Pointing Fingers: Reputational Contract Enforcement and Mutual Recriminations

Peter Lorentzen*

Stanford University Graduate School of Business

November 2005

Abstract

Cooperative behavior in economic relationships is often enforced in part by reputational concerns—players behave well with one partner in order to ensure that they will be given opportunities to engage in profitable transactions with other partners in the future. However, players not involved in a particular relationship often cannot observe misbehavior by either partner. What they can see is whether the relationship breaks up or continues. This paper explores the consequences of this assumption in an environment of bilateral agency where the value of a relationship increases over time, and reputations are created by Bayesian updating over types. It shows that partners in a relationship will be less willing to enforce cooperative behavior than they would if outsiders could observe behavior within the relationship. Since breaking off the relationship is the only available punishment, to punish one’s partner is to punish oneself. Second, the opportunity to find new partners who are unfamiliar with a player’s record will diminish this reluctance to exit a bad relationship, possibly increasing attainable levels of cooperation. These results provide a possible reason for caution when attempting to increase transparency in economic environments suffering from weak formal institutions.

1 Introduction

Almost every economic transaction provides opportunities for one or both parties to cheat the other, improving his position at the expense of the partner. Where effective judicial institutions are in place, parties to a transaction can

*Stanford Graduate School of Business, 518 Memorial Way, Stanford CA 94305, USA; peterlorentzen@stanford.edu. Research support from the John M. Olin Program in Law and Economics, and from the Center on Democracy, Development, and Rule of Law at the Freeman Spogli Institute of International Studies is gratefully acknowledged.
write contracts and commit to mutually beneficial actions. Yet in many contexts such institutions are incapable of verifying contract compliance or enforcing judgments. This problem is particularly acute among small firms in developing and transitional economies. Indeed, economists have identified the inability to support mutually beneficial transactions as one of the fundamental institutional problems leading to underdevelopment (Greif, 2005; Acemoglu and Johnson, 2005).

Informal contract enforcement through reputational mechanisms can mitigate this problem. Where cheaters run sufficient risk of being found out by members of their community and punished for their bad behavior, they can be kept on the straight-and-narrow (Kandori, 1992). As a consequence, it has been proposed that societies can benefit from the flow of gossip (Dixit, 2003) and that its formalization in private organizations that can adjudicate disputes and pass on information about deviators (Milgrom, et al., 1990). Yet in practice, gossip is often inaccurate, and even formal courts in developing countries can be both slow and unreliable. Community members may be unable to identify which party is truly at fault in a dispute, simply that a dispute is taking place. In addition, these models focus on the limiting case where each trade occurs with a new partner. While the ability to safely exploit one-shot trading opportunities with a variety of partners is an important aspect of market expansion and development, it is also important for economic actors to maintain longer-term productive relationships.

This paper shows that an economy where disputes are publicly observable but fault is not can actually be less efficient than one where no information is available about past relationships. When each partner can point the finger of blame at the other, a wronged party will be reluctant to punish the partner by ending the relationship. Because outsiders cannot assign fault for the dispute, to punish one’s partner is to punish oneself. This in turn reduces incentives for cooperative behavior in the relationship. By a very similar logic, a market that is expanding may actually see higher levels of cooperation and fewer relationship breakdowns.

These issues are explored in a simple two-period model. In the first period, players are matched randomly to play a two-sided agency game with private monitoring. The population of players includes both good types who always exert full effort on behalf of their partner, and opportunistic types who incur the costs of cooperative behavior only to the extent it will benefit them in the future. After interacting with a partner in the first round and privately observing the success or failure resulting from the partner’s (unobserved) efforts, players can choose to stay with that partner, accruing the benefits of experience, or leave that partner to find a new one. Potential second-round partners include both other players whose first-round relationships broke up, and new entrants. Two different information regimes are considered, "clean slates," and "mutual recriminations." The benchmark case is "clean slates." In this case we assume

\footnote{The other core problem they raise is the inability of governments and their agents to commit to avoid expropriating their citizens.}
the community knows nothing about a player’s past history, so rematches are anonymous and random.\textsuperscript{2} We then consider what happens if there are mutual recriminations. That is, the existence and breakup of past relationships is observable, but potential rematching partners cannot attribute fault for the breakup.

The main results are as follows. Partners are reluctant to punish one another by rematching, even following a bad outcome, because they will lose the benefits gained from continuing cooperation and they may or may not match with a better partner in the rematching market. The more closely opportunists mimic the behavior of good players, the less informative a bad outcome will be. This implies that for either information structure, no equilibrium can be sustained in which opportunists exert full effort on their partners’ behalf, even with constant marginal costs.

More importantly, this reluctance to punish means that more information flowing to the community may not be good. Anonymous rematching means that players who end their relationships can start again with a clean slate, giving them a relatively better chance of getting a good partner. On the other hand, if potential future partners can distinguish breakdowns from untainted new entrants, they will avoid the breakups. Breakup information has two effects on incentives within the relationship. First, it makes the punishment of a breakup harsher—the quality of a potential rematch will be lower than under anonymity. If players always left partners who failed, this would increase effort. However, the second effect of breakup information is to reduce the willingness of partners to break up following bad performance. This will reduce the incentives for effort for opportunists.

