeful type $v = 0$
Fitness function	$(1-e)(1-(2+v_H)\,\alpha)$	$e(2\alpha - 1)$	$(1-e)(1-(2+v_H)\,e)$	$(1-e)(1-2e)$
Value in example	0.418	- 0.036	0.760	0.855

proportions of the vengeful type ($L(x) \geq 3.30$ or $x \leq 0.036$) we have a Bad Pooling equilibrium: both types of Self try to opt out and Other tries to defect regardless of signal. For an overlapping range of vengeful type proportions ($L(x) \in [1.85, 6.24]$ or $x \in [0.002, 0.136]$) we have the separating equilibrium. No separating or pooling PBE exists for higher values of x until we reach $x = 0.75$ after which point we have the Good Pooling equilibrium.

5 Evolutionary Perfect Bayesian Equilibrium

The numerical example spotlights an evolutionary problem. In the separating PBE, the vengeful type has higher fitness (0.418) than the unvengeful type (-0.036). Therefore, by the basic principle of evolution, the fraction x of vengeful types should increase. But the separating PBE disappears when x gets above 13.6%. On the other hand, when x is large enough, we have a different equilibrium, the good pooling one. Here the unvengeful type is fitter (0.855) than the vengeful type (0.760), so x should decrease until it falls below 75% and the good pooling equilibrium disappears. Neither equilibrium seems stable.

The evolutionary problem is not due to an unfortunate parameter choice in the numerical example. In the separating PBE, the vengeful type always achieves positive fitness; otherwise she would not try to play T. The unvengeful type achieves negative fitness in this equilibrium because, with observational error rate $a < 1/2$, the payoff -1 is more frequent than +1. (See Table 2 for the general fitness expressions.) Hence evolutionary forces will increase x in the separating PBE. In the good pooling PBE, the vengeful type always has lower fitness because of the extra cost $(1-e)v_H e$ of reacting to Other's trembles, so evolutionary forces will decrease x. But no PBE exists in the intermediate region, so it seems that evolution undermines perfect Bayesian equilibrium.

5.1 Equal Fitness Principle

The problem is not due just to the peculiarities of our noisy trust game. Games of incomplete information generally have multiple types, and numerous mechanisms tend to increase the prevalence of high payoff types relative to low payoff types. For example, in an industry where firms with high quality products compete with those with low quality, one expects the market

share of the less profitable type of firms to decrease over time because they expand less rapidly or exit, or switch types. As another example, a type of worker with lower full compensation (earnings, benefits and perks net of effort cost and opportunity cost) should become less prevalent due to earlier retirements, lower accession rates, etc.

The point is that payoffs should be equalized across surviving types in long run equilibrium. We have no quarrel with PBE (or refinements such as sequential equilibrium) as a short run equilibrium concept, but in the long run the types and their relative prevalence should adjust so that only those types with highest payoff remain. This is precisely the "survival of the fittest" principle of evolutionary theory. It is also precisely the textbook distinction between short run and long run competitive equilibrium. To formalize it (in English for the time being, to avoid heavy notation not needed later), we propose the following general definition for games of incomplete information.

Definition. An evolutionary perfect Bayesian equilibrium (EPBE) is a PBE distribution over extensive form game strategy profiles such that in each population the strategies in the support of the distribution achieve equal and maximal expected fitness.

Several remarks are in order before applying the definition to the noisy reduced trust game.

- The original evolutionary equilibrium concept (Maynard Smith and Price, 1973) is evolutionary stable strategies (ESS); it applies to symmetric bimatrix games. EBPE is a generalization to games of incomplete information. EPBE, like ESS, is a static equilibrium concept that leaves implicit the evolutionary dynamics.

- EPBE appears to be new. There is already a literature on evolution in games of incomplete information, but it allows an arbitrary distribution of types that generally have different fitnesses. For example, N*ö*ldeke and Samuelson (1997) and Jacobsen et al (2001) fix the proportions of two seller types (high quality and low quality) and model the evolution of buyer beliefs regarding costly signals sent by the sellers. It seems to us that such analysis applies to short or perhaps medium run equilibrium before the more profitable types can increase their market share.

- We regard EPBE as an appropriate concept for long run equilibrium whenever (a) material payoffs such as income or evolutionary fitness can

be compared across types, and (b) adjustment mechanisms can affect existing types and their prevalence. Earlier definitions of evolutionary equilibrium might be interpreted as long run equilibria when the types are determined by last minute circumstance, and evolutionary selection applies to complete type-contingent strategies rather than to the types themselves. For example, an animal might find itself at different points of its life to be the large or small type, or the owner or intruder type. However, the contingent type interpretation is hard to justify in most applications we have seen, including the seller signaling game and our reduced trust game.

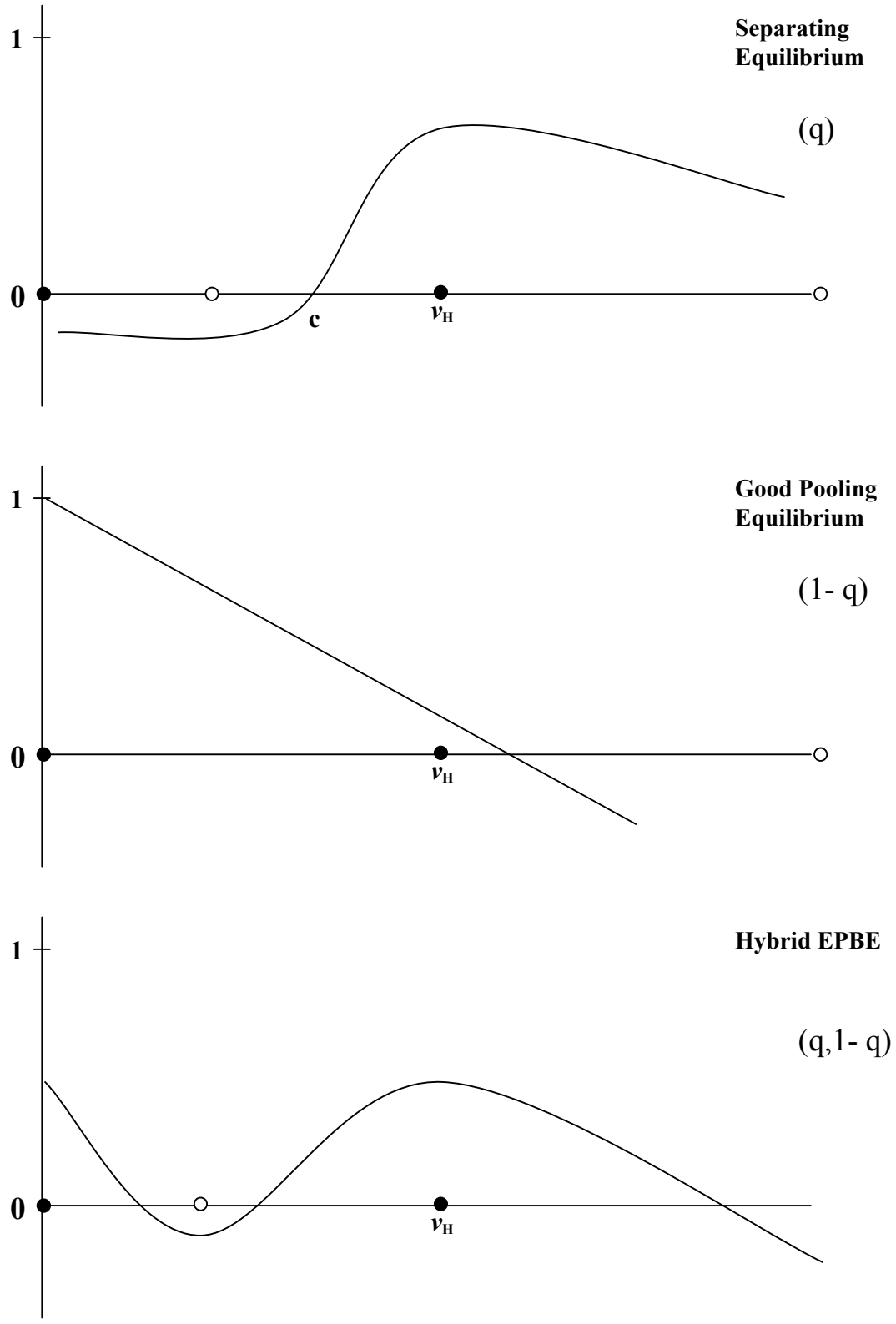
- Appealing features of EPBE are that it endogenizes crucial variables and selects among multiple equilibria. One often has a multiplicity of PBE that depend rather sensitively on arbitrary exogenous specifications of the types and their distribution. As we will illustrate shortly, EPBE can greatly reduce the equilibrium set while endogenizing the distribution over types.

5.2 New Ingredients for the Model

Returning to the reduced trust game, one soon notices that the Bad Pooling PBE evolves to the $x = 0$ extreme, which is a trivial EPBE. That is, when vengeful types are so rare that Other always tries to play D, then all types of Self try to play N. Now the less vengeful types are more fit, and the vengeful types become extinct, i.e., $x \to 0$. In equilibrium, the Self population consists entirely of $v = 0$ type Selfs who all try to play N, and the Other population always tries to play D regardless of signal. This is still a Bad Pooling PBE, and from Table 1 the payoffs are $-e(1-2e) < 0$ for Self and $e(2-e) > 0$ for Other. The strategy profile distributions in both the Self population and the Other population have support on a single point, so it is easy to check the equal and maximal expected fitness property that ensures EPBE.

