The Economics of Trust and Reputation: A Primer

Luís M B Cabral

New York University and CEPR

First draft: May 2002; This draft: June 2005

(alas, still preliminary!)

Preface

The purpose of these lecture notes is to provide an introduction to the economics literature on reputation. The current literature is quite vast and increasing at a fast pace. It is therefore prudent to start with a disclaimer that the present text is *not* a survey of the literature. I will present a series of simplified models that attempt to give a flavor of the main themes in the literature and some of its applications in economics. At some later stage (that is, in a future draft), I will also attempt to write more complete notes on the literature. But for the time being, the goal is limited to providing a brief introduction.

Despite its size, the literature is still relatively recent: more than half the papers I have looked at were published in the past ten years or so. Moreover, there are relatively few secondary sources, that is, what there is is mostly included in the original articles. For this reason, my main goal here is to "distill" some of the main ideas put forward in the original research. Specifically, my strategy is to develop a number of simplified models which, while lacking the generality of many of the original contributions, are sufficient to highlight the crucial points.

A first draft of this text was distributed in May of 2002 as support material for a lecture series at the University of Zurich. I am grateful to the host institution, and in particular to Armin Schmutzler, for their warm hospitality.

The present notes are work in progress. Comments, corrections, and suggestions of additional references to include are most welcome. Please send them to lcabral@stern.nyu.edu

Contents

1 Introduction

The terms "trust" and "reputation" are employed in the economics literature in many different ways (not to speak of daily usage). Although our purpose here is not to settle matters of terminology, it will be useful to start with a brief note on semantics.

There are essentially two mechanisms that lead to economic notions of trust and reputation. One is based on repeated interaction; we call it the bootstrap mechanism. The other one is based on Bayesian updating of beliefs; we call it the Bayesian mechanism. Although these two different mechanisms have indistinctly been referred to as "reputation" mechanisms, they are actually very different mechanisms. For reasons that hopefully will become clear as we move along, the term "trust" is relatively more appropriate for the bootstrap mechanism, leaving the term "reputation" for the Bayesian updating context.

Sometimes these mechanisms are referred to as moral hazard and adverse selection mechanism. However, we can have adverse selection models featuring the bootstrap mechanism and Bayesian mechanisms featuring moral hazard. In fact, many models of reputation involve both hidden actions and hidden information; but even then one can normally detect what the main driver is, and accordingly classify them as models of "trust" or "reputation."

In summary, we have essentially two mechanisms where reputation/trust takes place:

- Trust. This is the situation "when agents expect a particular agent *to do* something." Typical models feature moral hazard. The essence of the mechanism is repetition and the possibility of "puninshing" off-the-equilibrium actions.

- Reputation: This is the situation "when agents believe a particular agent *to be* something." Typical models feature adverse selection. The essence of the mechanism is Bayesian updating and possibly signalling as well.

I will try to keep this distinction throughout the text. In some cases, this will imply using the term "trust" when other authors have used the term "reputation." (An alternative way of stating the difference between the two mechanisms might be to refer to bootstrap reputation and Bayesian reputation.)

The economic applications of the notions of trust and reputation are numerous. For example, oligopoly competitors might trust each other regarding an implicit or explicit price fixing agreement; a developing country would like to be considered a trustworthy borrower; more generally, implicit contracts (employment, partnership, lending, etc) involve some amount of trust between contracting parties. Turning to reputation, governments would like to be perceived as rigorous on inflation; some incumbent monopolist have the reputation for being tough with respect to entrants; a seller would like to have the reputation for selling good quality products; and so forth.

The economics literature has addressed these and several other applications. In order to focus on the main mechanisms underlying the models of trust and reputation, I will consider a limited number of applications. Specifically, I will primarily focus on the notion of seller trustworthiness and reputation in experience goods markets, that is, situations when consumers can only observe quality ex-post (and even then with noise). In a few cases, I will also consider the problem of lending. In a future draft, Section 6 (Notes on the literature) will include applications the notions of trust and reputation to other economics issues as well.

The rest of the text is organized as follows. Sections 2 and 3 form the bulk of the lecture notes. They address the two main mechanisms I referred to above: the bootstrap mechanism and the Bayesian mechanism. Section 4 introduces a series of further topics in the economics of trust and reputation. Section 5 looks at empirical and practical aspects of reputation and trust. Finally, Section 6 includes notes on the literature that forms the basis of Sections 2 to 5.

2 Bootstrap models

The main idea in this section is that, if players interact frequently enough, then they can trust each other. In other words, frequent repetition of a game creates the possibility of equilibrium action profiles that would be impossible in a static context.

2.1 Repeated games, the folk theorem, and trust

Consider the game played between a seller and a buyer. The seller decides whether to put effort into offering a high quality product. Effort costs $e > 0$. If the seller makes effort, then the product it sells "works." If the seller does not make any effort (cost zero), then the product does not work. The buyer is willing to pay 1 for a product that works, zero for one that does not. Importantly, the buyer cannot observe, before buying, whether the seller put effort into the product it is selling.

Suppose that $e < 1$. This implies that it is efficient for the seller to offer a high quality product: the cost of doing so is less than the value for the buyers. However, selling a high-quality product is clearly not an equilibrium. Regardless of what the buyer believes the seller is going to do, the latter's best strategy is not to make any effort into selling quality products: by the time the seller finds out whether the product works or not, the seller will have received his payment. So the seller might as well save the cost e. Naturally, a rational buyer should take this into account and accordingly be unwilling to offer any positive price for the seller's product.

Consider now the a game played between an infinitely lived seller and an infinitely lived buyer (or a infinite series of one-period buyers who can observe history). In each period, seller and buyers play the game described above.

One possible equilibrium of this repeated game is simply the repetition of the one-period game considered before. But there may be more equilibria in the repeated game. Specifically, let us look for an equilibrium with the following structure: seller and buyer start in a "trust" phase, whereby the seller offers high-quality products and the buyer pays the expected value of a high quality product.[1] If in any period a product breaks down, then the game reverts to a "punishment" phase, whereby buyers stops buying (or offer zero).

[1] This corresponds to the assumption that the market is short on the seller's side; that is, there are two or more buyers competing with each other for the seller's product.

A seller's discounted payoff from keeping his trust is given by $(1-e)/(1-\delta)$. By cheating on quality, the seller gets 1 immediately but zero in the future. It follows that trust is an equilibrium if

$$\frac{1-e}{1-\delta} \geq 1,$$

or simply $\delta > e$.