Similar effects come about if we consider the effects of increasing market size or turnover in the setting of anonymous rematching. An increase in the number of new entrants to the market will make breakups less costly, possibly enhancing cooperation. This contrasts with the more common finding that smaller, more tightly-knit communities are better at enforcing good behavior (e.g. Dixit 2003). In models with this property it has been assumed that what one player might know about another is simply whether or not he is guilty, whereas the assumption in this paper that information coming through the flow of gossip will be inaccurate in a particular way. Which model better fits reality is an empirical question that may have different answers in different settings.

This analysis suggests that institution-building might be even harder than it seems. Simply facilitating information flow among businesses, for example by encouraging the formation of business associations or informal networks, may be insufficient or even detrimental if this gossip cannot be substantiated and blame cannot be clearly assigned. Encouraging the use of a formal legal system that is slow or inaccurate could have similar negative consequences.

\textsuperscript{2}Ghosh and Ray (1996) and Rob and Yang (2005) analyze the sustainability of cooperation markets with anonymous rematching in infinitely repeated games, but do not explore alternative community information structures.
1.1 Mutual recriminations and learning-by-doing

Before proceeding to the model, I will further motivate the two key assumptions underlying the results discussed. The "mutual recriminations" assumption says that the existence of past relationships and disputes can be observed, but guilt cannot. The "learning-by-doing" assumption says that, conditional on good behavior, a partnership that persists will be more productive than a new partnership.

In a survey of Vietnamese small businessmen, a typical view expressed was "if you treat customers fairly when they have difficulties you will have that reputation. People will do business with you because they think you will not kill them when they have difficulties" (McMillan and Woodruff, 1999). This desire to avoid public conflict appears in a variety of economic environments. Participants in online auctions are reluctant to give negative feedback about transaction partners, for fear that the partner will "flame" them back (Baron, 2002). Where disgruntled partners can easily seek an authoritative judgment of fault, this will be less of an issue, but it can be highly salient when enforcement is informal, occurring through social networks that have no investigative power. African firms report that clients will "use small defects as an excuse to delay payment or renegotiate prices" (Fafchamps, 2004: 172). Such disputes are difficult or impossible for any outsider to resolve. Clay (1997) finds among early Californian merchants, the historical record contains "abundant commentary about allegedly untruthful allegations of misconduct."

This difficulty of assigning blame has practical consequences for social behavior and institution-building. An influential study of private order among cattle ranchers found that they were very reluctant to get in conflicts with their neighbors (Ellickson 1991). When they observed apparent bad behavior, they were very cautious and gradual in implementing punishments. When a rancher found it necessary to harshly punish neighbor’s misbehavior, he would do so in a fashion that would be costly to himself as well as to the neighbor, in an attempt to demonstrate that he was not simply trying to take advantage of the neighbor. In a very different context, an association of Mexican shoe manufacturers devoted substantial effort to investigating its own members in order to ensure that the breakdown of a customer’s relationships with one member firm need not preclude its dealing with other firms (Woodruff, 1997). Some firms in the association became known for having "trouble with everyone," leading other members to disregard their attempts to blacklist customers. This further highlights how problematic it is to assume that a community gossip network can easily assign fault for a breakup and provides an example of how this affects formal and informal institutional development.

In addition to neglecting this important informational feature, most theo-

\[3\] Note also that for this possibility to arise, there must be an element of mutual agency. That is, both partners must have opportunities to cheat the other. If agency were purely one-sided, the ending of the relationship would be sufficient to imply guilt. The examples provided above suggest that a two-sided model is more appropriate for many interactions.

\[4\] This phenomenon is explored and explained in a different framework in Bernheim et al, (2005).
ries of multilateral enforcement have focused on repeated, identical transactions. This approach highlighted the ability of communities to enforce good behavior even when the wronged individual would never be able to personally retaliate. Yet these spot-market transactions are only one possibility. Many long-term relationships increase in productivity over time through learning-by-doing and the increased ability to integrate operations, as is commonly noted in analyses of the Japanese auto industry (Helper and Levine, 1992). While the automotive industry is quite sophisticated, there can be substantial benefits to longer-term cooperation even in industries with what might appear to be much simpler technologies. A manufacturer in the garment industry reported: "If we have a factory that is used to making our stuff, they know how it’s supposed to look. They know a particular style. It is not always easy to make a garment just from the pattern, especially if we rushed the pattern. But a factory that we have a relationship with will see the problem when the garment starts to go together. They will know how to make the fabric to make it look the way we intended. A factory that is new will just go ahead and make it. They won’t know any better" (Uzzi, 1997). Merchants in Mexican California often failed to participate in communal punishment (ostracism) of their business partners because any alternative partner would lack important knowledge about the partnership (Clay, 1997). Such effects will be less important in industries with standardized transactions, such as the trade of commodities, but will be more important where close cooperation between buyer and supplier is needed.5

2 The Economy

2.1 Timing and actions

The economy consists of a continuum of players with measure $1 + \alpha$. Play has three phases: a first transaction round, a breakups and rematching phase, and a second transaction round. A mass of players with measure 1 enters the market in the first phase and is randomly matched into pairs. In the transaction rounds each player simultaneously undertakes an effort level $e$ on a task that benefits the partner. In the rematching phase, continuing players first choose simultaneously whether or not to continue with the same partner or to rematch. If either player chooses to break up, the relationship is terminated and both players enter the rematching market. The remaining mass $\alpha$ of new players enters the rematching market at the same time as the old players. In the rematching phase, each player declares a preference ordering over possible partners, based on the available information (to be discussed further below). Players are put in pairwise stable matches based on these preferences.6 All

5 This learning-by-doing is also distinct from endogenously increasing relationship-specific levels of trust as in Watson (1999), since it happens automatically. In practice there is probably always some of both.