Can there be another EPBE that supports mutual gains? We have seen that the separating and good pooling PBE are evolutionarily unstable, but thinking about their long run fate suggests a promising hybrid. Suppose that a fraction q of the Other population tries to play DC (i.e., D if $s = 0$ and C if $s = 1$) while the other $(1-q)$ of them tries to play CC (i.e., C regardless of perception). The fraction q must be such that unvengeful and vengeful Selfs achieve equal payoff, and the fraction x of vengeful Selfs must be such

Figure 6: Fitness *W* as a Function of Vengefulness *v*



that Others achieve the same fitness with DC as with CC. The intuition can be seen in Figure 6: by mixing the fitness landscape in panel A with that in panel B, we might hope to get the landscape in panel C with equal local (and global) maxima at $v = 0$ and v_H.

———fig 6 about here———

What strategy profiles should we look for? As before, Other's strategy set is $\{\Pr[C|s = 0] \in [e, 1 - e]\} \times \{\Pr[C|s = 1] \in [e, 1 - e]\}$, and now we are hypothesizing a distribution with support in this set consisting of the two points $(1 - e, 1 - e)$ (corresponding to CC) and $(e, 1 - e)$ (corresponding to DC) of respective mass $(1 - q)$ and q.

The matter is more complicated for Self. In the simple mix of equilibria in Figure 6, some unvengeful Selfs try to play T expecting C and a positive payoff, while the rest try to play N expecting D and a negative payoff. This is incoherent because the unvengeful Selfs who expect C are no more likely than those who expect D to match up with the Others who play CC. In equilibrium, unvengeful Selfs must either all try to play T, or all try to play N, or all must be indifferent. It can be shown that indifference is possible only if Self's fitness is 0, and in evolutionary equilibrium the vengeful type would also have fitness zero. Likewise, all unvengeful Selfs try to play N only when their fitness is negative. Since we are interested in mutual gains, we seek an evolutionary equilibrium with positive fitness and therefore where all unvengeful as well as vengeful Selfs try to play T. That is, the equilibrium strategy we seek involves $\Pr[T] = 1 - e$ for both types of Self.

Now we face a more fundamental complication: the value of v_H is no longer exogenous. The idea is that social and perhaps genetic forces shape Self's emotional response to violation of trust. In EPBE v_H maximizes fitness in an appropriate space of types, which we shall take to be $[0, v^{\max}]$. Here $v^{\max}$ is finite but large enough not to be a binding constraint in our analysis; see Friedman and Singh (1999) for a discussion that supports this assumption. In general one considers a distribution or measure over the space of types, but the EPBE distribution that we seek consists of an atom of mass x at some $v_H \in (0, v^{\max})$ with the remaining mass $(1 - x)$ at $v = 0$.

Finally, it no longer makes sense to regard the observational error amplitude a as constant. Extremely vengeful types should be easier than slightly vengeful types to distinguish from $v = 0$ types. Accordingly, we assume a perception technology $a = A(v)$ for the symmetric[6] probability of misperceiving

[6]Error symmetry is assumed for simplicity throughout the paper. Separate functions

Self's true type. We assume that $A(v)$ is a smooth, positive and decreasing function, with $A(v) \to 0$ as $v \to \infty$ and $A(0) = 1/2$. Thus the types cannot be distinguished in the limit as the vengeful type becomes completely unvengeful, and can be distinguished perfectly in the limit as the vengefulness becomes extreme.

Concrete results require a parametric form for the perception technology A. Our choice is a simple Gaussian function with precision parameter $k > 0$,

$$A(v) = 0.5 \exp(-kv^2), A' = -2kva. \tag{1}$$

5.3 Results

We seek a hybrid EPBE that supports mutual gains in the noisy reduced trust game. It will be characterized by values (v_H, a, q, x) with the following properties. First, $a = A(v_H)$ is the endogenous observational error rate between 0 and 1/2. Second, the preference parameter $v_H > c$ maximizes Self's expected fitness given the exogenous tremble rate e, the endogenous signal error amplitude a, and Other's equilibrium attention probability q, i.e.,

$$v_H = \arg \max_{v \in [c, v^{\max}]} \{E_q W^S(v|A(v), e)\}. \tag{2}$$

W^S is the maximal fitness Self can attain in the behavioral noise constrained strategy set $[e, 1-e]$, and we seek an equilibrium where it is attained at $(1-e)$, i.e., Self tries to play T. The restriction $v \in [c, v^{\max}]$ reflects the inefficacy of positive v less than c.

Third, applying the equal fitness principle to Self, the value of q must allow the unvengeful type to achieve the same (maximal) expected fitness as the vengeful type,

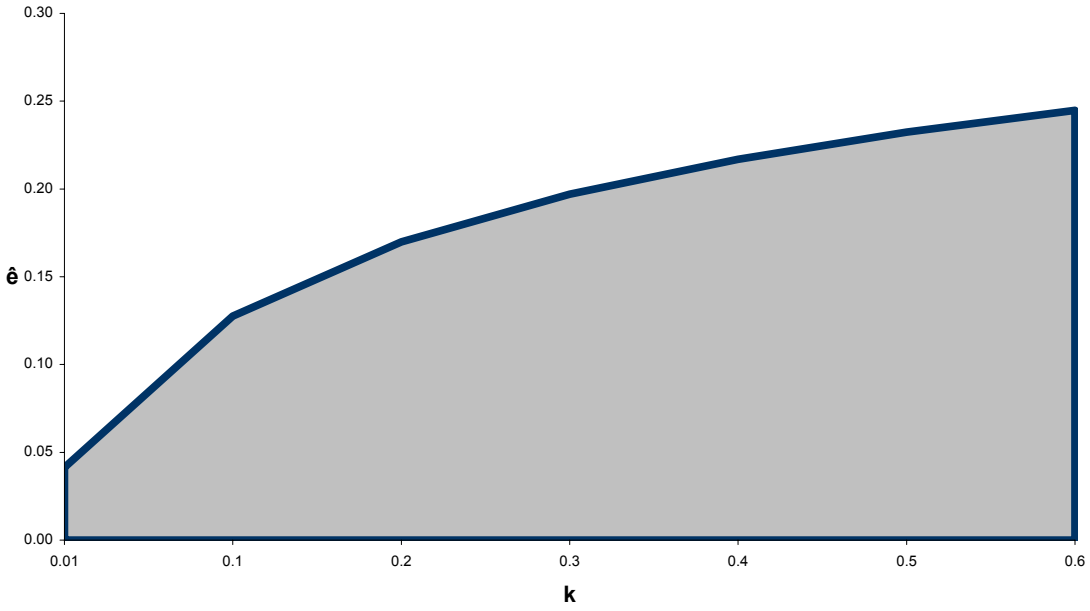
$$E_q W^S(0|A(0), e) = E_q W^S(v_H|A(v_H), e). \tag{3}$$

Finally, applying the equal fitness principle to Other, the fraction x of vengeful Selfs must allow Others who attend the signal to receive the same expected fitness as those who do not,

$$E_x W^O((1-e, 1-e)|a, e) = E_x W^O((e, 1-e)|a, e). \tag{4}$$

for type I and type II error rates would complicate the analysis but as far as we can see would not change our main conclusions. Equal behavioral error rates for Self and Other are assumed for the same reason.

Figure 7: Feasible Region of e and k



Here W^O denotes Other's fitness for the given strategy, usually abbreviated CC or DC. The principle also implies that W^O will be no higher for the unused strategies DD and CD.

The main result is that such efficient equilibria do exist and are unique over a wide range of the exogenous parameters. For example, the equilibrium exists for any choice of the precision parameter $k \in (0.1, 0.6)$, the tremble rate $e \in (0, .1)$ and the marginal punishment cost $c \in (0, 1)$.

————Figure 7 about here————

Proposition 2 establishes a larger feasible domain, shown in Figure 7. It is based on the following considerations. The parameter k is bounded above by the point $\bar{k} \approx 0.612$ where the second order condition for Self's fitness holds with equality. (By comparison, in the unit Normal distribution $k = 0.5$.) A finite value of $v^{\max}$ defines a lower bound; for example $v < v^{\max} = 10$ implies $k > 0.028$. Figure 7 shows that the upper bound $\hat{e}(k)$ for the tremble rate increases from about 10% to about 20% as k goes from about 0.1 to 0.6. The equilibrium exists over the entire natural range (0,1) of the marginal punishment cost parameter c. Higher values of c (for which Self's fitness reduction is larger than Other's) can tighten the upper bound on k due to the constraint $v_H > c$. For example, when $c = 2.0$ the upper bound is near $k = 0.3$.

Proposition 2. Given marginal punishment cost $c \in (0, 1)$, behavioral error rate $e \in (0, \hat{e}(k))$, and signal technology (1) with precision parameter $k \in (0, 0.6)$, there is a unique hybrid EPBE whose characteristics (v_H, a, q, x) depend smoothly on the exogenous parameters. There is also the trivial (Bad Pooling) EPBE with proportion $x = 0$ of vengeful types.