More generally, one of the central results of the theory of repeated games, the folk theorem, states that, if players are patient enough, then any feasible, individually rational set of payoffs can be sustained as the Nash equilibrium of a repeated game. The folk theorem can be interpreted as a model of trust: *If players are patient enough (that is, if the future matters a lot), then there exist equilibria of mutual trust.* The basic idea is very simple and intuitive: there is a trade-off between a short-term gain from doing what is myopically best and the long-term loss from squandering trust (many people would say "reputation"—see the Introduction).

The applications of these basic idea in economics are multiple. A particularly important one is to oligopoly collusion. In fact, many of the extensions considered in the remainder of this section were first developed in the context of oligopoly collusion. However, as mentioned earlier, throughout the text I will attempt to apply the various models to the problem of seller trust and reputation.

2.2 The dynamics of trust

The model presented in the previous section assumes that the basic stage game (seller picks effort, buyer offers price) is the same in every period; and that the history of past play is perfectly observed by all players. In this section, I present two extensions of the basic model of trust. The first extension assumes that history is observed with noise (Section 2.2.1). The second extension (Section 2.2.2) considers the possibility the stage game changes over time.

2.2.1 Noisy signals

The model in the previous section assumes that past actions are perfectly observable. In reality, however, observability is frequently imperfect. Consider again the case of seller trustworthiness. In practice, even high quality products break down—sometimes due to consumer carelessness. In this context, it would be rather inefficient for buyers to boycott sellers forever whenever a product were to break down: no seller would survive such drastic treatment.

If drastic punishment may be an inefficient way to maintain trust, total absence of punishment won't do either: then we are back in the basic prisoner's dilemma. The solution must be somewhere in between. In this section, we consider optimal trust equilibria when there is noisy observation of players' actions. The main idea is that, even among honest players, occasional "loss of trust" may be the price to pay for creating and maintaining trust.

Let us continue with the case of seller trustworthiness introduced in the previous section. Suppose now that a high-quality product breaks down with probability α (as before, a low-quality product always breaks down).

Consider the following equilibrium of the repeated game. Seller and buyers start in a "trust" phase whereby the seller produces high-quality products at cost e and buyers pay $(1-\alpha)$ (the expected value once we factor in the probability a high-quality product will break down). If the product does break down, then the equilibrium switches to a "punishment" phase where the seller offers low quality products (for which buyers pay zero) during T periods, upon which they revert to the "trust" phase again.

Let V^+ be the seller's value along a trust phase and V^- the seller's value at the start of a "punishment" phase. We have

$$\begin{aligned} V^+ &= 1-\alpha-e+(1-\alpha)\,\delta\,V^+ + \alpha\,\delta\,V^- && (1)\\ V^- &= \delta^T\,V^+ && (2) \end{aligned}$$

The seller's no-deviation constraint during the cooperative phase is given by

$$V^+ \geq 1-\alpha+\delta\,V^-, \tag{3}$$

or, using, (1),

$$V^+ - V^- \geq \frac{e}{(1-\alpha)\,\delta}. \tag{4}$$

Solving (1)–(2), we get

$$V^+ = \frac{1-\alpha-e}{1-(1-\alpha)\,\delta-\alpha\,\delta^{T+1}} \tag{5}$$

Substituting (5) and (2) into (4) and simplifying, we get

$$\delta\,\frac{1-\delta^T}{1-\delta^{T+1}} > \frac{e}{(1-\alpha)}. \tag{6}$$

Notice that (6) is not satisfied for $T=0$ (the right-hand side becomes zero). As $T\to\infty$, the left-hand side converges to δ. Therefore, if $\delta > e/(1-\alpha)$ there exists a T such that (6) is satisfied.[2] Specifically, if $\delta > e/(1-\alpha)$ then the optimal equilibrium is the one that minimizes the value of T consistent with (6). In fact, from (5), V^+ is decreasing in T.

Notice that, while a T-period punishment does the job of achieving some level of seller quality with a simple mechanism, this is not the only possible equilibrium to achieve a given value of V^+. In fact, when deriving equation (3) we made no use of (2). In other words, the maximization problem is given by (1) and (3):

$$\begin{aligned} \max_{V^-} \quad & V^+ = 1-\alpha-e+(1-\alpha)\,\delta\,V^+ + \alpha\,\delta\,V^- \\ \text{s.t.} \quad & V^+ \geq 1-\alpha+\delta\,V^-. \end{aligned}$$

Since an increase in V^- increases V^+ and tightens the no-deviation constraint, the above inequality will be an equality at the optimum. Solving for V^+ and V^- we get

$$V^+ = \frac{1-\alpha}{1-\delta} - \frac{e}{(1-\alpha)(1-\delta)} \tag{7}$$

$$V^- = \frac{1-\alpha}{1-\delta} - \frac{e}{(1-\alpha)(1-\delta)\,\delta} \tag{8}$$

Therefore, any continuation phase that leads to this value of V^- leads to the same value V^+. One way of generating a V^- is the one we considered above: consumers stop purchasing for T periods and then the trust phase resumes, where T is such that (8) holds (I am ignoring integer constraints here). One alternative that seems plausible in the context of seller/buyer relations is to assume that, upon the observation of low quality, a fraction θ of consumers stop purchasing the product. If all consumers are identical and the price extracts the consumer surplus, then this is an equilibrium strategy on the part of consumers. All we need to do is then to find the minimum value of θ such that the no-deviation constraint holds.

2.2.2 Changing market conditions

The credit market is an interesting application of the the basic mechanism of bootstrap equilibrium trust. The difference with respect to seller trustworthiness is that the strategic party is now the "buyer" (the borrower); but the basic mechanism is the same.

[2]Notice the condition $\delta > e/(1-\alpha)$ generalizes the condition from the previous section, $\delta > e$, which obtains for the case $\alpha = 0$.

Specifically, consider the following simple model. In each period, the borrower has an investment opportunity that requires \$1 of capital. The lender's opportunity cost is r and the lending market is competitive, so the lender requires a repayment of $1 + r$. The project yields $1 + R$. The borrower has two choices: to pay back, in which case the borrower's payoff is $R - r$, or to run away with the capital borrowed, in which case payoff is 1.[3] Assume that $0 < R - r < 1$. In words, the project is viable but not profitable enough that investing is a dominant strategy for the borrower. In other words, if the game between lender and borrower is played only once then the borrower would prefer not to invest and not to pay back (in which case the lender would not lend).