6 For a noncooperative game leading to this outcome in a continuum with two types, see Appendix 1.
pairs (continuing and new) play a second transaction round, and the game ends.

2.2 Players, payoffs and costs

Effort in the transaction phase has cost $ce$, where $e \in [0, 1]$ is the effort level.\(^7\) A fraction $\gamma$ of the players are "good," constrained to choose $e = 1$, while the remaining fraction $(1 - \gamma)$ are "opportunists" who can choose $e \in [0, 1]$. This is the only difference between good types and opportunists. These efforts result in success, $S$, with probability $\epsilon p$, or failure, $F$, with probability $(1 - \epsilon p)$, where $p < 1$. If a player succeeds, his partner receives some payoff. If he fails, his partner receives 0. The agency problem is thus bilateral. Both sides are expected to put in their best efforts and adapt to unforeseen circumstances.\(^8\) $p < 1$ ensures that even good players will sometimes perform badly.

In the first round, the payoff for a success is normalized to 1. A crucial feature of the model, as discussed earlier, is that potential payoffs from cooperation are increasing over time. To capture this, we will make the following assumption.


We will also consider at times what happens when that assumption does not hold, so $A = 1$. The possibility of declining value from relationships will not be considered here.\(^9\)

Assumption 2 (Common shock): After playing in the first round, but before deciding whether or not to leave the current partner, both players in each partnership learn about an arbitrarily small common shock to the value of that partnership, $A$. Actual potential payoffs to continuing each relationship are distributed uniformly $U[A - \epsilon, A + \epsilon]$.

This assumption makes certain calculations more straightforward and eliminates possible multiple equilibria that could arise for some parameter values, as will be discussed more below. It does not affect the qualitative nature of the results.

2.3 Information

When matching in the first round, players know only their own types and the probability of a randomly-selected partner being a good type, $\gamma$. In the trans-

\(^7\) Typically, we assume convexity in costs to ensure a solution. As we will see, this is unnecessary here. This cost structure can also be thought of as representing the situation where there are only two actions, "good" behavior at cost $c$, and "bad" behavior at cost 0 and a randomly chosen fraction $e$ of the opportunists choose good behavior.

\(^8\) Note that there may well be some fully observable and contractable dimensions of the transaction. We disregard these here in order to focus on the non-contractible dimensions. By requiring all players to match every round, I am implicitly assuming that the contractible value obtained from even a potentially bad relationship is high enough to make it better than any outside option.

\(^9\) Although see Sobel (2002) for a related model addressing exactly this possibility.
action phase, after choosing their effort levels, they see whether their partners’ efforts have resulted in a success or a failure. Monitoring is private, so neither player observes his own success or failure in the eyes of the partner.

The next two alternative assumptions are the most important to this paper.

Assumption 3 (Clean Slates): Players on the rematching market know nothing about any prospective partners.

Assumption 4 (Mutual Recriminations): On the rematching market, new players are distinguishable from old players who have left their first-round partners.

2.4 Strategies and Equilibrium

The solution concept is symmetric perfect Bayesian equilibrium. A strategy for each of the \((1 - \gamma)\) opportunistic players starting in the first round will consist of four elements:

1) A first-round effort level \(e_1\).
2) A punishment rule consisting of a probability of rematching (punishing the partner) conditional on the partner’s success, \(\pi(S)\), and a probability conditional on the partner’s failure \(\pi(F)\).
3) A preference declaration over potential partners on the rematching market, based on whatever information will be available about them and the other players on the market.
4) A second-round effort level \(e_2\).

A strategy for the \(\alpha(1 - \gamma)\) opportunistic players starting on the second round will include only elements 3 and 4. Strategies for the \(\gamma\) good types starting in the first round need only specify elements 2 and 3, since their effort levels are predetermined, and the \(\alpha\gamma\) good second-round entrants will only need to specify element 3. Under assumption 3, Clean Slates, element 3 can be neglected, as potential rematching partners will be indistinguishable.

3 Base Case: Clean Slates

We will first explore the properties of this economy under assumption 2, anonymous rematching. Under this assumption, new entrants cannot be distinguished from players who have just broken up with their previous partners. Intuitively, this would be the case in a large market where there is no easy way to find out about any player’s past dealings. For instance, in the New York apparel business, "manufacturers can play hit and run for years before their reputation catches up with them." This possibility arises because there is a large number of firms and turnover in the industry is high (Uzzi, 1997). In addition, firms

\[10\] The symmetry assumption is not a strong one. Equilibrium effort levels \(e\) will be symmetric by type in any case. Incentives to break up will be identical for each player, given the actions of all other players, but where players are indifferent over the ending of relationships (pursuing mixed strategies) they could choose asymmetric randomizations.
may be reluctant to help potential competitors by helping them avoid deadbeats or identify good prospective partners (Uzzi, 1997; Fafchamps, 2004).

### 3.1 Equilibrium Analysis

We proceed through backward induction. First, it is clear that opportunist types will never cooperate in the second round of play, $e_2 = 0$, as there is no future to the relationship and $ce > 0$. The expected payoff for any player entering the second round is simply the probability of the partner being good times the expected payoff from a good partner’s efforts. A match with a new partner will yield a payoff of 1 with probability $p$ if the partner is a good type. The conditional probability of a partner being good given the available information we will denote $\theta(.)$. The expected utility of rematching with a new partner from the market will therefore be $p\theta^M$.