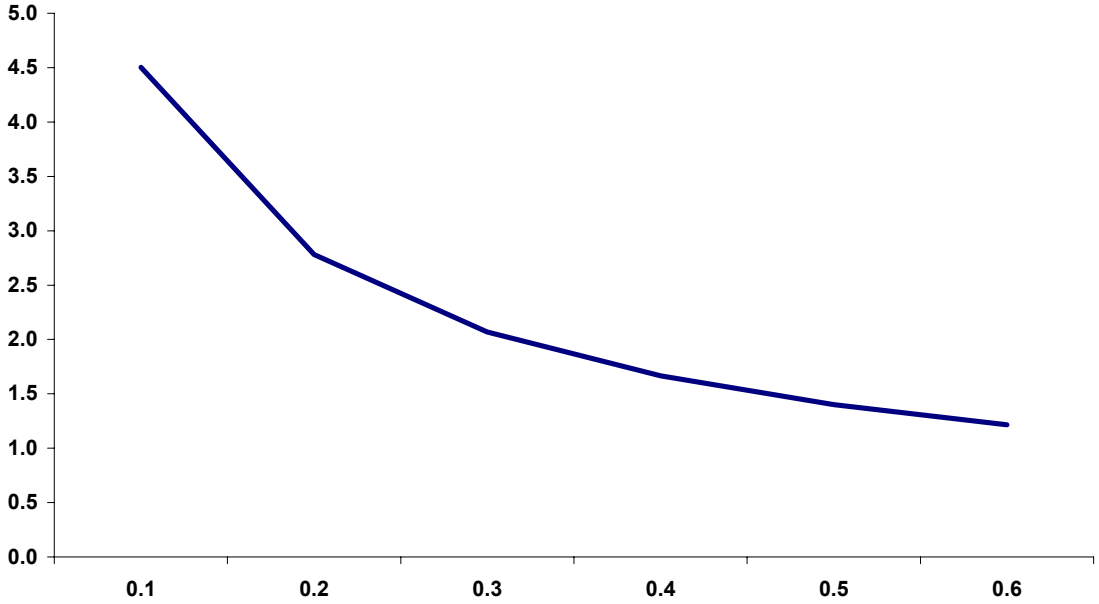
The proof again appears in the Appendix. It is constructive, and proceeds by writing explicitly three equations corresponding to the three EPBE necessary conditions listed above, solving them in terms of the exogenous parameters, and checking that no side conditions are violated for the given ranges of exogenous parameters. It turns out that the equilibrium values v_H and a depend on k but are independent of e and c, while q is independent of c, and x is independent of e.

————Figure 8 about here————

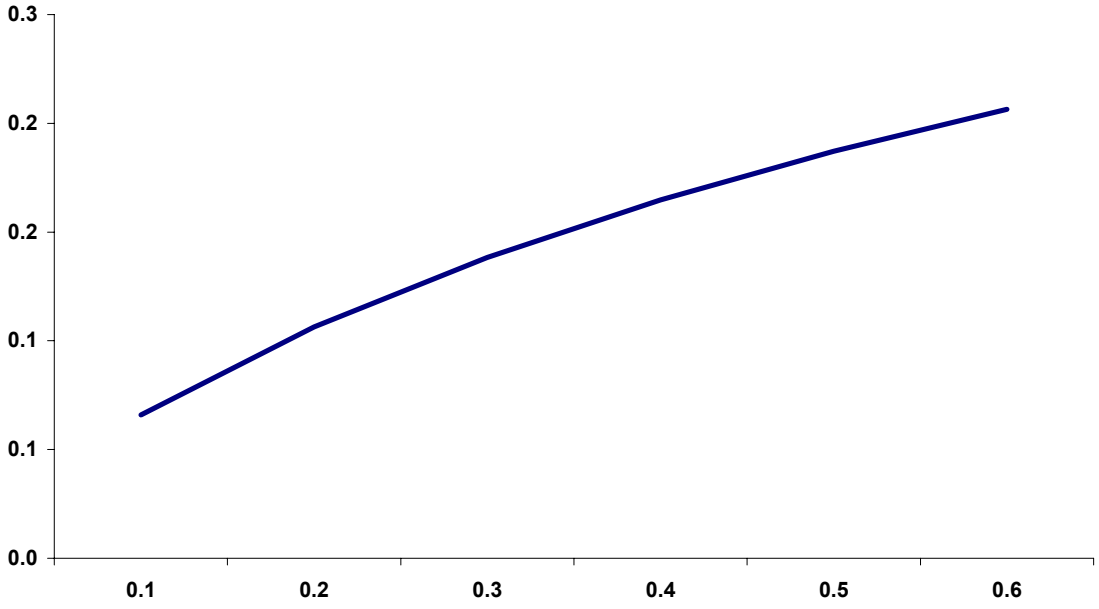
The comparative statics are given by the four functions $v_H = v^*(k)$, $a = a^*(k)$, $q = q^*(e, k)$ and $x = x^*(c, k)$, graphed in Figure 8. While proving Proposition 2, the Appendix also shows that $v^*(k)$ decreases in k, that is, the equilibrium level of vengeance always goes down as the precision of the