Suppose that lender and borrower are in a repeated relationship whereby new investment opportunities arise each period. A simple bootstrap trust equilibrium is one where the lender will continue supplying one unit of capital each period so long as the borrower keeps paying back. The condition for this to be an equilibrium is that $\frac{R-r}{1-\delta} \geq 1$, or simply

$$\delta \geq \delta' \equiv 1 + r - R. \tag{9}$$

If $\delta < \delta'$, then the efficient solution (lending \$1 each period) is not feasible. Might there be a second-best equilibrium where there is *some* lending? To consider this possibility, suppose that the investment opportunity is divisible: any investment $l \leq 1$ leads to a payoff of $l(1+R)$. If $\delta < \delta'$, can the lender offer an amount lower than 1? The no-deviation constraint would now be given by

$$l\,\frac{R-r}{1-\delta} \geq l.$$

But this is equivalent to the previous condition. We thus conclude that, if $\delta \geq \delta'$, then full lending is possible in equilibrium; whereas, if $\delta < \delta'$, then no amount of lending can be part of a bootstrap equilibrium.

Suppose now that all periods are not equal. Specifically, in each period, with probability $\frac{1}{2}$ the borrower has an investment opportunity yielding R_H; and with probability $\frac{1}{2}$ an investment opportunity yielding R_L, where $R_H > R_L$ and $R = \frac{1}{2}(R_H + R_L)$. The particular realization of $\tilde{R}$ is observable at the time for deciding on a loan. Moreover, the distribution of $\tilde{R}$ is i.i.d. across periods.

[3] I assume that, if the borrower invests and defaults, then the lender liquidates the investment and recovers $1 + r$, whereas the borrower gets zero; whereas, if there is no investment, then the lender gets nothing back.

Given the assumption of i.i.d. shocks, the expected continuation value when in the trust phase is independent of the current state. Suppose the discount factor is sufficiently large that the no-deviation constraints are satisfied. Then, along the trust phase the borrower receives value

$$V = \frac{1}{2}\,\frac{R_H - r}{1-\delta} + \frac{1}{2}\,\frac{R_L - r}{1-\delta} = \frac{R - r}{1-\delta}. \tag{10}$$

The no-deviation constraints for the borrower are therefore given by $R_i - r + \delta V \geq 1$ $(i = H, L)$. The most binding of these two constraints is for $i = L$, so we have, substituting (10) for V in (9) and simplifying,

$$\delta \geq \delta'' \equiv \frac{1 + r - R_L}{1 + R - R_L}.$$

In order to understand the relation between δ'' and δ', it is useful to write $R_L = R - \epsilon$ and $R_H = R + \epsilon$. Then $\delta'' = \frac{1+r-R+\epsilon}{1+\epsilon}$. If $\epsilon = 0$ (no fluctuations), then not surprisingly we get $\delta'' = \delta'$. Moreover, δ'' is increasing in ϵ. We thus conclude that $\delta'' > \delta'$.

In words, fluctuations around the value R make it more difficult to maintain trust between lender and borrower. The intuition is that, since the payoff from investing is relatively lower in a low state, the temptation to walk away with the loan is relatively higher. While the average return over a number of periods is the same regardless of ϵ, the relevant binding constraint is the possibility of walking away with the loan in each possible state, not on an average state.

If $\delta \geq \delta''$, then we continue with the same equilibrium as before. Suppose however that $\delta' < \delta < \delta''$. In this case, there would be full investment if there were no fluctuations, whereas full investment is not possible under fluctuating market conditions. Is there an equilibrium with positive amount of lending under market fluctuations? The previous analysis for the certainty case might suggest a negative answer. However, we will now see that there may indeed exist equilibria with positive (though less than one) levels of lending.

Consider a value of δ slightly lower than δ''. Suppose that, in equilibrium, there is full lending in state H but only l_L is the low state. The no-deviation constraint in the low state is given by

$$l_L(R_i - r) + \delta\left(\frac{1}{2}\,\frac{R_H - r}{1-\delta} + \frac{1}{2}\,\frac{l_L(R_L - r)}{1-\delta}\right) \geq l_L$$

If $l_L = 1$, then this reduces to $\delta > \delta''$, which we know is not true. Suppose however that we lower l_L away from 1. The derivative of the right-hand side with respect to l_L is equal to 1. The derivative of the left-hand side, in turn, is less to 1. To see this, suppose that $(R_H - r)$ were also multiplied by l_L. Then the entire left-hand side would be multiplied by l_L and the derivative would be equal to 1 too. We thus conclude that, by reducing l_L, the right-hand side drops faster than the left hand side; and for l_L low enough the inequality is satisfied. We have thus found an equilibrium.

More generally, let l_i be the amount lent in state i (at this point, we may consider any number of states). The borrower's no-deviation constraint in state i is given by $l_i(R_i - r) + \delta V \geq l_i$, or simply

$$l_i \leq \frac{\delta V}{1 + r - R_i}. \tag{11}$$

The equilibrium value for the borrower, V, is increasing in l_i. Moreover, by the argument above, reducing l_i increases the slack in the ith constraint. It follows that, for each constraint (11), either $l_i = 1$ or $l_i < 1$ and the constraint binds as an equality. And since the right-hand side is increasing in R_i, we conclude that $l_i \geq l_j$ if $R_i > R_j$: in periods of lower market conditions, borrowers can't be trusted as much.

2.2.3 Free entry

Consider again the equilibrium proposed in Section 2.1. If trust is associated to a name and names can be costlessly created, then an optimal strategy for a seller would be to (a) create a name and the expectation that the high-quality equilibrium is to be played; (b) sell a cheap, low-quality product for a high price; (c) "exit" the market under the current name and (d) reappear next period under a different name. Naturally, consumers should know about this possibility, in which case they would not be willing to pay a high price in the first place. We conclude that the equilibrium of Section 2.1 falls through under the possibility of costless entry.

In order to make trust an equilibrium there must be some cost for a seller from creating a new name and restarting the trust process. One possibility is advertising. Suppose that the equilibrium is as follows. If sellers spend an amount A before the first period of their existence, then buyers coordinate on playing the trust equilibrium described in Section 2.1; otherwise, they won't. Again, we are using the power of coordination in creating bootstrap equilibria. Clearly, if A is sufficiently high then the job is done.

Suppose that there is a trust equilibrium as described above. By continuing to honor his trust, a seller expects a discounted payoff of $(1 - e)/(1 - \delta)$.

Alternatively, a seller could cheat consumers on quality, make a profit 1 this period, change its name, and restart the process next period. In this case, discounted future profit would be $-A+(1-e)/(1-\delta)$. We conclude that the seller's no-deviation constraint is given by

$$\frac{1-e}{1-\delta} \geq 1+\delta\left(-A+\frac{1-e}{1-\delta}\right)$$

or simply $A > e/\delta$. In words, if the initial advertising requirement is high enough, then sellers prefer to keep their name to squandering their reputation and start a new name.