Defining $\theta^g$ to be the probability of a partner being a good type, conditional on that partner having succeeded in the first round, payoffs from staying with a success will be $A\theta^S$. The multiplier $A > 1$ is the consequence of Assumption 1, Learning-by-Doing. Analogously, the expected payoff from staying with a partner who failed the first time is $A\theta^F$. Note that since all opportunists will choose no effort, $e_2 = 0$, and all good types will choose $e_2 = 1$, $c(e_2)$ will be independent of the match for each player. It will not enter into the breakup or rematching decisions. Henceforth, we will neglect mention of second-period costs of effort and drop the subscript on first period effort $e$.

In this setting, all players on the rematching market will have no information with which to distinguish one potential partner from another, so they will be rematched randomly, with expected payoff $p\theta^M$. Now consider the choices facing old players when they decide between staying with the current partner or choosing a new one. First, if a player’s partner will end the relationship, the player’s decision is meaningless. Thus, there will always be an equilibrium where all players choose to end their first-round relationships because they know this decision is not pivotal. However, such an equilibrium is not robust to plausible refinements such as a small possibility of trembles, so we will instead focus on the equilibria where players choose to stay or go assuming that their decisions will be pivotal.

Note also that the partner’s decision to leave will not be an effective signal. Since good and opportunistic players receive the same expected payoffs from rematching with a good partner, they will have exactly the same incentives to leave a relationship or to try to maintain it. A player will therefore choose to stay in a relationship with a successful partner rather than go on the rematching market if $A\theta^S \geq p\theta^M$, and will be willing to stay with a failed partner if $A\theta^F \geq p\theta^M$.

It will be useful to define $\rho = (\gamma + (1 - \gamma)e)$. Thus $p\rho$ is the probability of a randomly selected partner succeeding, given the first-round effort choices of opportunists $e$. By Bayes’ rule, we know that $\theta^g = \frac{p\rho}{p\rho + \gamma(1-p)}$, and $\theta^F = \frac{\gamma(1-p)}{p\rho + \gamma(1-p)}$. Unsurprisingly, $\theta^g > \gamma > \theta^F$ where $e < 1$. That is, successes are more trusted
than unknown quantities, who are in turn trusted more than failures.

We now consider the probability of getting a good partner from going on the rematching market, $\theta^M$. This can be decomposed into a convex combination of the reputation of the new entrants, $\gamma$, and the endogenously determined quality of the players who break up after the first round of play and go on the market, $\theta^B$. These reputations are weighted by the mass of new entrants and the mass of breakups, $\alpha$ and $\beta$, yielding

$$\theta^M = \frac{\alpha \gamma + \beta \theta^B}{\alpha + \beta}.$$ 

**Proposition 1** In equilibrium, all players will stay with successful partners if possible, $(S) = 0$, and they will break up with failures with probability $\pi(F) \in [0,1]$. Where $\pi(F) > 0$, $\theta^B = \gamma \frac{(1 - p^2 \rho)}{1 - (p \rho)^2}$, and $\beta = \left(1 - (p \rho)^2\right) \pi(F)$.

The proofs of all results are in the appendices. A success means that the updated probability of being a good type increases, relative to the general run of the market, $\gamma$, making the decision to stay with that partner obvious. A failure indicates the opposite. Breaking up with a failure is the only disincentive to cheating available, so we will refer to this as "punishment." Since $(S) = 0$, we will henceforth refer to the probability of punishing a failure $\pi(F)$ as $\pi$.

Defining "willingness-to-punish" as $W(\gamma, p, e, \pi, A, \alpha) = \frac{\theta^M}{\alpha} - A$, the following observations are immediate.

**Lemma 1 (Willingness-to-punish conditions)** Where $W > 0$, players always punish, $\pi^* = 1$. Where $W = 0$, a mixed strategy in punishment, $\pi^* \in [0,1]$, is possible, and where $W < 0$, players never punish failed partners, $\pi^* = 0$.

Assumption 2 comes into play when $W = 0$. The small common shock to $A$ essentially coordinates the reactions of partners to failure. Setting $\pi$ amounts to choosing a cutoff point in the distribution of $A$, such that above the cutoff point, a player will always end the relationship following a failure, and below the cutoff, the player will always continue in the relationship, in the spirit of Harsanyi’s Purification (1973). Thus, where $\pi < 1$, a relationship will break down with the same probability regardless of whether one or both players fail. This assumption simplifies some of the calculations and proofs provided below, without significantly altering the qualitative properties of the results.

We now explore some of the properties of $W$.

**Lemma 2** $W$ is decreasing in $e$, for a fixed $\pi$.

Intuitively, as the effort of opportunists increases, more failures and resulting breakups will be the result of bad luck as opposed to shirking ($e < 1$). This will reduce the appeal of breaking up and trying to find a new match relative to staying with the current partner despite a failure.

**Lemma 3** $W$ is decreasing in $\pi$, for a fixed $e$. 

This comes about because as more breakups go onto the rematching market, there will be less chance of matching with a new player and more chance of being stuck with another breakup.

**Proposition 2** There exists a cutoff \( \bar{e} \in [0,1) \) such that for \( e < \bar{e} \), \( \pi = 1 \). Where \( \bar{e} > 0 \), it is the solution to \( W = 0 \) where \( \pi = 1 \). For \( e \in [\bar{e}, e^{\text{max}}] \), there exists a unique \( \pi(e) \) satisfying the condition \( W = 0 \), and \( \pi'(e) < 0 \). For \( e \geq e^{\text{max}}, \pi = 0 \). \( e \) decreases in \( A \) and increases in \( \alpha \).