Figure 8: Comparative Statics Graphs

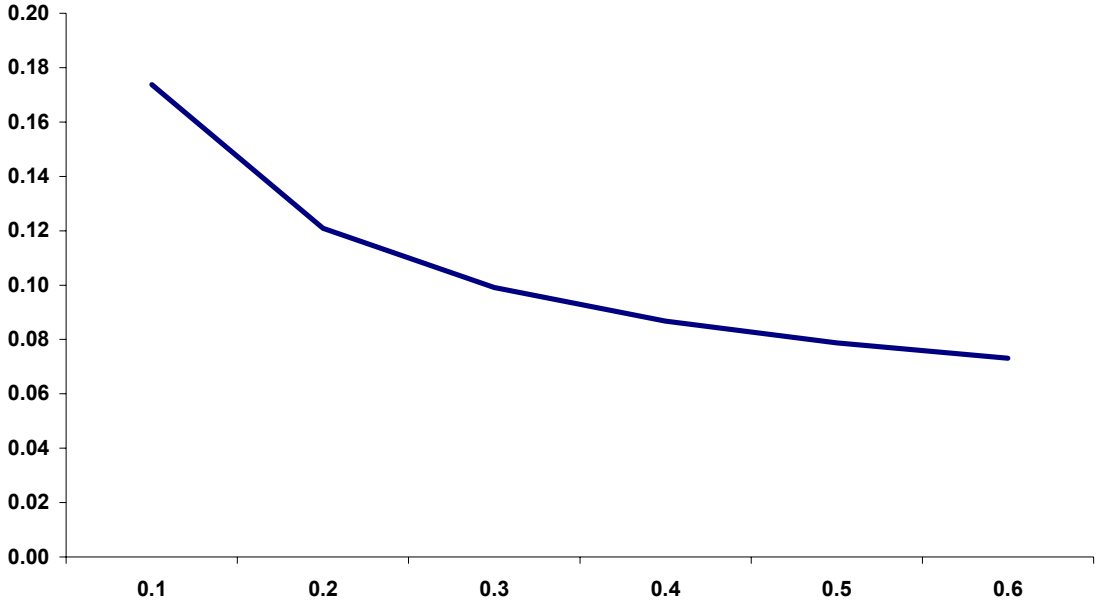
Graph of $v^*(k)$



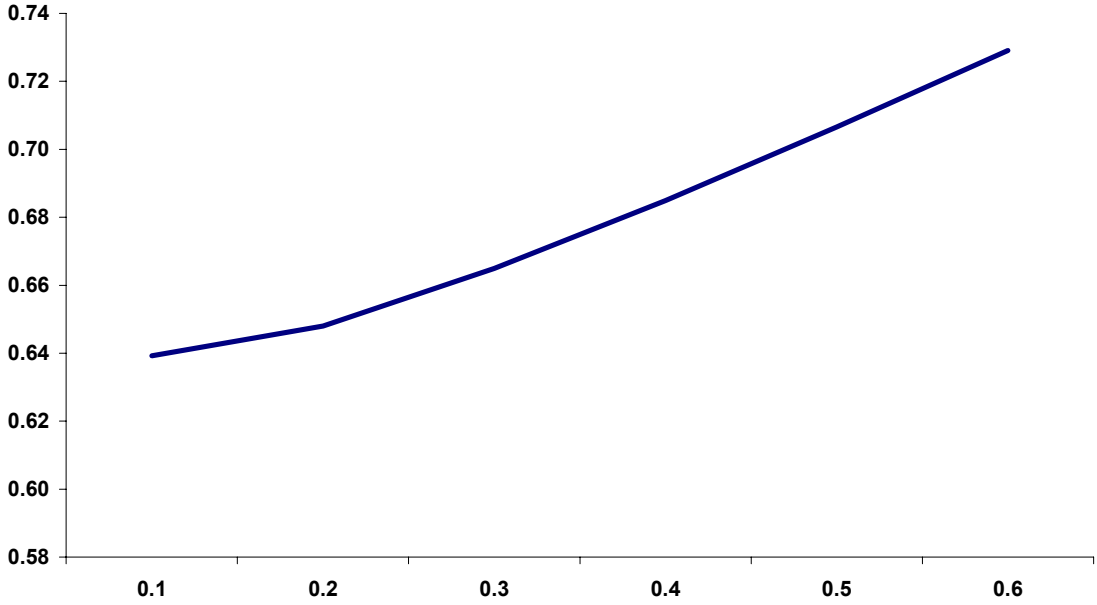
Graph of $a(k)$



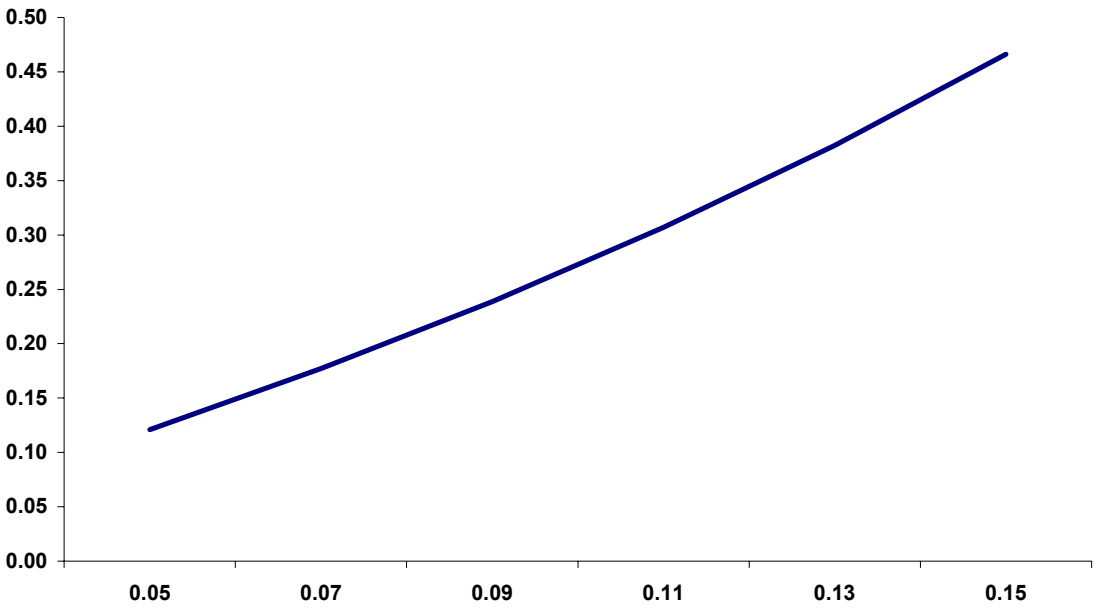
Graph of $q^*(0.05, k)$



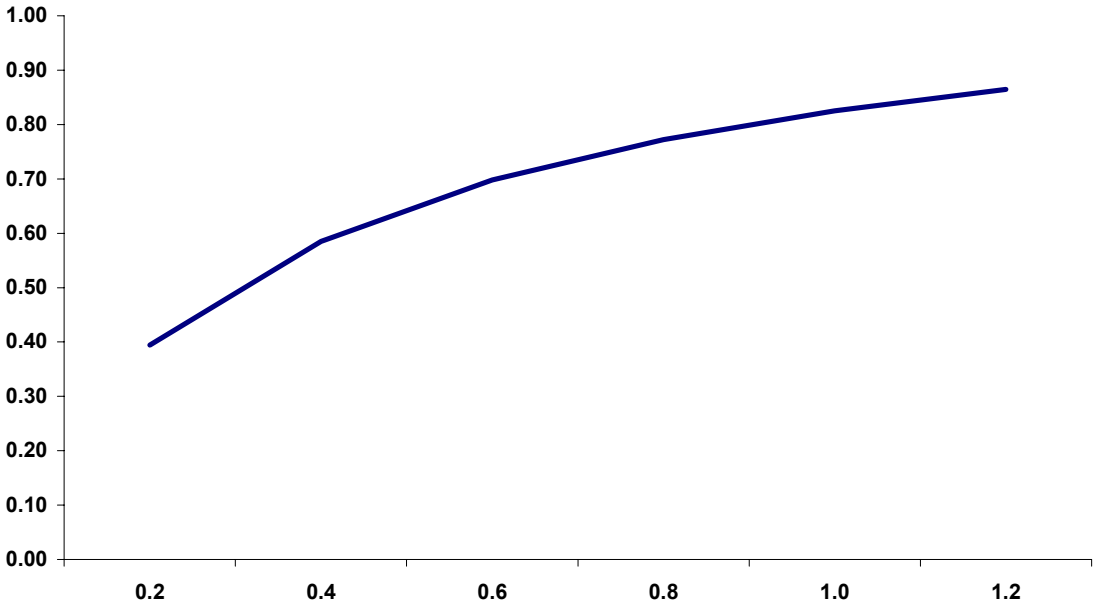
Graph of $x^*(0.5, k)$



Graph of $q^*(e, 0.2)$



Graph of $x^*(c, 0.2)$



observation goes up. Perhaps surprisingly, a^* is increasing in k, that is, the equilibrium observational error rate goes up as the precision increases. It turns out that the indirect effect via $v^*(k)$ dominates the direct effect of k.

How about the equilibrium rate at which Other attends to perceptions, q^*? It increases in the tremble rate e as a consequence of the Self's indifference condition (3), and it turns out to decrease in the precision of perceptions, k. Finally, the equilibrium fraction x^* of vengeful Selfs increases in the cost of punishment c as a consequence of the Other's indifference condition (4). However, the precision parameter k can have either a positive or negative effect on x^* depending on the level of c.

6 Discussion

We may summarize our work as follows. Economists need to come to grips with human motives such as vengeance. Since vengeance generally reduces own material payoff or fitness, the persistence of such motives is an evolutionary puzzle. We therefore construct a model in which a taste for vengeance survives in a long run evolutionary equilibrium.

Our model incorporates such tastes as what we call ESDUCs, or emotional state dependent utility components. The existence of ESDUCs is the proximate answer to the question of why individuals may want to harm (or help) others. However, the deeper questions of why certain ESDUCs exist and how they survive requires an analysis of their indirect fitness consequences. As noted in the introduction, this is called the indirect evolutionary approach. Studying vengeance is just one (interesting and complicated) application of the indirect evolutionary approach.

Our answer to the evolutionary puzzle proceeds in three stages. First, we construct a simple but representative situation in which ESDUCs matter, viz., an extended version of the Trust game. Second, we derive a perfect Bayesian Equilibrium, and characterize conditions under which pooling and/or separating equilibria exist. We note that fitnesses of different types of individuals (venegeful or not) are not equalized in the PBE, leaving room for evolutionary pressures to operate. The third stage, therefore, is to introduce a new long run equilibrium concept called evolutionary PBE, which allows adjustment in the proportion of vengeful types, as well as the intensity of their vengefulness. We characterize the EPBE for a nontrivial range of

parameter values.

At least three important issues remain. First, is the interesting EPBE dynamically stable? The answer may depend the specific form of adjustment dynamics. We have in mind gradient dynamics for continuous variables (like vengeance parameter v) and monotone dynamics for discrete variables (like the strategies C and D, or N and T) because these are standard in the biological literature (Eshel, 1983) and evolutionary games literature (Weibull, 1995). Our current conjecture is that, for a wide range of the basic parameters k (the signal precision parameter), e (the tremble or behavioral noise amplitude) and c (the marginal punishment cost),

- the trivial (and inefficient) EPBE will have an open basin of attraction and so will be the long-run fate of situations where most Selfs initially are not very vengeful and most Others try to play D, at least when they perceive $s = 0$.

- The hybrid EPBE (which reaps a major portion of the potential mutual gains) will have a one dimensional stable manifold and a two dimensional center manifold with a large basin of attraction.

A second issue springs from the trivial EPBE: how can one get a critical mass to escape from the trivial EPBE? Or more simply, in the context of the simple game in Figure 2, how can one get $v > c$ starting from $v = 0$? Friedman and Singh (2001) suggests a possible answer to this threshold problem. Subthreshold $v < c$ is not adaptive in a large population, but in small groups one can show that it works together with the discount factor δ to increase fitness. Thus positive values of v could get started in smaller groups and eventually become advantageous in larger groups.

Third, one should consider endogenizing the other parameters. Keeping punishment technology c constant (or doing comparative exercises) seems to make sense. The tremble rate parameter e trades off trivially against the endogenous probability q that Other attends to the signal, as can be seen from equations (8) and (5). However, there is every reason to take seriously the evolution of perception technology. A mutant Self with true vengeance parameter $v = 0$ who could somehow mimic the vengeful type would receive a major fitness boost. Friedman and Singh (2001) refers to this possibility as the Viceroy problem, a reference to Monarch butterflies that correspond to vengeful types and their mimics known as Viceroys. That paper sketches

an elaborate solution to the problem that involves interactions within and across small groups.

Eventually the theoretical work presented here should be applied to substantive problems in social science, and then must grapple with several additional issues. For example, we did not distinguish between genetic evolution and social evolution; their time scales and transmission mechanisms generally are quite distinct. We also considered only a single form of social interaction, the trust game, while in reality people play many different social games. Of course, our methods apply directly to any stable mix of games, and comparative statics apply to one-time shifts in the mix.[7] Continuing shifts in the mix of games played evidently require a truly dynamic analysis.

The present paper focused on two ideas, each of which we believe has widespread applicability independent of the other. Emotional state dependent utility components (ESDUCs) offer a tractable and flexible modelling approach to other-regarding preferences that is capable of dealing with many of the leading issues in behavioral economics. In particular, the vengeful components emphasized in the present paper may be the key ingredient in models giving new insights into "irrational" conflicts ranging from employment relations to international struggles. Friendly components may enter models providing insight into behavior within the family, teams, charitable giving, etc.