Notice that, while I've treated initial investment as advertising, I am not assuming any informative or persuasive effects of this initial investment. In fact, the only purpose of advertising is to create an opportunity for "money burning." Any other form of money burning would do equally well in terms of creating the desired bootstrap equilibrium. For example, it could be the case that sellers are expected to sell at a lower price during first n periods.

■ **Building trust.** Why does it take time for agents (friends, trading partners, etc) to establish a trust relationship? One possible answer is that gradual trust building is precisely a way to solve the free-entry and exit problem considered above.

Consider again the credit market bootstrap equilibrium presented in Section 2.2.2. There, the borrower has a short-run incentive not to repay. However, the borrower finds it optimal to repay his loans so as to be able to borrow again in the future. But consider now the case of credit in developing countries with large populations. One problem with the above equilibrium is that the borrower could simply not repay and, in the future, borrow from some other lender. If lenders cannot easily communicate with each other, this strategy might indeed work out, breaking down the trust equilibrium considered earlier. A similar situation is found in online markets where it is relatively cheap to create a new identity. How can agents punish a guilty deviant if the latter can simply change his identity and continue as before?

In order for trust to be an equilibrium, we must impose some cost on borrowers to start a new relationship. Consider the lending model from Section 2.2.2. Suppose that equilibrium dictates that, during the first T periods, the borrower only gets a loan $l < 1$, whereas from T on it gets the full loan of \$1. In this context, running away with a loan implies that the borrower will need to start a new relationship with a new lender, going back to the low level l.

Let V^+ be the borrower's value when in the full lending phase, and V^- the borrower's value of a new relationship. We then have

$$V^+ = \frac{R - r}{1 - \delta} \tag{12}$$

$$V^- = \frac{1 - \delta^T}{1 - \delta}\, l\,(R - r) + \delta^T\, V^+ \tag{13}$$

The no-deviation constraint along the full-lending phase is given by

$$V^+ \geq 1 + \delta\, V^-. \tag{14}$$

Substituting (12)–(13) for V^+ and V^-, and simplifying, we get

$$l \leq \frac{\delta - (1 + r - R) - (R - r)\delta^{T+1}}{(R - r)\,\delta\,(1 - \delta^T)} \tag{15}$$

If $\delta < \delta' \equiv 1 + r - R$, then (15) has no solution. In fact, even if there were no threat of a borrower running away and finding a new lender, $\delta < \delta'$ implies that no equilibrium exists, as we saw in Section 2.2.2. For $\delta > \delta'$, there exist values of l satisfying (15). In fact, as $T \rightarrow$, the right-hand side of (15) converges to

$$\frac{\delta - (1 + r - R)}{(R - r)\,\delta},$$

which is positive. In order to find an optimal equilibrium, we would find value of l and T such that (15) holds as an equality. Notice that there are multiple pairs (l, T) that do the job. In fact, an optimal pair is one that maximizes V^- (the value starting from day 1). But the no-deviation constraint constraint (14), holding as an equality, implies a specific value of V^-.

3 Bayesian models

As I mentioned in the Introduction, one can think of reputation as the situation when agents (e.g., buyers) believe a particular agent (e.g., seller) *to be* something (e.g., a high-quality seller). In this section, I introduce models of reputation, understood in this sense. Typical models feature adverse selection. The essence of the reputation mechanism is then the Bayesian updating process by the uninformed party (buyers); and possibly signalling by the informed party (seller).

I start the section with a very simple model that depicts seller reputation as a Bayesian updating process (Section 3.1). In Section 3.2, I make things a bit more complicated by allowing the seller to take actions that influence the probabilities which in turn influence the buyers' beliefs (investing in reputation). Finally, in Section 3.3 I study how sellers can buy, sell, discard or leverage their reputation (a name and its associated record) to their advantage.

3.1 Reputation as a Bayesian belief

Suppose that a seller can be of two types. A "good" seller's products work with probability α_H; a "bad" seller's products work with probability α_L, where $0 < \alpha_L < \alpha_H < 1$. As before, consumers are risk neutral and offer a price equal to their willingness to pay. A product that works is worth 1, one that doesn't is worth zero.

Let μ be the buyers' belief that the seller is good. Then buyers are willing to pay

$$p = \mu\,\alpha_H + (1-\mu)\,\alpha_L.$$

Clearly, p is increasing in μ, the belief that the seller is good. We may refer to this posterior as the seller's *reputation*. In fact, the greater a seller's reputation, the more buyers are willing to pay for its products.

Let μ_0 be the initial prior that the seller is good. Suppose that the seller's history consists of S sales when the product worked and F sales when it didn't. It follows that the buyers' posterior that the seller is good is given by

$$\mu = \frac{\mu_0\,\alpha_H^S\,(1-\alpha_H)^F}{\mu_0\,\alpha_H^S\,(1-\alpha_H)^F + (1-\mu_0)\,\alpha_L^S\,(1-\alpha_L)^F}.$$

In other words, for a given prior μ_0, there is a correspondence between a seller's record (number of Ss and number of Fs) and the sellers reputation (the value of μ).

Notice that μ is increasing in S and decreasing in F. So, after the sale of a product that works the seller's reputation, μ, increases, and so does sale price, p. And after the sale of a product that does not work the seller's reputation, μ, decreases, and so does sale price, p. By the law of large numbers, μ converges to 1 or 0, depending on the seller's type, and price converges to α_H or α_L, respectively.

At this stage, this is pretty much all that can be said about Bayesian reputation. Notice that in this simple model all of the action is on the buyers' side, through Bayesian updating of beliefs. The seller, in turn, does not play any active role. In the remainder of this section, we will focus on two areas where the seller may play an important role. First, in Section 3.2 we consider the case when the values α_i are not exogenous, rather a function of seller's effort. Then, in Section 3.3, we consider the possibility of trading reputations. For example, the seller might abandon his record (and its associated reputation), or better still, sell it; and then buy an existing record or start a new one from scratch.

3.2 Investing in reputation

One of the most fascinating issues in the study of animal behavior (including human behavior) is the dichotomy between nature and nurture. How much of what we do is the result of our genetic code (nature); and how much of it is a function of our environment, including the actions of those around us and our own past actions (nurture)?

In the context of models with asymmetric information, an analogous dichotomy is that of adverse selection and moral hazard. Under adverse selection, one player does not know the other player's *type* (nature); under moral hazard, its the other player's *actions* that are unobserved (nurture). The models in Section 2 are based on the latter paradigm. The model in Section 3.1 is based on the nature paradigm.