Moving back a step, we can now examine what the effort levels of opportunists, \( e \), will be in the first round of play.

**Lemma 4** First-round effort in any interior equilibrium will be characterized by the first-order condition \( \pi p^3 \rho \left[ A \theta^S - \theta^M \right] = c \).

The right-hand side of the expression is of course the marginal cost of effort. The left hand side of this expression captures the marginal benefit to an opportunist of his effort on behalf of his partner. \( p \left( A \theta^S - \theta^M \right) \) is the potential benefit from sticking with a partner who succeeded rather than rematching. Extra effort will only increase the probability of receiving this benefit if two events coincide. First, the partner must in fact be a success, which occurs with probability \( p \). Second, the player’s partner must choose to punish the player for failure, which occurs with probability \( \pi \). For convenience, we will define \( MB(e, \pi) = \pi p^3 \rho \left[ A \theta^S - \theta^M \right] \). Note that \( MB \) is actually constant from the perspective of any individual player, because \( \rho, \theta^S \), and \( \theta^M \) are functions of the effort levels of all of the other opportunistic players on the market. Note also that \( \theta^M \) is also a function of \( \pi \).

**Lemma 5** \( MB_1(e, \pi) < 0 \).

The marginal benefit of effort, for given punishment probabilities, is decreasing in the effort of other opportunistic players. This arises because as opportunists exert more effort, the probability of one’s partner succeeding increases, but since more opportunists are succeeding, the value of that signal is lessened.

**Lemma 6** \( MB_2(e, \pi) > 0 \).

This arises for two reasons. The direct effect is that the more likely one’s partner is to punish following a bad outcome, the stronger is the incentive. There is an also an indirect effect through \( \theta^M \). Where more relationships are breaking up after failure, the quality of potential rematches on the market decreases.

**Proposition 3** \( MB(e, \pi(e)) \) is decreasing in \( e \).
For $e < \bar{e}$, $\pi = 1$ so this follows from the earlier lemma. Where $e \in [\bar{e}, e^\max]$, $W = 0$ and the implicit function theorem tells us that $e$ and $\pi(e)$ will decrease together, reducing $MB$. For $e > e^\max$, $MB = 0$. We will henceforth refer simply to $MB$ or $MB(e)$.

**Proposition 4** There exists a unique equilibrium $e^* \in [0, 1)$. Where $c > MB(0)$, $e^* = 0$, $\pi^* = 0$. Where $c \in (MB(\bar{e}), MB(0)]$, $e^*$ solves $MB = c$, and $\pi^* = 1$. Where $c \leq MB(\bar{e})$, $c \in [0, 1]$, $e^*$ solves $MB(e) = c$.

The first type of equilibrium is a corner solution, where $c > MB(0)$ costs are so high that opportunists never exert effort. This is the case for $c = c'$ in Figure 1. The second type of equilibrium, where costs are as in $c''$, is where equilibrium $e^* < \bar{e}$ so the punishment constraint $W \geq 0$ does not bind. In that case all successful players will break up with failures, providing the maximal possible incentive for cooperation. The third type of equilibrium is where players are indifferent about staying together or ending the relationship following a failure, $W = 0$, limiting the amount of effort that can be induced. In Figure 1, this would be where costs are $c'''$.

### 3.2 Comparative Statics

Having established the properties of the equilibrium for any fixed set of exogenous parameters, we will now consider the comparative statics of changes in these parameters on $MB$ and thus on $e^*$, for interior solutions. These will depend on the type of equilibrium we start from. If $e^* < \bar{e}$, then punishment is certain, $\pi = 1$, so we need only assess the changes to $MB(e, 1)$. Anything making the relationship more valuable relative to rematching will increase cooperation. On the other hand, if $e^* > \bar{e}$, punishment is not certain. In this range, changes in the parameters will affect both the payoffs, conditional on punishment, and the willingness to punish. This leads to somewhat counterintuitive results. Harsher punishments, in the sense of worse outside options, will reduce incentives for cooperation within the relationship, because to punish one’s partner is to punish oneself.

#### 3.2.1 Equilibria with certain punishment

We will first consider equilibria where $e^* \in (0, \bar{e})$, so $\pi = 1$, a breakup is certain if either partner fails. Taking this into account, and recalling that $\theta^S = \frac{\alpha^2}{pp}$, we can rewrite $MB = p^2(A\gamma p - pp\theta^M)$. Most of the action will be in the term $pp\theta^M$, the probability of a randomly chosen partner succeeding $(pp)$ times the expected quality of a randomly chosen new partner met on the rematching market $(\theta^M)$. As shown earlier, $\theta^M = \frac{\alpha\gamma + \beta\theta}{\alpha + \beta}$, a weighted combination of the reputations of new players, $\gamma$, and breakups, $\theta^B = \gamma \frac{(1-p^2)}{(1-(pp))^2}$. The number of new entrants, $\alpha$, is exogenous, and $\beta$ is the number of breakups on the market, $(1 - (pp)^2)$. The comparative statics of these parameters are summarized in the following proposition.
Proposition 5 Where \( e^* \in (0, \bar{e}), \frac{de}{dp} > 0, \frac{de}{dA} > 0, \frac{de}{d} < 0, \frac{de}{dR} \geq 0 \).