Our emphasis has been a discipline on such other regarding components: they must directly or indirectly bring evolutionary fitness. This theoretical discipline, together with the empirical discipline already favored by behavioral economists, should help to sharpen behavioral models.

The second idea is evolutionary perfect Bayesian equilibrium (EPBE). We wrote a general verbal definition and worked it out explicitly for a particular (and not especially simple) game of incomplete information. We believe that EPBE is an appropriate characterization of long run behavior when there are multiple "types" and some opportunity for entry, exit and/or switching among types. Many games of incomplete information could be reconsidered in this light.

[7] A caveat. A point equilibrium in the original mix defines the initial state following a shift (due perhaps to a regime change). If this initial state lies in the basin of attraction for the corresponding equilibrium for the new mix, then the overall change is described by comparative statics parallel to those accompanying Proposition 2. However, the comparative statics are misleading if the shift is large enough to put the initial state in a different basin of attraction.

7 Appendix A. Mathematical Details.

7.1 Preliminary computation of Self's fitness.

Figure 3 indicates that Self's fitness has a local maximum at $v = 0$, a global minimum near $c-\sigma$ and a global maximum near $c-\sigma$. An argument supporting this conclusion is as follows. Let $e \geq 0$ be the behavioral noise amplitude as in the text. Assume that it is small, in particular that $e < 1/(2 + v^{\max})$, where $v^{\max}$ is a finite upper bound on the vengeance parameter. To define the observational noise amplitude $\sigma \geq 0$, begin with a continuous random variable z located at zero (i.e., mean=mode=median=0) and otherwise arbitrary density function $h(z)$ and cdf $H(z)$. For example, z could have a uniform or a Normal distribution. Other perceives not Self's true vengefulness v but rather a noisy version $u = v + \sigma z$.

Key to the analysis is the probability $P(v)$ that Other will try to play D against Self with true parameter v, or equivalently, that c will exceed Other's posterior expectation of v. Computation is straightforward when Other has a uniform prior distribution for v. In this case, Other's posterior expectation of v is simply u and so $P(v) = \Pr[u < c|v] \equiv \Pr[\sigma z < c - v] = H(\frac{c-v}{\sigma})$. Then $P'(v) = -\sigma^{-1}h(\frac{c-v}{\sigma}) < 0$; its minimum is attained at $-\sigma^{-1}h(0)$ when $v = c$. Hence the inverse Mills ratio $P(v)/|P'(v)|$ attains a positive minimum of approximately $\sigma/(2h(0))$ near $v = c$. When Other has prior on v with positive density in the relevant neighborhood but otherwise arbitrary, the computation is much messier. It still can be shown that $P(v)/|P'(v)|$ attains a positive minimum of approximately $\kappa\sigma$ near $v = c$. (Now κ depends on the prior density as well as on h.) The approximations hold exactly in the limit as $\sigma \to 0$.

The preceding computation helps characterize the fitness function $W^S(v|\sigma, e)$. The probability that Other will actually play D (not just try) is $\alpha(v) = e + (1 - 2e)P(v)$, and so Self will achieve fitness $\beta(v) = 1 - (2 + v)\alpha(v)$ if she actually plays T and fitness 0 otherwise. Self will try to play N when $\beta(v) < 0$ and will try to play T when $\beta(v) \geq 0$. Thus

$$\begin{aligned} W^S(v|\sigma, e) &= e\beta(v) && \text{if } \beta(v) < 0, \\ &= (1 - e)\beta(v) && \text{otherwise.} \end{aligned}$$

When σ is small, $P(0) \approx 1$ and $\beta(0) \approx -(1 - 2e) < 0$, while for v moderately above c, $P(v) \approx 0$ and $\beta(v) \approx 1 - (2 + v)e > 0$. Suppose

that β has two regular critical points, one near $c-\sigma$ and the other near $c+\sigma$. Since $\beta'(0) = -(1-e) < 0$ we see that β and W^S slope downward from 0 to the first critical point, upward between the critical points, and downward beyond the second critical point. It follows that β is zero only at one point between the two critical points, and hence W^S indeed has the shape indicated in Figure 3.

Thus it remains only to verify the critical points. They are given by the first order condition (FOC) $0 = \beta' = 2\alpha + (2+v)\alpha'$. After straightforward algebraic manipulation the FOC can be rewritten as $P(v)/|P'(v)| = v/2 + (1-3e)/(1-2e)$. The Right Hand Side (RHS) of this last expression has slope $+1/2$ and $v=0$ intercept a bit below 1. As noted above, the LHS (the inverse Mills ratio) has a unique minimum near c and (since $P' \to 0$ as v moves away from c in either direction) and increases without bound on either side. The minimum value is of order σ so for σ sufficiently small there are exactly two regular solutions to the FOC and the verification is complete.

Figure 3 graphs W^S using the indicated values of e and σ, a uniform prior and the unit triangular density function for z.

7.2 Proof of Proposition 1

Proposition 1. Given perceptions with error rate a and choices with tremble rate e, and given types $v=0$ and $v=v_H > c$ constituting respectively Self population fractions $(1-x)$ and $x \in (0,1)$, assume that $0 < a, e < 1/2$ and $\alpha = a + e - 2ae \leq 1/(2+v_H)$. Then the separating PBE given in Table 1 exists iff $L(c/v_H) + L(e) - L(a) \leq L(x) \leq L(c/v_H) + L(e) + L(a)$; the Good Pooling equilibrium exists iff $L(x) \leq L(c/v_H) - L(a)$, and the Bad Pooling equilibrium exists iff $L(x) \geq L(c/v_H) + L(a)$.There is no PBE if $L(c/v_H) - L(a) < L(x) < L(c/v_H) + L(e) - L(a)$.

Proof. Recall from the text that Other is indifferent between C and D iff $E(v|s) = c$, and suppose first that Other observes $s=0$. Then (using the probabilities noted in the text) $c = E(v|s=0) = v_H \Pr[v = v_H | s = 0] + 0 = v_H[x(1-e)a/(x(1-e)a + (1-x)e(1-a))]$. Solving for x we obtain $x^s = 1/(1 + (\frac{a}{1-a})(\frac{1-e}{e})(\frac{v_H - c}{c}))$, which can be written $\frac{1-x^s}{x^s} = (\frac{a}{1-a})(\frac{1-e}{e})(\frac{1-c/v_H}{c/v_H})$. Recall $L(y) = \ln(\frac{1-y}{y})$, so $\ln(\frac{y}{1-y}) = -L(y)$. Hence Other is indifferent after seeing $s=0$ when $L(x^s) = -L(a) + L(e) + L(c/v_H)$, and prefers D when the prior odds $L(x)$ that $v = v_H$ are longer.

Other will see $s=1$ when $v = v_H$ with probability $x(1-e)(1-a)$

and with probability $(1-x)ea$ when $v=0$. Similar algebra shows that $L(x) \leq L(c/v_H) + L(e) + L(a)$ motivates Other to play C in this case. This verifies Other's part of the separating PBE.

Given that Other will try to play C iff $s=1$, Self with $v=v_H$ will face D with probability $\alpha = (1-e)a + e(1-a) = a + e - 2ae$ when playing T; a simple calculation shows that Self's expected payoff is nonnegative (and therefore she will try to play T) as long as $\alpha \leq 1/(2+v)$. Self with $v=0$ will face D with probability $1-\alpha$ when playing T; and she will avoid doing so as long as $\alpha \leq 1/(2+v) = 1/2$, a redundant condition. This completes the separating PBE verification.

If Other finds it worthwhile to ignore the $s=0$ signal because unvengeful types are quite rare, then those types will also try to play T. In the posterior probability calculation the expression $(1-x)e(1-a)$ thus is replaced by $(1-x)(1-e)(1-a)$ and the $1-e$ factors cancel. Calculations a bit simpler than those in the first part of the proof show that Other wants to play C even when $s=0$ iff $L(x) \leq L(c/v_H) - L(a)$. The condition ensuring that Self indeed wants to play T is the same as before taking $a=0$, so it holds *a fortiori*. This completes the Good Pooling PBE verification.

If Other finds it worthwhile to ignore the $s=1$ signal because vengeful types are quite rare, then those types will also try to play N. In the posterior probability calculation the factors involving e again drop out, and the condition $L(x) \geq L(c/v_H) + L(a)$ ensures that Other prefers to play D even when $s=1$. The condition $e < 1/2$ ensures that both types of Self prefer to play N when Other always tries to play D. This completes the Bad Pooling PBE verification.

Finally, if $L(c/v_H) - L(a) < L(x)$ then Other prefers to play D when $s=0$ and both types of Self try to play T. Hence unvengeful types' best response in this case is to try to play N. However, if unvengeful types try to play N, then Other's best response is to play C even when $s=0$ as long as $L(x) < L(c/v_H) + L(e) - L(a)$. Hence when both inequalities hold no PBE can exist. ♦

7.3 Proof of Proposition 2, and comparative statics

For convenience, the derivations of comparative statics are also included in the proof of the proposition. The proposition defines a parameter domain using the following functions: $R(k) = (kv(2+v) - \frac{1}{2})\exp(-kv^2)$, $B(k) = \frac{1}{(2+v)(1+a/R(k))}$ and $\hat{e}(k) = \min\{B(k), R(k)/(1+2R(k))\}$. These are functions

of the exogenous parameter k because in equilibrium v and a are specific functions (derived below) of k only.