As it turns out, the richest case is the one that combines both adverse selection and moral hazard. In this case, a seller knows that his actions influence the outcome of today's transaction; and, through the buyers' Bayesian updating, the seller's actions today influence the seller's future reputation. In other words, combining moral hazard and adverse selection provides a framework in which we can study the seller's incentives to *invest in reputation*.

What are a seller's incentives to invest in reputation? How do they vary along the seller's lifetime? In this section I deal with these and related questions. In Section 3.2.1, I consider the case when sellers improve their reputation by pooling with "excellent" sellers. In Section 3.2.2, I deal with the case when improving reputation means showing one is not a "lousy" seller (separation). As we will see, even though there is a certain symmetry between these two cases, their implications are drastically different.

3.2.1 Pooling

Consider a seller who lives for $T < \infty$ periods. In each period, the seller offers a product that works with probability e, a probability that is chosen by the seller at a cost. Specifically, let $\phi(e)$ be the cost function of effort, which I assume has the following properties: $\phi(0) = 0, \phi'(0) = 0, \phi'(1) = \infty, \phi''(e) > 0$.

Buyers are willing to pay 1 for a product that works and zero for a product that does not work. Moreover, buyers are risk neutral, so the price is determined by the buyer's belief that the product will work.

Consider first the case when the seller lives for one period. In this situation, regardless of the price paid by the buyer (which is a function of the latter's beliefs), the seller's dominant strategy is not to make any effort. A rational buyer should know that, and so in equilibrium only low quality products are sold (for zero). Notice this is an inefficient solution: since $\phi(0) = 0$ and $\phi'(0) = 0$, it would be optimal for some positive quality level to be offered—namely the level that maximizes $e - \phi(e)$.

Suppose now that the seller were to sell in two consecutive periods. Again, the only equilibrium would be for the seller not to make any effort. A first period buyer should anticipate that, regardless of what happens in the first period, it will be in the seller's best interest not to make any effort in the second period. In other words, the second-period outcome is independent of what happens in the first period. The logic of the one-period game therefore applies and only the zero effort equilibrium remains. More generally, given any finite repetition of the one-period game, the only equilibrium is for the seller not to make any effort.

Suppose however that buyers are unsure about the type of seller they face. The seller may be of the type described above, which I will refer to as an "action" type, or the seller may be of a type who always sells products that work (a "good" type). Let μ_0 be the buyers' prior that the seller is of a good type. In what follows, I will refer to the buyers' belief that the seller is good as the seller's reputation.

A tantalizing possibility is that the action seller may now have an incentive to offer high quality in the first period in the hope of "passing by" a good seller and making a profit in the second period. I will now show that this is indeed a possibility.

Let us first consider the case when the seller lives for two periods, $T = 2$. Suppose that, at $t = 1$, buyers expect an action seller to make effort at level $\hat{e}$. Then they are willing to pay $\mu_0 + (1 - \mu_0)\,\hat{e}$ in the first period. If the product breaks down in the first period, then buyers will know they are facing an action seller; and, by the reasoning above, $e = 0$ in the second period. If the product works, however, then buyers update their belief that the seller is a good type. We now have

$$\mu_1 = \frac{\mu_0}{\mu_0 + (1 - \mu_0)\,\hat{e}}.$$

At $t = 2$, buyers expect an action seller to choose $e = 0$ (regardless of what happens at $t = 1$). Therefore, conditional on a product working in the first period, buyers are willing to pay μ_1 in the second period.

Putting all of this together, the seller's expected payoff is given by

$$V = \Big(\mu_0 + (1 - \mu_0)\,\hat{e}\Big) + (1 - e)\,0 + e\left(\frac{\mu_0}{\mu_0 + (1 - \mu_0)\,\hat{e}}\right) - \phi(e). \tag{16}$$

The first-order condition is given by

$$\frac{\mu_0}{\mu_0 + (1 - \mu_0)\,\hat{e}} = \phi'(e). \tag{17}$$

Given our assumptions regarding ϕ, there exists a unique interior solution to the first-order condition, which is a maximum.

An equilibrium is finally determined by the condition that $\hat{e} = e$. Notice that the left-hand side is strictly positive for $\hat{e} = 0$ and decreasing in $\hat{e}$. It follows there exists a unique solution to (17) such that $\hat{e} = e$.

Notice that the first term in the right-hand side of (16), which corresponds to current payoff, does not depend on e. The action seller's choice of e is an *investment in reputation.* Its payoff will come in the second period through (with positive probability) a higher belief by buyers—a better reputation.

Another important feature of (17) is that the left-hand side is increasing in μ_0. (To see this, notice the left-hand side is less than one and converges to one as $\mu_0 \to 1$.) Together with the arguments in the paragraphs above, this implies that the equilibrium level of e is increasing in μ_0. In words, the better the seller's initial reputation, the greater incentives the seller has to invest in reputation. The intuition is that, the higher μ_0, the more the seller has to lose from squandering his reputation.

Consider now a seller who lives for $T > 2$ periods. If ever a product breaks down, the seller will be identified as an action type and quality drops to zero. Let V_t be the seller's future expected value conditional on its product not having broken down before. By analogy with the analysis above, in the last period we have $V_T = \mu_{T-1}$ and $e_T = 0$. For $1 \le t < T$, V_t is given by

$$V_t = \Big(\mu_{t-1} + (1 - \mu_{t-1})\, e_t\Big) + e_t\, V_{t+1} - \phi(e_t),$$

whereas e is (implicitly) given by

$$V_{t+1} = \phi'(e_t).$$

Consider the difference between V_2 and V_3. If T is very large, the future looks similar starting from $t = 2$ or $t = 3$ except that the initial level of reputation μ_1 and μ_2 are different. In other words, if $\mu_1 = \mu_2$ then $V_1 \approx V_2$. But we know that

$$\mu_2 = \frac{\mu_1}{\mu_1 + (1 - \mu_1)\, e_2} > \mu_1.$$

Since $\mu_2 > \mu_1$ and V_2 and V_3 only differ with respect to μ, we conclude that $V_3 > V_2$: a seller's value increases after a second successful sale. We finally conclude that $e_2 > e_1$.

Let us summarize the reputation dynamics in this case. Initially, while the seller's record is made of product successes only, reputation increases and so does effort in quality. Whenever a product failure is observed, reputation shoots down and effort drops to the minimum. In other words, reputation and investment in reputation are positively correlated: both increase initially and drop eventually.