\( MB \) and therefore \( e^* \) are increasing in \( p \), because as the connection between effort and success becomes tighter, effort becomes more worthwhile. Similarly, the value of effort is increasing in \( A \)– as the value of staying with a good partner becomes higher, the incentives to work hard to keep the partner satisfied are higher. On the other hand, as \( \alpha \) increases, there is more chance of rematching with a new player after a breakup, making that punishment less harsh. This reduces incentives. \( e^* \) can be either increasing or decreasing in \( \gamma \), depending on other parameters. More good players in the market means that it will be more valuable in an absolute sense to stay with the current partner, but it also means that the quality of players on the rematching market will be better.

3.2.2 Equilibria with reluctant punishment

Many of these effects are reversed where \( e^* \in (\bar{e}, e^{\text{max}}) \). Here, the effect of a variable on \( W = \frac{de}{dp} - A \), the willingness-to-punish, will be crucial.

Proposition 6 Where \( e^* \in (\bar{e}, e^{\text{max}}), \frac{de}{dp} > 0, \frac{de}{dA} < 0, \frac{de}{d} > 0 \).

As \( p \) increases, players’ judgments are increasingly accurate. Fewer failures will be false positives, and the quality of breakups will be lower. Holding all else equal, this would increase \( W \) above 0. This would imply a jump in \( \pi \) up to 1, implying a sharp rise in \( e \) unless we are precisely at the point where \( e^* = \bar{e} \). Thus, in equilibrium, both \( e \) and \( \pi \) will increase slightly until \( W = 0 \) binds again. Thus, effort is strictly increasing in \( p \) regardless of whether \( W \) is binding.

The same is not true for \( A \) and \( \alpha \). An increase in \( A \) will reduce \( W \)– as the potential gains from maintaining the same relationship increase, players will be more reluctant to punish a failure, giving each other the benefit of the doubt. Effort will therefore decline until \( W = 0 \) binds. An increase in \( \alpha \) will have the opposite effect. As the number of new entrants increases, players will be more sensitive to deviations, increasing \( W \). This implies an increase in effort level. In both cases, there will be a slight offsetting effect from the change in \( MB(e, \pi) \), but this will not be enough to change the sign.

3.2.3 Effects of learning-by-doing

When comparing two kinds of transactions or industries, a low-\( A \) transaction where partners are relatively interchangeable, and a high-\( A \) transaction where there are significant benefits to keeping the same partner, the comparative statics will depend on whether willingness-to-punish binds (as in figure 2, where \( c = c' \)) or not (where \( c = c'' \)). Where \( W = 0 \) binds, transactions with higher \( A \) (\( MB'' \) in the figure) will have lower equilibrium effort levels than those with lower \( A \) (\( MB' \) in the figure). Where it does not, the effect is reversed.

This can be seen another way in figure 3. As \( A \) increases, this initially increases the incentives for performance, as opportunistic partners hope to impress
their partners and thus be able to stay around to reap the benefits. However, when $A$ increases past the point that $W = 0$ becomes binding, we see a reduction in effort. The threat of punishment becomes less credible, and effort begins to decline, eventually reaching 0 at the point where players will disregard any failure because the potential benefits of staying with the same partner are too large to abandon.

### 3.2.4 Effect of new players

The parameter $\alpha$ represents the entry of new players into the market. The larger $\alpha$ is, the more chances a player who leaves a relationship has of matching with a new partner who is untainted by that stigma. An increase in $\alpha$ thus has a qualitatively opposite effect from an increase in $A$. This can be seen most clearly in figure 4, where we compare equilibrium effort levels plotted against $A$ for two different levels of $\alpha$. Where $A$ is low, players easily rematch away from bad partners. A larger supply of untainted prospective new partners simply makes the punishment of rematching less harsh. Where $A$ is high, and willingness to punish is an issue, the availability of fresh partners helps strengthen players’ resolve to punish failure, increasing incentives to effort.

This result contrasts qualitatively with that of Dixit (2002), which finds that market expansion has a strictly negative effect on the enforceability of cooperative behavior, by making it easier for cheaters to escape the consequences of their actions. While a similar dynamic occurs here, the increasing value of the relationship with time and the inability of the market to distinguish cheater from cheated means the cheated must also think twice about moving on. By providing an escape for both parties, an expanding market decreases for spot market transactions, but increases punishment (and therefore effort) for relationships with a greater learning-by-doing component.

### 4 Clean Slates and Mutual Recriminations

In the benchmark case, we maintained assumption 2, anonymous rematching. Players who ended on relationship would be randomly matched with another player. We now consider what would happen under assumption 3, mutual recriminations. Under this assumption, any breakdown in a relationship becomes known to the community of potential rematching partners. However, each player still has private information about his own effort level and the success or failure resulting from his partner’s actions. Thus, when potential partners meet on the rematching market, they can distinguish between old players, ones who have broken up with their earlier partners, and new players.