Proposition 2. Given marginal punishment cost $c \in (0,1)$, behavioral error rate $e \in (0, \hat{e}(k))$, and signal technology (1) with precision parameter $k \in (0, 0.6)$, there is a unique hybrid EPBE whose characteristics (v_H, a, q, x) depend smoothly on the exogenous parameters. There is also the trivial (Bad Pooling) EPBE with proportion $x = 0$ of vengeful types.

Proof. Begin with the trivial case. Suppose there are indeed only unvengeful type Selfs. Then Other maximizes expected fitness by trying to play D irrespective of the signal. Self then maximizes expected fitness by trying to play N. Thus the strategy profile is a PBE. In this PBE, vengeful type Selfs will have lower fitness because of the extra cost they incur when inadvertently playing T; and Other lowers own fitness by trying to play C. Thus both populations have only a single type present and that type achieves maximal fitness, so the the strategy profile is indeed an EPBE.

We now construct the desired hybrid EPBE. The first and most laborious step is to derive the equilibrium value of v_H for a given k. Using the probabilities in Table 1, one sees that vengeful types trying to play T attain fitness $E_q W^S(v) = (1-e)[q(1-\alpha-(1+v)\alpha) + (1-q)((1-e)-(1+v)e)]$ $= (1-e)[1-(2+v)e - qa(2+v)(1-2e)]$. The first order condition (FOC) $0 = dE_q W^S/dv$ for the maximization problem (2) simplifies slightly to $0 = -e - qA'(2+v)(1-2e) - qa(1-2e)$ or, separating variables,

$$[\frac{e}{1-2e}]q^{-1} = -(2+v)A' - a. \tag{5}$$

The second order condition is $(2+v)A'' + 2A' \geq 0$. Substituting in the A' expressions from (1), the FOC is

$$[\frac{e}{1-2e}]q^{-1} = [2kv(2+v)-1]a = (kv(2+v) - \frac{1}{2})\exp(-kv^2) \tag{6}$$

and the SOC is

$$kv^3 + 2kv^2 - \frac{3}{2}v - 1 \geq 0. \tag{7}$$

The EPBE condition (3) says that vengeful and unvengeful type Selfs coexist in the EPBE because they have equal fitness. Recall that $E_q W^S(v)$ $= (1-e)[1-(2+v)e - qa(2+v)(1-2e)]$. Recall also that we are looking

for an EPBE in which even the unvengeful try to play T, so $E_qW^S(0) = (1-e)[q(\alpha-(1-\alpha)) + (1-q)((1-e)-e)] = (1-e)[1-2e-q(2-2a-4e+4ae)]$. Thus (3) reduces to $ve = q[2(1-2\alpha) - av(1-2e)] = q(1-2e)[2-a(4+v)]$. Separating variables again, we obtain

$$[\frac{e}{1-2e}]q^{-1} = (2-a(4+v))/v. \tag{8}$$

Note that (6) and (8) have the same left hand side. Equating the right hand sides, we get $2kv(2+v)a - a = (2-4a)/v - a$ or

$$kv^3 + 2kv^2 + 2 = 2\exp(kv^2) = 1/a. \tag{9}$$

This equation holds trivially for $v = 0$ and $a = 1/2$, but we now show that it also implicitly defines a candidate equilibrium level of vengefulness $v^*(k) > 0$.

Lemma 1. Equation (9) has a unique positive solution $v^*(k)$ for any positive k. The solution $v^*(k)$ decreases in k over the range where the second order condition (7) is valid.

Proof of Lemma. At $v = 0$ both sides of (9) are equal to 2, and have equal slopes of 0. The LHS has slope $4kv(1 + \frac{3}{4}v)$ and the RHS has slope $4kv\exp(kv^2) = 4kv(1 + kv^2 + ...)$. For small positive v (up to approximately $v = \frac{3}{4k}$) the LHS has steeper slope but the reverse is true for larger v (indeed, the slope ratio tends towards ∞). Hence RHS = LHS at some $v \approx \frac{3}{4k}$ (with this approximation being better for larger k and smaller v), so (9) indeed has a unique positive solution $v^*(k)$ for any positive k.

Implicitly differentiate (9) to get

$$v^{*\prime}(k) = -[v^3 + 2v^2 - 2v^2\exp(kv^2)]/[3kv^2 + 4kv - 4kv\exp(kv^2)]. \tag{10}$$

Use (9) to substitute for the exponential term and rearrange to obtain

$$-kv^{*\prime}(k)/v = [kv^2 + 2kv - 1]/[2kv^2 + 4kv - 3]. \tag{11}$$

The RHS of (11) is $[g + \frac{1}{2}]/[2g]$ for $g(k) = kv^2 + 2kv - \frac{3}{2}$. Rewrite the second order condition (7) as $g \geq 1/v$, and since $v > 0$, we have $g > 0$. Hence the RHS of (11) is positive. Since v and k are also positive, we conclude from (11) that $v^{*\prime}(k) < 0$ when the SOC holds.♦

We now show that the SOC (7) holds over the indicated range of k and is independent of the other exogenous parameters.

Lemma 2. Let $v = v^*(k)$ and $g(k) = kv^2+2kv-\frac{3}{2}$, and define $S(k) \equiv vg$. Then the equation $S(k) = 1$ has a unique solution $k = \bar{k} \approx 0.612$, and the second order condition (7) holds as an equality iff $k = \bar{k}$, and holds as a strict inequality iff $k \in (0, \bar{k})$.

Proof of Lemma. Write (7) as $S(k) \geq 1$. We first show that S strictly decreases in an open set U containing $S^{-1}[1, \infty)$. By direct computation we get $S'(k) = v^3 + 2v^2 + (v')(3kv^2 + 4kv - \frac{3}{2})$. Use (11) and simplify to write the RHS in the form $[vM]/[2kg]$, where v and k are positive and g is positive in U. The messy factor reduces to $M = -[(kv^2 + \frac{1}{2})g + \frac{3}{4}]$, which is strictly negative in U. Hence S indeed strictly decreases in U.

Use $v = O(1/k)$ from the proof of Lemma 1 to conclude that $S \to \infty$ as $k \to 0$ and $S \to 0$ as $k \to \infty$. Hence by the intermediate value theorem there is some $k \geq \varepsilon > 0$ such that $S(k) = 1$; let $\bar{k}$ be the smallest such k. We have $S'(\bar{k}) < 0$ and by the definition of U and continuity we have $S'(k) < 0$ $\forall k > \bar{k}$ s.t. $S(k) \geq 1 - \epsilon$. It follows that S is strictly bounded above by $1 - \epsilon$ on $(k + \delta, \infty)$. Therefore $\bar{k}$ is the unique solution to $S(k) = 1$ and the SOC fails for $k > \bar{k}$. Numerical solutions give $\bar{k} \approx 0.612$. ♦

Equations (6), (8) and (9) together with Lemmas 1 and 2 show that $v_H = v^*(k)$ and $v = 0$ indeed both maximize Self's fitness, and that $v_H = v^*(k)$ has the indicated comparative statics. The remaining steps in the proof are to find corresponding values of a, q and x, and to verify that all EPBE conditions and comparative statics hold. The perceptual error probability is simply $a = a^*(k) \equiv A(v^*(k))$. The text asserted that a increases in the precision parameter k. To verify, insert $v^*(k)$ into $A(v) = 0.5\exp(-kv^2)$ and differentiate to get $\frac{da^*}{dk} = -(2kvv'+v^2)A$. Use (11) to get $2kvv'+v^2 = v^2/(3-4kv-2kv^2) = -v^2/(2g) < 0$. Hence $\frac{da^*}{dk} > 0$ and the second comparative statics result is verified.

Next, obtain Self's mixing probability q from the left hand side of either (6) or (8). Use the right hand side of (6) with $v = v^*(k)$ to get the desired function of k only, $R(k) \equiv (kv(2 + v) - \frac{1}{2})\exp(-kv^2)$. $R(k)$ has the same sign as $2kv^2 + 4kv - 1 = 2g + 2$, which is positive over $(0, \bar{k}]$. It therefore makes sense to rewrite (6) as

$$q = q^*(e, k) \equiv \frac{e}{(1 - 2e)R(k)}. \tag{12}$$

Note that the condition $0 < e < \min\{1/2,\ R(k)/(1 + 2R(k)\}$ ensures that $0 < q < 1$. It is obvious from (12) that $q^*(e, k)$ is increasing in e. To show that $q^*(e, k)$ is decreasing in k, use (12) to write $q^* = \frac{e}{(1-2e)}\frac{\exp(kv^2)}{4(1+g)}$, differentiate

and simplify using (11). Eventually one obtains $\partial q^*/\partial k = \frac{ve}{(1-2e)}\frac{\exp(kv^2)}{8g(1+g)}[1 - vg-2g]$. All factors are positive except $[1-vg-2g]$, which is negative because $vg > 1$ by the SOC and $2g > 0$, so indeed $\partial q^*/\partial k < 0$.