3.2.2 Separation

In this section, I consider a variation of the model with which I started Section 3.2.1. As before, the seller may be an "action" type, in which case the product works with probability e, a probability that is chosen by the seller at cost $\phi(e)$. But now, instead of assuming the other type is a good type, suppose it is a "bad" type instead, one who always produces bad products. Let μ be the buyers' belief that the seller is an "action" type. (Notice that in Section 3.2.1

μ is the belief the seller is "good." I am now denoting it as the belief the seller is an action type so that higher μ continues to denote better reputation.)

As before, consider a seller who lives for $T < \infty$ periods. First notice that, if the product ever works, then buyers will know that the seller is an action type. And, following the analysis in Section 3.2.1, no effort will be made thereafter: $e_t = 0$ for all future t. But if success leads to zero future profits, then the seller has no incentives to ever invest in quality. It follows that $e_t = 0$ *in every period!*

The contrast with respect to the previous section could not be greater. Before, we had a situation when the action seller wants to be taken as a good seller, that is, wants to pool with a good seller. This induced the action seller to invest in reputation, in fact, to invest more the higher the buyer's belief that the seller may be a good type. Now, we have a seller who, by investing in effort, will separate himself from a bad type. And separation, insofar as it leads to precise information, destroys the seller's incentives for investment.

The type space we considered so far is a bit extreme. In particular, an action type who makes no effort is as bad as a bad type. Suppose instead that $\phi(\underline{e}) = 0$, that is, even if the action seller makes no effort, its product works with probability $\underline{e}$. The situation is now different. As before, if buyers know for sure that the seller is an action seller, they know the seller has no incentives to invest in quality. However, differently from before, buyers are willing to pay $\underline{e} > 0$ for such a seller's product. This is more than what buyers are willing to pay for a bad seller's product (zero). An action seller thus has an incentive to be known as an action seller.

Things are a bit better, in the sense that, initially, an action seller has some incentive to spend a positive amount in quality. But once a product is successful those incentives disappear.

Suppose the seller lives for T periods. If a product works at time t, then $\mu_\tau = 1$ for all $\tau \geq t$. If instead the product has always failed, then the buyers' type beliefs are governed by the updating process

$$\mu_t = \frac{\mu_{t-1}\,(1 - e_t)}{\mu_{t-1}\,(1 - e_t) + (1 - \mu_{t-1})}. \tag{18}$$

If the seller has been identified as an action seller at time t (that is, $\mu_t = 1$), then its future value is given by $(T - t)\,\underline{e}$. In fact, the seller has no further incentives to invest in quality—and buyers know it. Regardless of his past record, the seller has no incentive to invest in the last period: $e_T = \underline{e}$, leading to a seller payoff of $V_T = \mu_{t-1}\,\underline{e}$. For $t < T$, let V_t be the seller's value at time t if all past transactions have been failures. V_t is obtained by maximizing

$$V_t = \mu_{t-1}\,\hat{e} + e_t\,(T-t)\,\underline{e} + (1-e_t)\,V_{t+1} - \phi(e_t).$$

with respect to e_t, where $\hat{e}$ is the buyers' belief regarding the seller's effort. The first-order condition (which is sufficient) is given by

$$(T-t)\,\underline{e} - V_{t+1} = \phi'(e_t). \tag{19}$$

Notice that V_t is increasing in μ_{t-1}, whereas the latter is decreasing in t (which follows from (18) and the fact that $e_t < 1$). Suppose that T is very large. By an argument similar to the previous section's, the main difference between V_2 and V_3 is that the initial priors μ_0 and μ_1 are different. But since $\mu_1 < \mu_0$, we conclude that $V_3 < V_2$. Finally, it follows from (19) that $e_2 > e_1$.

Let us summarize the reputation dynamics in this case. Initially, while the seller's record is made of product failures only, reputation declines, whereas effort in quality increases. Whenever a product success is observed, reputation shoots up, whereas effort drops to the minimum. In other words, reputation and investment in reputation are inversely correlated. Notice that this is the exact opposite of what happens when the action seller tries to pool with a good seller (Section 3.2.1).

■ **Changing types.** Our drastic results regarding separation and incentives for investment in reputation depend on a somewhat extreme set of assumptions. Once buyers know that a seller is an action type, the latter has no incentives for further investment: the seller has nothing more to prove.

But suppose instead that the seller's type is governed by a Markov process: each period, the seller type changes with probability λ. Suppose moreover that the seller lives for T periods. I will now argue that, in this setting, the seller has incentives to invest in quality in every period except the last.

Suppose that, along the equilibrium path, a product that works is sold at $t = 1$. Following the reasoning above, the seller would have no incentive to invest further. But given the λ transition probability, the subgame following the first period is equivalent to a $T-1$ lived seller and a buyers' prior $\mu_0 = 1-\lambda$ that the seller is an action seller. This suggests that even after buyers discover they have bought from an action seller, the latter's incentives to invest in reputation persist. Specifically, consider the case when $T = 3$. Then, following a product that works at $t = 1$, the subgame is equivalent to a $T = 2$ game without type changes and a prior $\mu_0 = 1 - \lambda$ that the seller is an action seller.

And from our previous analysis we know that first period effort by an action seller is positive.

In sum, the possibility of type changes keeps an action seller on his toes. Even after buyers find that they have bought from an action seller in period t, the latter continues having incentives to separate from a bad seller. As a result, positive investment will take place up to the last period.

3.3 Names and reputations

In his highly quotable *Afterthoughts* (1931), essayist Logan Pearsall Smith wrote that "Our names are labels, plainly printed on the bottled essence of our past behavior." As we saw in the previous section, the economic notion of reputation is the Bayesian belief that an uninformed party forms based on the past record of the informed party. Records of this sort are associated to names—Smith's labels. One is therefore led to think of the possibility of managing names for the value of the reputation they have attached. One can create names, sell names, buy names, discard names, and so forth. In this section, I deal with the economics of some of these transactions.

3.3.1 Changing names

Consider a market like eBay. Anecdotal evidence suggests that it is relatively cheap to change one's identity. What implications does this have for equilibrium prices and the seller's name management decisions (namely that of creating a new name)?

Consider the following pure-adverse selection model of product quality. Each period, a measure one of sellers is born. Each seller lives for two periods and can be of two types. Good sellers (fraction μ_0 of the population) succeed in their transactions (outcome S) with probability α_H. Bad sellers only succeed with probability $\alpha_L < \alpha_H$. Buyers value a successful transaction at 1, and an unsuccessful one at zero. Assuming in addition that buyers are risk neutral, they are willing to pay $\mu\,\alpha_H + (1-\mu)\,\alpha_L$, where μ is the belief that the particular seller they are dealing with is good. In particular, $\mu = \mu(\mu_0, h)$, where h is the history associated to that particular name.