As discussed in the introduction, many community settings will have this property. When a business approaches a new supplier for a component or product that it had already been using, that supplier will be aware that there was probably another supplier before. When a merchant looks for a new agent to represent his business in a region, that agent may know that this merchant’s
products were previously available through another source. When two regular co-authors no longer acknowledge each other in the hallways, their colleagues will notice.\footnote{In a slightly more realistic model, a bad reputation would probably evolve over several periods. That is, the stigma for the first breakup would be small, but if the inability to sustain partnerships emerged as a pattern over time, potential future partners would take note. Two periods of cooperation is simply the minimum with which to capture this effect of past cooperation on future performance.}

Institutions designed to promote information transmission and transparency may actually exacerbate this problem. Where a slow or inaccurate court system becomes involved in disputes, an accurate judgment may never emerge, but the existence of the dispute will become a matter of public record. Formal business associations or informal networks may also facilitate the spread of unsubstantiated rumors. If one partner begins spreading rumors following a breakup, the other would be unwise not to present another side of the story. This paper considers this extreme case, where the cycle of mutual recriminations leaves outsiders unable to assign blame for the breakup.

We shall see that there is not a strict ranking of this scenario relative to that of anonymous rematching. That is, more information may not necessarily be a good thing. The logic will be almost identical to that used in the earlier comparative statics analysis of $\alpha$, the number of new players.

The key change here will be in the rematching phase. With anonymous rematching players were forced to accept an essentially random assignment. Now, from the perspective of a rematching player, there are two possible classes of partners they might be matched with—new players (with reputation $\gamma$) and old players (with reputation $\theta^B$). The difference between these two reputations, $\gamma - \theta^B$, can be thought of as the stigma of being a breakup. Since $\gamma > \theta^B$, the expected utility of matching with a new player will be greater than that of matching with another breakup.

I assume that positive assortative matching will emerge in equilibrium. This satisfies the commonly applied criterion of pairwise stable matching, and can be derived as the equilibrium of a non-cooperative subgame as shown in Appendix 1.

This means that a change from the "hit-and-run" environment of the anonymous market to one where relationship histories are observable to the community will be equivalent to setting $\alpha = 0$. Players from breakups will be forced to match only with each other. As we noted when we interpreted a change in $\alpha$ as a market expansion, the effect of this will depend on the importance of the willingness to punish, $W$. If $W = 0$, going from $\alpha > 0$ to $\alpha = 0$ will make players even more reluctant to punish, reducing effort levels. If $W > 0$, then going from $\alpha > 0$ to $\alpha = 0$ will partly act to toughen incentives, but if the shift is large enough to make $W = 0$, this effect will be counteracted by an increasing reluctance to punish.

Two competing effects arise when analyzing the welfare effects of these different informational regimes. In particular, when the willingness-to-punish condition is binding, reducing the amount of information available will increase
the attainable effort level, but at the cost of increasing the frequency of breakups and discarding the benefits of learning-by-doing.

The total surplus in equilibrium will be:

\[ V(A, p, \gamma, e, \pi) = p\gamma(A + 1) - \gamma p(A - 1)\pi^* (1 - p^2\rho^*) + (1 - \gamma)(p - c)e^* \]

The first term, \( p\gamma(A + 1) \), is the maximal possible productivity of good types if they stayed with their partners for two rounds. This remains unchanged in any equilibrium. The second term is the cost of breakups. \( p(A - 1) \) is the lost productivity for each good type that leaves his partner. This loss occurs if either partner fails and they choose to punish each other, which has probability \( \pi(1 - p^2\rho^*) \). The third term is the effort increase. This is the amount produced by inducing the \( (1 - \gamma) \) opportunists to exert effort.

Welfare would be maximized if players could somehow precommit to \( \pi = 0 \) and to maximal effort \( e = 1 \). In equilibrium, \( \pi > 0 \) is necessary to induce positive effort. This leads to a tradeoff between the breakup costs incurred as partners become more sensitive to failures, and the effort generated by opportunists as they work to maintain their good standing.

**Proposition 7** Equilibrium social welfare under assumption A2, Clean Slates, will be higher than social welfare under assumption A3, Mutual Recriminations, only if \( \pi^* < 1 \) under A3. [Tentative:] If \( \pi^* < 1 \) under A3, there exists \( \bar{\alpha} > 0 \) s.t. for \( \alpha \in [0, \bar{\alpha}) \) equilibrium social welfare under Clean Slates will be higher.

What will be the welfare effects of going from the world of clean slates \( (\alpha > 0) \) to the world with mutual recriminations \( (\alpha = 0) \)? If punishment is already certain \( (\pi^* = 1) \) despite mutual recriminations, then moving to the clean slates regime will be strictly harmful. A reduction in the stigma of breakup will make the consequences of failure less severe, so opportunists will reduce their effort levels.

On the other hand, if \( \pi^* < 1 \), the constraint \( W = 0 \), willingness to punish, is binding. In that case, the fact that the Clean Slates regime provides opportunities to match with new players will loosen that constraint, making players more willing to break up following a failure. This in turn will increase effort, as shown in [the earlier proposition regarding \( \frac{d\pi}{d\alpha} \)]. The increase in effort will have a direct impact on social welfare through \( (1 - \gamma)(p - c)e^* \). It will also provide an indirect benefit because the number of good players who fail or whose partners fail, \( \gamma(1 - p^2\rho) \), will decrease. Counteracting these beneficial effects, however, will be an increase in the rate of breakups \( \pi \) for those pairs that do not have good outcomes. I am still working on an analytical characterization of when this negative effect might outweigh the positive effect. For examples calculated under a variety of different parameters, there does not appear to be any such case.

To see why there might be some \( \bar{\alpha} \) above which Clean Slates would reduce welfare, consider a case where \( \pi^* \) is close to 1 under Mutual Recriminations. For a small \( \alpha \) the move to Clean Slates would increase \( \pi \) and therefore \( e \). Once \( \pi \) reaches 1, any further increase in \( \alpha \) will not increase incentives by increasing
Instead, it will increase \( \theta^M \), decreasing \( MB(e, \pi) = \pi \rho^3 \rho [A \theta^S - \theta^M] \) and thus reducing effort.