Self's mixing probability x comes from the condition (4), which says that Others who attend the signal (strategy DC) coexist in EPBE with those that do not (strategy CC). Using Table 1 and simplifying, one finds that $\frac{1}{1-e}E_x W^O(\text{DC}) = 2-\alpha+x(-1+2\alpha-\frac{v}{c}\alpha)$ while $\frac{1}{1-e}E_x W^O(\text{CC}) = 1+e-x(\frac{v}{c}e)$. Equating these expressions, substituting $\alpha = a + e - 2ae$ and solving for x one obtains

$$x = \frac{1-a}{1+(\frac{v_H}{c}-2)a}. \tag{13}$$

Conditions already imposed, viz., $v_H > c > 0$ and $0 < a < 1/2$, ensure that $0 < x < 1$. Since a and v_H are independent of c, inspection of (13) reveals that x is increasing in c. Simulations show that x can be increasing or decreasing in k, depending on the value of c.

To verify that (v_H, a, q, x) defined by (9), $a = A(v^*(k))$, (12) and (13) constitutes the desired hybrid EPBE, we first need to confirm that vengeful Selfs (and a fortiori unvengeful Selfs) prefer T. With the tremble rate $e < 1/2$, an equivalent condition is that $E_q W^S(v)$ is positive, i.e., $1 \geq (2+v)e+qa(2+v)(1-2e)$ or $1/(2+v_H) \geq e+qa(1-2e)$. Note that the last inequality is the same as the corresponding condition in PBE except that $\alpha = e + a(1-2e)$ is replaced by an expression using qa instead of a. Hence the inequality is easier to satisfy than the PBE condition. Using (12), the right side of the inequality can be written $e+ea(1-2e)/(1-2e)R(k) = e(1+a/R(k))$. Thus a necessary condition for the good hybrid EPBE is

$$e \leq \frac{1}{(2+v)(1+a/R(k))} \equiv B(k). \tag{14}$$

Note that the positivity of v and $a/R(k)$ ensure that $B(k) < 1/2$. Combining (14) and the inequality following (12) we obtain a sharp bound on the tremble rate,

$$0 < e < \min\{B(k), R(k)/(1+2R(k)\} \equiv \hat{e}(k). \tag{15}$$

(It turns out that $\hat{e}(k) = B(k)$ for $0 < k < \bar{k}$.)

The rest of the proof is routine. Since the first order condition (6) has a unique solution in positive v, that solution $v_H = v^*(k)$ indeed solves the

maximization problem (2) whenever the second order condition (7) holds. We have already verified that the restrictions of k ensure that the second order condition is satisfied, indicated that c can be specified so that $v > c$, checked that x and q are well defined, and have built in the equal fitness condition for unvengeful and vengeful Selfs.

The only remaining chore is to verify that the unused strategies DD (always defect) and CD (perverse separating) do not increase Other's fitness. Using Table 1 and simplifying, one finds that $\frac{1}{1-e}E_xW^O(\text{DD}) = 2-e+x(-1+e)\frac{v}{c}$, so CC brings higher fitness as long as $0 < -(1-2e)+x(1-2e)\frac{v}{c}$,i.e., as long as $c < xv = X(k)$–this tightens the bounds on c, but since x approaches 1 as c approaches v, the effect is inconsequential. Since $v^*(k) > 1$ when $k = 0.6$, the restriction $c \in (0, 1)$ is easily sufficient for this final condition. Finally, comparing $\frac{1}{1-e}E_xW^O(\text{CD})$ to $\frac{1}{1-e}E_xW^O(\text{DC})$ term by term from Table 1, one sees that DC dominates as long as $\alpha < 1/2$ and $c < v$. ♦

References

[1] **Alchian, Armen.** "Uncertainty, Evolution and Economic Theory." *Journal of Political Economy.* 1950, 58, pp. 211-221.

[2] **Becker, Gary S.** "Irrational Behavior and Economic Theory." *Journal of Political Economy.* 1962, 70, pp. 1-13.

[3] ———**.** *The Economic Approach to Human Behavior.* Chicago: University of Chicago Press, 1976.

[4] **Bolton, Gary E. and Ockenfels**, **Axel.** "ERC: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, March 2000, 90(1), pp. 166-93.

[5] **Charness, Gary and Rabin**, **Matthew.** "Social Preferences: Some Simple Tests and a New Model." Discussion paper, University of California at Berkeley, 2001.

[6] **Cipolla, Carlo.** *The Basic Laws of Human Stupidity.* Bologna: The Mad Millers, 1976.

[7] **Cox, James C. and Friedman, Daniel.** "A Tractable Model of Reciprocity and Fairness." Manuscript, University of California at Santa Cruz, 2002.

[8] **Dufwenberg, Martin and Kirchsteiger**, **Georg.** "A Theory of Sequential Reciprocity." Discussion paper, CentER for Economic Research, Tilburg University, 1999.

[9] **Ely, Jeffrey C. and Yilankaya, Okan.** "Nash Equilibrium and the Evolution of Preferences." *Journal of Economic Theory*, 97, pp. 255-272, 2001.

[10] **Eshel, Ilan**. "Evolutionary and Continuous Stability." *Journal of Theoretical Biology*, 103, pp. 99-111, 1983.

[11] **Falk, Armin and Fischbacher, Urs.** "Distributional Consequences and Intentions in a Model of Reciprocity." *Annales d'Economique et de Statistique*, 63-64 (Special Issue), July-December 2001.

[12] **Fehr, Ernst and Schmidt**, **Klaus M.** "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, August 1999, 114(3), pp. 817-68.

[13] **Frank, Robert.** *Passions within Reason: The Strategic Role of the Emotions*, New York: WW Norton, 1988.

[14] **Friedman, Daniel and Singh, Nirvikar.** "On the viability of vengeance." UC Santa Cruz Working Paper, 1999, http://econ.ucsc.edu/faculty/workpapers.html.

[15] **Friedman, Daniel and Singh, Nirvikar.** "Vengeful Preferences." Paper presented at the UC Davis conference on Preferences and Social Settings, May 18-19, 2001.

[16] **Friedman, Milton.** "The Methodology of Positive Economics." In *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953.

[17] **Fudenberg, Drew and Tirole, Jean.** *Game Theory*, Cambridge, MA: MIT Press, 1991.

[18] ——**-, and Maskin.** "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica,* 1986, 54:3, pp. 533-554.

[19] **Geanakopolis, John , Pearce, David and Stacchetti, Ennio.** "Psychological Games and Sequential Rationality." *Games and Economic Behavior*, 1989, 1, pp. 60-79.

[20] **Guth, Werner and Kliemt, Hartmut.** "Competition or Cooperation: On the Evolutionary Economics of Trust, Exploitation and Moral Attitudes." *Metroeconomica*, 1994, 45:2, pp. 155-187.

[21] **Guth, Werner and Kliemt, Hartmut and Peleg, Bezalel.** "Coevolution of Preferences and Information in Simple Games of Trust." Manuscript, Humboldt University Berlin, 2001.

[22] **Guth, Werner and Yaari, Menachem.** "An Evolutionary Approach to Explaining Reciprocal Behavior," in U. Witt, ed., *Explaining Process and Change-Approaches to Evolutionary Economics*. Ann Arbor, The University of Michigan Press, 1992.

[23] **Hamilton, William D.** "The evolution of social behavior." *Journal of Theoretical Biology*, 1964, 7, pp. 1-5.

[24] **Huck, Steffen and Oechssler, Jorg.** "The indirect evolutionary approach to explaining fair allocations." *Games and Economic Behavior*, 1999, 28, pp. 13-24.

[25] **Jacobsen, Hans Jørgen, Jensen, Mogens and Sloth, Birgitte.** "Evolutionary Learning in Signalling Games." *Games and Economic Behavior*, 2001, 34:1, pp. 34-63.

[26] **Kaufman, Stuart.** *The Origins of Order: Self-Organization and Selection in Evolution*, NY: Oxford U Press, 1993.

[27] **Levine, David K.** "Modeling altruism and spitefulness in experiments." *Review of Economic Dynamics*, 1998, 1, pp. 593-622.

[28] **Mullainathan, Sendhil and Thaler, Richard** "Behavioral Economics." *MIT Working Paper 00-27*, September 2000. To appear in textitInternational Encyclopedia of the Social and Behavioral Sciences.

[29] **Noldeke, Georg and Samuelson, Larry.** "A Dynamic Model of Equilibrium Selection in Signaling Markets." *Journal of Economic Theory*, 1997, 73, pp. 118-156.

[30] **Rabin, Matthew.** "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 1993, 83, pp. 1281-1302.

[31] **Rilling, James K., Gutman, David A., Zeh, Thorsten R., Pagnoni, Guiseppe, Berns, Gregory S., and Kitts, Clinton D.** "A Neural Basis for Cooperation." *Neuron*, 2002, 35, pp. 395-405.

[32] **Rubin, Paul H. and Paul, C.W.** "An Evolutionary Model of Taste for Risk." *Economic Inquiry*, 1979, 17, pp. 585-596.

[33] **Samuelson, Larry and Swinkels, Jeroen.** "Information and the Evolution of the Utility Function." Mimeo, University of Wisconsin, 2001.

[34] **Samuelson, Larry.** "Introduction to the Evolution of Preferences." *Journal of Economic Theory*, 2001, 97, pp. 225-230.

[35] **Smith, John Maynard and Price, G.R. "**The Logic of Animal Conflict." *Nature*, 1973, 246, pp. 15-18.

[36] **Sobel, Joel.** "Social Preferences and Reciprocity." Mimeo, University of California at San Diego, 2000.

[37] **Trivers, Robert.** "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, 1971, 46, pp. 35-58.