Consider the following equilibrium. If a seller gets S in the first period, then he keeps the same name into the second period. If a seller gets an F in the first period, then he changes his name. Under this equilibrium hypothesis, a good seller exits with probability $1-\alpha_H$, a bad one with probability $1-\alpha_L$.

This implies that

$$\begin{aligned}
\mu(\emptyset) &= \frac{\mu_0\,(2-\alpha_H)}{\mu_0\,(2-\alpha_H)+(1-\mu_0)\,(2-\alpha_L)} \\
\mu(S) &= \frac{\mu_0\,\alpha_H}{\mu_0\,\alpha_H+(1-\mu_0)\,\alpha_L}
\end{aligned}$$

Notice that, if changing names were prohibitively costly, then

$$\begin{aligned}
\bar{\mu}(\emptyset) &= \mu_0 \\
\bar{\mu}(S) &= \frac{\mu_0\alpha_H}{\mu_0\alpha_H+(1-\mu_0)\alpha_L}
\end{aligned}$$

Straightforward computation shows that $\alpha_H > \alpha_L$ implies $\mu(\emptyset) < \bar{\mu}(\emptyset)$. The implication of inexpensive name changes is that there is a negative premium on sellers with new names.

3.3.2 Markets for names

To be completed.

3.3.3 Brand stretching

Suppose that sellers can be of two types. Good sellers always produce quality products which never break down. Bad sellers produce products that only work with probability α. Let μ_0 be the buyers' prior belief that a given seller is of a good type. In period 1 the seller is endowed with a new product and must decide how to name it. One possibility is to use its current brand name (to which the reputation μ is attached), a practice I will designate as "brand stretching." Another possibility is to create a new name. Suppose that, if the seller does not brand stretch, buyers have no way to know whether this is because the seller was not endowed with a new product; or whether, having been endowed with a new product, the seller decided to create a new name.[4]

Since buyers believe a genuinely new seller to be a good type with probability μ_0, they are willing to pay

$$p_0 = \mu_0 + (1-\mu_0)\,\alpha \tag{20}$$

for a new product.

[4]Technically, this is equivalent to assuming there is a continuum of sellers in the economy and only a finite number are endowed with a new product.

Suppose that, in equilibrium, a good seller always chooses to brand stretch, whereas a bad seller does so with probability ξ. If buyers hold these equilibrium beliefs, then their willingness to pay for a brand stretched new product is given by

$$p_1 = \frac{\mu}{\mu + (1-\mu)\,\xi}\,1 + \frac{(1-\mu)\,\xi}{\mu + (1-\mu)\,\xi}\,\alpha = \frac{\mu + (1-\mu)\,\xi\,\alpha}{\mu + (1-\mu)\,\xi}$$

Suppose first that $\xi = 1$. Then $p_1 = \mu + (1-\mu)\,\alpha$. The only difference with respect to (20) is that we have μ instead of μ_0. We thus conclude that, if $\mu > \mu_0$ and $\xi = 1$ and everything else is constant, then a seller is better off by brand stretching. The idea is straightforward: a seller with better than average reputation ($\mu > \mu_0$) is better off launching a new product under its valuable brand. We may refer to this as the direct reputation effect: other things constant, sellers with better reputation are more likely to brand stretch.

Now suppose that, in a second period, the firm sells a unit of its base product (to which the initial reputation μ was attached). Clearly, for a good type it is better to have stretched. In fact, since a good type always produces good products, its reputation must necessarily improve. Not necessarily so for a bad seller. If the stretched product works well, then stretching indeed improves period 2 payoffs. But if the stretched product fails, then the seller receives less than it would have received had it not stretched. We have a second effect here, which we may refer to as the reputation feed-back effect. Other things being equal, better sellers are more likely to stretch. Notice an important difference with respect to the direct reputation effect comparative statics: the latter are done with respect to seller reputation, the former with respect to seller quality (which are correlated but not necessarily equal).

The reputation feedback effect suggests that better sellers are more prone to stretch (everything else constant). This implies that, in equilibrium, $\xi < 1$. That is, whereas a good seller stretches with probability one, a bad sellers does so with probability less than one. But then the mere fact a seller brand stretches signals (if only partially) that it is a high quality seller; call it the signalling effect of brand stretching.

4 Further topics in the economics of trust and reputation

To be completed.

5 Trust and reputation in practice

To be completed.

6 Notes on the literature

Important note: The notes in this section are essentially limited to the ideas discussed in the present version of these lectures notes. They cannot therefore be considered a survey of the literature. There are dozens of articles on the economics of reputation which I do not include here. In a future version of the text, I hope to provide a more complete set of notes on the literature. The notes below follow the order with which topics are presented.

■ **Bootstrap models.** The bootstrap mechanism for trust is based on a general result known as the folk theorem (known as such because of its uncertain origins). For a fairly general statement of the theorem (and its proof) see Fudenberg and Makin (1986). One of the main areas of application of the folk theorem has been the problem of (tacit or explicit) collusion in oligopoly. This is a typical problem of trust (or lack thereof): all firms would prefer prices to be high and output to be low; but each firm, individually, has an incentive to drop price and increase output. Friedman (1971) presents one of the earliest formal applications of the folk theorem to oligopoly collusion. He considers the case when firms set prices and history is perfectly observable.

Both of the extensions presented in Section 2.2 were first developed with oligopoly collusion applications in mind. The case of trust with noisy signals (2.2.1) was first developed by Green and Porter (1984). A long series of papers have been written on this topic, including the influential work by Abreu, Pearce and Stacchetti (1990).

Rotemberg and Saloner (1986) proposed a model of oligopoly collusion with fluctuating market demand. In this case, the intuition presented in Section 2.2.2 implies that firms collude on a lower price during periods of higher demand. This suggests that prices are counter-cyclical in markets where firms collude. Rotemberg and Saloner (1986) present supporting evidence from the cement industry.

A number of papers have built on Rotemberg and Saloner's analysis. Kandori (1992) shows that the i.i.d. assumption simplifies the analysis but is not crucial. Harrington (19??) considers a richer demand model and looks at how prices vary along the business cycle.

The basic idea of repetition as a form of ensuring seller trustworthiness is developed in Klein and Leffler (1981). See also Telser (1980) and Shapiro (1983). When considering the problem of free entry, Klein and Leffler (1981) propose advertising as a solution, whereas Shapiro (1983) suggests low introductory prices.