5 Conclusion

This model illustrates a surprising fact–better information about players’ past histories may diminish rather than enhance incentives. While it is commonly suggested that increasing "transparency" will facilitate market development, this model suggests matters may not be so easy. Simply establishing common knowledge of the easily verifiable facts, such as the existence of a dispute, can actually damage incentives for cooperative effort. This helps explain why Mexican footwear manufacturers are careful to police their own (Woodruff, 1997) and why cattle ranchers in conflict with their neighbors must take costly actions to signal to the community that they are not in the wrong (Ellickson, 1991). Formal or informal institutions may be as necessary to protect the reputations of accusers as they are to indict the guilty.

6 Appendix A: Rematching process with two types

We consider here a matching process with two types, A and B, both of which prefer to match with type A. Suppose that rematching proceeds through a form of the deferred acceptance algorithm as follows: Each player simultaneously ranks his preferences ("A over B" or "B over A"). All possible mutual matches where each player is of the other’s first-choice type are made. Remaining players are then selected in a random order and given their first choice, if available. Once no player’s first choice is available, the remaining players are randomly matched.

Claim 1 In equilibrium, matching will be positively assortative by reputation and no players will be unmatched.

Proof. Matching players have only two strategies–to prefer A or to prefer B. Let \( \alpha^A \) be the measure of new players who declare a preference for A, \( \alpha^B \) the measure of A players who declare "N," \( \beta^A \) the measure of B players declaring "A," and \( \beta^B \) the measure of B players declaring "B." All members of the group \( \alpha^A \) will match with each other in the first round, as will all the members of the group \( \beta^B \). Since there is a continuum of players, no member of these groups will be left unmatched after the first round. Suppose \( \alpha^B > \beta^B \), then all remaining B players will match with young players, and young players will match with B with probability \( \frac{\beta^B}{\alpha^B} \). The remaining \( \alpha^B - \beta^B \) young players will be forced to match with each other. Symmetrically, if \( \alpha^B < \beta^B \) then \( \beta^B - \alpha^B \) B players will be forced to match with one another. If \( \alpha^B = \beta^B \) then all players will receive their first choice. \( \blacksquare \)
It is therefore weakly dominant for every player to declare the same preference ordering: A over B.

Consider an A player. If he declares "A" he will match with another A if \( \alpha^A > 0 \). If \( \alpha^A = 0 \) and \( \beta^A > 0 \), he will match with a B if \( \alpha^B < \beta^A \), otherwise he will be matched with another A from the group \( \alpha^B \). On the other hand, if he declares "B," he will match with a B with probability \( \max \left\{ \frac{\beta^A}{\alpha^A}, 1 \right\} \) and otherwise will match with an A. If in equilibrium \( \alpha^A > 0 \), declaring "A" is strictly preferable. If \( \alpha^A = 0 \), and \( \beta^A > 0 \), he will be matched with an B regardless of his declaration. If \( \alpha^A = 0 \), and \( \beta^A < \alpha^A \), then he will be matched with an A if he requests one, or with a B with probability \( \frac{\beta^A}{\alpha^A} \) if he requests one.

The case of \( \alpha^A = 0 \), and \( \beta^A = \beta^A \) is ill-defined in this setting, since a measure 0 of players will be left over for our individual player to match with. We cannot simply disregard this as a generically non-occurring event, since \( \alpha^B \) and \( \beta^A \) emerge from equilibrium choices. Therefore, I assume in this case that our player will be randomly matched with another player, getting a good match with probability \( \frac{\alpha^B + \beta^A}{\alpha^A + \beta^A + \beta^B} \). His choice will then be irrelevant. Declaring "A" is strictly preferred unless \( \alpha^A = 0 \), and \( \beta^B < \beta^A \), in which case the player will be indifferent.

Now consider a B player. His options will be symmetric to the A player's, but of course he would prefer to match with a A player. If \( \alpha^B > \beta^A \) he will get an A partner if he requests one, and will get one regardless if \( \beta^B = 0 \). If \( \alpha^B < \beta^A \) then he can match with an A player with probability \( \frac{\alpha^B}{\beta^A} \) if he desires, otherwise he will match with a B. Thus, requesting an A partner is strictly preferred if \( \beta^B > 0 \). Since having players request B partners means \( \beta^B > 0 \), such a strategy cannot be supported in equilibrium.

There are two possible equilibria of this process. The first, and more plausible, is the equilibrium selected by elimination of weakly dominated strategies, where all players declare a preference for A partners, but A players only match with each other (\( \alpha^B = \beta^B = 0 \)). If \( \alpha \leq \beta \), it is also possible to construct the equilibrium where \( \alpha^A = \beta^B = 0 \). In such an equilibrium, A players declare a preference for "B" because they are indifferent, but this is not robust to the possibility of some fraction \( \varepsilon \) of "A" types deviating and expressing their true preference to match with another A player. I thus assume that the first equilibrium will hold, leading to perfect positive assortative matching.

7 Appendix B: Proofs of Propositions

[currently rewriting–available upon request]
8 Works Cited

References


Figure 1: Equilibrium effort levels
Figure 2: Increasing A
Figure 3: Equilibrium effort and $A$
Figure 4: Equilibrium effort with increasing $\alpha$