[38] **van Winden, Frans.** "Emotional Hazard Exemplified by Taxation-induced Anger" *Kyklos*, 2001, 54, pp. 491-506.

[39] **Weibull, Jorgen W.** *Evolutionary Game Theory.* Cambridge, MA: MIT Press, 1995.

[40] **Wittman, Donald.** "Why Democracies Produce Efficient Results." *Journal of Political Economy,* 1989, 97(6), pp. 1395-1424.

[41] **Wright, Sewall.** "Adaption and Selection," in L. Jepsen, G.G. Simpson, and E. Mayr eds., *Genetics, Paleontology, and Evolution.* Princeton, N.J.: Princeton University Press, 1949.

CESifo Working Paper Series

---

695 Hans Gersbach, Financial Intermediation and the Creation of Macroeconomic Risks, April 2002

696 James M. Malcomson, James W. Maw, and Barry McCormick, General Training by Firms, Apprentice Contracts, and Public Policy, April 2002

697 Simon Gächter and Arno Riedl, Moral Property Rights in Bargaining, April 2002

698 Kai A. Konrad, Investment in the Absence of Property Rights: The Role of Incumbency Advantages, April 2002

699 Campbell Leith and Jim Malley, Estimated General Equilibrium Models for the Evaluation of Monetary Policy in the US and Europe, April 2002

700 Yin-Wong Cheung and Jude Yuen, Effects of U.S. Inflation on Hong Kong and Singapore, April 2002

701 Henry Tulkens, On Cooperation in Musgravian Models of Externalities within a Federation, April 2002

702 Ralph Chami and Gregory D. Hess, For Better or For Worse? State-Level Marital Formation and Risk Sharing, April 2002

703 Fredrik Andersson and Kai A. Konrad, Human Capital Investment and Globalization in Extortionary States, April 2002

704 Antonis Adam and Thomas Moutos, The Political Economy of EU Enlargement: Or, Why Japan is not a Candidate Country?, April 2002

705 Daniel Gros and Carsten Hefeker, Common Monetary Policy with Asymmetric Shocks, April 2002

706 Dirk Kiesewetter and Rainer Niemann, Neutral and Equitable Taxation of Pensions as Capital Income, April 2002

707 Robert S. Chirinko, Corporate Taxation, Capital Formation, and the Substitution Elasticity between Labor and Capital, April 2002

708 Frode Meland and Gaute Torsvik, Structural Adjustment and Endogenous Worker Recall Probabilities, April 2002

709 Rainer Niemann and Caren Sureth, Taxation under Uncertainty – Problems of Dynamic Programming and Contingent Claims Analysis in Real Option Theory, April 2002

710 Thomas Moutos and William Scarth, Technical Change and Unemployment: Policy Responses and Distributional Considerations, April 2002

711 Günther Rehme, (Re-)Distribution of Personal Incomes, Education and Economic Performance Across Countries, April 2002

712 Thorvaldur Gylfason and Gylfi Zoega, Inequality and Economic Growth: Do Natural Resources Matter?, April 2002

713 Wolfgang Leininger, Contests over Public Goods: Evolutionary Stability and the Free-Rider Problem, April 2002

714 Ernst Fehr and Armin Falk, Psychological Foundations of Incentives, April 2002

715 Giorgio Brunello, Maria Laura Parisi, and Daniela Sonedda, Labor Taxes and Wages: Evidence from Italy, May 2002

716 Marta Aloi and Huw Dixon, Entry Dynamics, Capacity Utilisation and Productivity in a Dynamic Open Economy, May 2002

717 Paolo M. Panteghini, Asymmetric Taxation under Incremental and Sequential Investment, May 2002

718 Ben J. Heijdra, Christian Keuschnigg, and Wilhelm Kohler, Eastern Enlargement of the EU: Jobs, Investment and Welfare in Present Member Countries, May 2002

719 Tapio Palokangas, The Political Economy of Collective Bargaining, May 2002

720 Gilles Saint-Paul, Some Evolutionary Foundations for Price Level Rigidity, May 2002

721 Giorgio Brunello and Daniela Sonedda, Labor Tax Progressivity, Wage Determination, and the Relative Wage Effect, May 2002

722 Eric van Damme, The Dutch UMTS-Auction, May 2002

723 Paolo M. Panteghini, Endogenous Timing and the Taxation of Discrete Investment Choices, May 2002

724 Achim Wambach, Collusion in Beauty Contests, May 2002

725 Dominique Demougin and Claude Fluet, Preponderance of Evidence, May 2002

726 Gilles Saint-Paul, Growth Effects of Non Proprietary Innovation, May 2002

727 Subir Bose, Gerhard O. Orosel, and Lise Vesterlund, Optimal Pricing and Endogenous Herding, May 2002

728 Erik Leertouwer and Jakob de Haan, How to Use Indicators for 'Corporatism' in Empirical Applications, May 2002

729 Matthias Wrede, Small States, Large Unitary States and Federations, May 2002

730 Christian Schultz, Transparency and Tacit Collusion in a Differentiated Market, May 2002

731 Volker Grossmann, Income Inequality, Voting Over the Size of Public Consumption, and Growth, May 2002

732 Yu-Fu Chen and Michael Funke, Working Time and Employment under Uncertainty, May 2002

733 Kjell Erik Lommerud, Odd Rune Straume, and Lars Sørgard, Downstream Merger with Oligopolistic Input Suppliers, May 2002

734 Saku Aura, Does the Balance of Power Within a Family Matter? The Case of the Retirement Equity Act, May 2002

735 Sandro Brusco and Fausto Panunzi, Reallocation of Corporate Resources and Managerial Incentives in Internal Capital Markets, May 2002

736 Stefan Napel and Mika Widgrén, Strategic Power Revisited, May 2002

737 Martin W. Cripps, Godfrey Keller, and Sven Rady, Strategic Experimentation: The Case of Poisson Bandits, May 2002

738 Pierre André Chiappori and Bernard Salanié, Testing Contract Theory: A Survey of Some Recent Work, June 2002

739 Robert J. Gary-Bobo and Sophie Larribeau, A Structural Econometric Model of Price Discrimination in the Mortgage Lending Industry, June 2002

740 Laurent Linnemer, When Backward Integration by a Dominant Firm Improves Welfare, June 2002

741 Gebhard Kirchgässner and Friedrich Schneider, On the Political Economy of Environmental Policy, June 2002

742 Christian Keuschnigg and Soren Bo Nielsen, Start-ups, Venture Capitalits, and the Capital Gains Tax, June 2002

743 Robert Fenge, Silke Uebelmesser, and Martin Werding, Second-best Properties of Implicit Social Security Taxes: Theory and Evidence, June 2002

744 Wendell Fleming and Jerome Stein, Stochastic Optimal Control, International Finance and Debt, June 2002

745 Gene M. Grossman, The Distribution of Talent and the Pattern and Consequences of International Trade, June 2002

746 Oleksiy Ivaschenko, Growth and Inequality: Evidence from Transitional Economies, June 2002

747 Burkhard Heer, Should Unemployment Benefits be Related to Previous Earnings?, July 2002

748 Bas van Aarle, Giovanni Di Bartolomeo, Jacob Engwerda, and Joseph Plasmans, Staying Together or Breaking Apart: Policy-makers' Endogenous Coalitions Formation in the European Economic and Monetary Union, July 2002

749 Hans Gersbach, Democratic Mechanisms: Double Majority Rules and Flexible Agenda Costs, July 2002

750 Bruno S. Frey and Stephan Meier, Pro-Social Behavior, Reciprocity or Both?, July 2002

751 Jonas Agell and Helge Bennmarker, Wage Policy and Endogenous Wage Rigidity: A Representative View From the Inside, July 2002

752 Edward Castronova, On Virtual Economies, July 2002

753 Rebecca M. Blank, U.S. Welfare Reform: What's Relevant for Europe?, July 2002

754 Ruslan Lukach and Joseph Plasmans, Measuring Knowledge Spillovers Using Patent Citations: Evidence from the Belgian Firm's Data, July 2002

755 Aaron Tornell and Frank Westermann, Boom-Bust Cycles in Middle Income Countries: Facts and Explanation, July 2002

756 Jan K. Brueckner, Internalization of Airport Congestion: A Network Analysis, July 2002

757 Lawrence M. Kahn, The Impact of Wage-Setting Institutions on the Incidence of Public Employment in the OECD: 1960-98, July 2002

758 Sijbren Cnossen, Tax Policy in the European Union, August 2002

759 Chandima Mendis, External Shocks and Banking Crises in Developing Countries: Does the Exchange Rate Regime Matter?, August 2002

760 Bruno S. Frey and Lars P. Feld, Deterrence and Morale in Taxation: An Empirical Analysis, August 2002

761 Lars Calmfors and Åsa Johansson, Nominal Wage Flexibility, Wage Indexation and Monetary Union, August 2002

762 Alexander R. W. Robson and Stergios Skaperdas, Costly Enforcement of Property Rights and the Coase Theorem, August 2002

763 Horst Raff, Preferential Trade Agreements and Tax Competition for Foreign Direct Investment, August 2002

764 Alex Cukierman and V. Anton Muscatelli, Do Central Banks have Precautionary Demands for Expansions and for Price Stability? – Theory and Evidence, August 2002

765 Giovanni Peri, Knowledge Flows and Knowledge Externalities, August 2002

766 Daniel Friedman and Nirvikar Singh, Equilibrium Vengeance, August 2002