Section **??** is based on my own research notes. The general analysis of self-reinforcing agreements when there is an outside option of the kind considered here may be found in Ray (2002). Watson (1999, 2002) also considers models where the level of trust stars at a low level and gradually increases.

■ **Bayesian models.** The seminal contributions to the study of Bayesian models of reputation are Kreps and Wilson (1982) and Milgrom and Roberts (1982). The model in Section 3.2.1 includes elements from these papers as well as from Diamond (1989). Hölmstrom (1982/1999) makes the point that separation leads to reduced incentives to invest in reputation. The issue of reputation with separation and changing types is treated in detail in the forthcoming book by Mailath and Samuelson (2006).

In Section 3.3, I presented a series of models that deal with name as carriers of reputations. The part on changing names (Section 3.3.1) reflects elements from a variety of models, though, to the best of my knowledge, no study exists that models the process of secret, costless name changes in an infinite period adverse selection context. The study of markets for names follows the work by Tadelis (1999) and Mailath and Samuelson (2001). All of these papers are based on the Bayesian updating paradigm. Kreps (1990) presents an argument for trading reputations in a bootstrap type of model.

The analysis of brand stretching (Section 3.3.3) is adapted from Cabral (2000). The paper considers a more general framework where the direct reputation, feedback reputation and signalling effects are present; and shows that better sellers are always more likely to brand stretch. The comparative statics with respect to the initial reputation level, however, are not obvious. As we saw above, a higher reputation firm can earn a higher direct reputation effect premium. But a higher reputation firm also has more to lose. The trade-off between using one's reputation and protecting it can go both ways. For other papers on brand stretching and umbrella branding see Choi (1998), Anderson (2002).

■ **Topics.** There are a number of ideas and papers I plan to look at in the topics section. These include Fishman and Rob (2005) and Phelan (2005) on the rise and fall of reputation; Ely and Välimäki (2003) on "bad reputations;" Tirole (1996) on collective reputations. Suggestions of additional papers are welcome.

Bibliography

ABREU, DILIP, DAVID PEARCE AND ENNIO STACCHETTI (1990), "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica* **58**, 1041–1064.

ANDERSSON, FREDRIK (2002), "Pooling reputations," *International Journal of Industrial Organization* **20**, 715–730.

BERNHEIN, B. DOUGLAS AND MICHAEL D. WHINSTON (1990), "Multimarket Contact and Collusive Behavior," *Rand Journal of Economics* **21**, 1–26.

CABRAL, LUÍS M B (2000), "Stretching Firm and Brand Reputation," *Rand Journal of Economics* **31**, 658-673.

CHOI, J.P. (1998), "Brand Extension and Informational Leverage," *Review of Economic Studies* **65**, 655–69.

DIAMOND, DOUGLAS W (1989), "Reputation Acquisition in Debt Markets," *Journal of Political Economy* **97**, 828–862.

ELY, JEFFREY C., AND JUUSO VÄLIMÄKI (2003), "Bad Reputation," *The Quarterly Journal of Economics* **118**, 785–814.

FISHMAN, A., AND R. ROB (2005), "Is Bigger Better? Customer Base Expansion through Word of Mouth Reputation," forthcoming in *Journal of Political Economy.*

FRIEDMAN, JAMES (1971), "A Noncooperative Equilibrium for Supergames," *Review of Economic Studies* **28**, 1–12.

FUDENBERG, DREW AND ERIC MASKIN (1986), "The Folk Theorem in Repeated Games with Discounting or with Imperfect Public Information," *Econometrica* **54**, 533–556.

GREEN, ED AND ROBERT PORTER (1984), "Noncooperative Collusion Under Imperfect Price Information," *Econometrica* **52**, 87–100.

HOLMSTRÖM, BENGT (1999), "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies* **66**, 169–182. (Originally (1982) in *Essays in Honor of Professor Lars Wahlback.*)

KANDORI, MICHIHIRO (1992), "Repeated Games Played by Overlapping Generations of Players," *Review of Economic Studies* **59**, 81–92.

KLEIN, B, AND K LEFFLER (1981), "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy* **89**, 615–641.

Kreps, David (1990), "Corporate Culture and Economic Theory," in J Alt and K Shepsle (Eds), *Perspectives on Positive Political Economy*, Cambridge: Cambridge University Press, 90–143.

Kreps, David M., Paul Milgrom, John Roberts and Robert Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* **27**, 245–252.

Kreps, David M., and Robert Wilson (1982), "Reputation and Imperfect Information," *Journal of Economic Theory* **27**, 253–279.

Mailath, George J, and Larry Samuelson (2001), "Who Wants a Good Reputation?," *Review of Economic Studies* **68**, 415–441.

Mailath, George J, and Larry Samuelson (1998), "Your Reputation Is Who You're Not, Not Who You'd Like To Be," University of Pennsylvania and University of Wisconsin.

Mailath, George J, and Larry Samuelson (2006), *Repeated Games and Reputations: Long-Run Relationships*, Oxford: Oxford University Press.

Milgrom, Paul, and John Roberts (1982), "Predation, Reputation, and Entry Deterrence," *Journal of Economic Theory* **27**, 280–312.

Phelan, Christopher (2001), "Public Trust and Government Betrayal," forthcoming in *Journal of Economic Theory.*

Ray, Debraj (2002), "The Time Structure of Self-Enforcing Agreements," *Econometrica* **70**, 547–582.

Rotemberg, Julio, and Garth Saloner (1986), "A Supergame-Theoretic Model of Price Wars During Booms," *American Economic Review* **76**, 390–407.

Shapiro, Carl (1983), "Premiums for High Quality Products as Rents to Reputation," *Quarterly Journal of Economics* **98**, 659–680.

Tadelis, S. (1999), "What's in a Name? Reputation as a Tradeable Asset," *American Economic Review* **89**, 548–563.

Tadelis, Steven (2002), "The Market for Reputations as an Incentive Mechanism," *Journal of Political Economy* **92**, 854–882.

Telser, L G (1980), "A Theory of Self-enforcing Agreements," *Journal of Business* **53**, 27–44.

Tirole, Jean (1996), "A Theory of Collective Reputations (with applications to the persistence of corruption and to firm quality)," *Review of Economic Studies* **63**, 1–22.

WATSON, JOEL (1999), "Starting Small and Renegotiation," *Journal of Economic Theory* **85**, 52–90.

WATSON, JOEL (2002), "Starting Small and Commitment," *Games and Economic Behavior* **38**, 176–199.

WERNERFELT, BIRGER (1988), "Umbrella Branding as a Signal of New Product Quality: An Example of Signalling by Posting a Bond," *Rand Journal of Economics* **19**, 458–466.