GROUP REPUTATIONS, STEREOTYPES, AND COOPERATION IN A REPEATED LABOR MARKET

Paul J. Healy*

February 7, 2007

Abstract

Reputation effects and other-regarding preferences have each been used to predict cooperative outcomes in markets with inefficient equilibria. Existing reputation-building models require either infinite time horizons or publicly observed identities, but cooperative outcomes have been observed in several moral hazard experiments with finite horizons and anonymous interactions. This paper introduces a *full reputation equilibrium* (FRE) with stereotyping (perceived type correlation) in which cooperation is predicted in early periods of a finitely repeated market with anonymous interactions. New experiments generate results in line with the FRE prediction, including final-period reversions to stage-game equilibrium and non-cooperative play under unfavorable payoff parameters. (**JEL** C72, C91, D52, D64)

How can cooperation persist in the absence of enforceable performance contracts? With infinitely-lived relationships, cooperation can emerge when the long-term cost of damaging a valuable relationship outweighs the immediate benefit of poor performance (see, for example, the models of Benjamin Klein & Keith B. Leffler (1981) and W. Bentley MacLeod & James M. Malcomson (1989), or the 'folk' theorems of Drew Fudenberg & Eric Maskin (1986) and others). Even with finitely-lived relationships, David M. Kreps *et al.* (n.d.) demonstrate that the standard unraveling arguments can be avoided and cooperation maintained for some length of time if there is a small degree of uncertainty about players' preferences. Specifically, selfish (rational) players prefer to build a false reputation for being a 'tit-for-tat' player in early periods, though they must reveal their true stripes by the final period.

In these reputation-based 'folk theorem' arguments with a finite horizon, it is essential that players know the identity of their opponents.[1] Experimental studies show, however, that cooperation can emerge in finitely repeated games even when interactions are anonymous. In several tests of moral hazard in repeated labor markets (see Ernst Fehr, Georg Kirchsteiger & Arno Riedl (1993, 1998b), Fehr & Armin Falk (1999), Simon Gächter & Falk (2002), R. Lynn Hannan, John H. Kagel, & David V. Moser (2002), and Gary Charness (2004), among others), wages and effort levels are observed substantially higher than the stage game equilibrium prediction, even though transactions are anonymous. Consequently, many authors have concluded that players must have preferences for fairness, inequity, or reciprocity that lead to cooperative outcomes, even in one-shot games.

In this paper, we demonstrate that folk theorems for finitely-repeated games can be extended to the case of anonymous matching to predict the cooperation observed in the repeated labor market experiments. The basic argument works as follows: assume, *à la* Kreps *et al.* (n.d.), that some percentage of workers are in fact fair-minded players whose effort is always positively correlated with their wage. If it is common knowledge that firms believe workers' types are correlated (*i.e.*, firms stereotype the workers) then a single

[1] In the 'contagion' equilibrium of Michihiro Kandori (1992) interactions are anonymous but the time horizon is infinite.

defection by one worker leads firms to believe that other workers are more likely to be selfish as well. This one defection can sufficiently damage the reputation of every group member so that firms offer only low wages in all subsequent periods. Depending on the payoff structure, firms' prior beliefs, and the degree of perceived correlation among types, selfish workers may prefer to imitate a reciprocal worker in early periods of the repeated game, *even when his actions are not linked to his identity*, because damaging the group's reputation means damaging his own future outcomes. Consequently, if a selfish worker would prefer to imitate the reciprocal type in a two-player repeated game, he would also prefer to do so in a repeated game with many players and anonymous matching.

Note the following about this argument: First, we assume firms believe a non-trivial fraction of workers have other-regarding preferences, which is best supported by assuming that the percentage of other-regarding workers is in fact non-trivial. Thus, we interpret this as a 'mixed' model in which other-regarding preferences and repeated-game effects operate together to generate cooperative outcomes. Second, we do not assume a particular form of other-regarding preferences; any preference-based model that predicts a positive wage-effort correlation can be inserted into the above argument. Third, the assumption of correlation in firms' beliefs is quite necessary; we show in Proposition 2 that such reputation-building equilibria exist for only a vary small set of firms' prior beliefs, and that this set shrinks quickly in the number of workers. Fourth, we predict that the selfish workers revert to defection by the final period. This end-game reversion is not observed in some experimental studies, and a failure to revert to the selfish equilibrium is consistent with our theory only when all workers are in fact non-selfish. Finally, the existence of this reputation-building equilibrium is sensitive to the payoff parameters of the game and the (unobservable) beliefs of the firms.

We find support for our theory in a series of new repeated labor market experiments (see Sections III and IV). Specifically, we observe cooperation in early periods, with a pronounced 'crash' towards the stage game equilibrium in the final period, and we find that cooperation fails to emerge when the payoff parameters are made more 'stringent', where the reputation-building equilibrium exists only when firms believe that nearly all workers have other-regarding preferences. The effect of changing the payoff parameters is most pronounced in one experimental session where a group of subjects exhibit no cooperation under the stringent parameters, but cooperation subsequently emerges (and then crashes) for the same subjects under less stringent parameters.

Taken individually, our experimental results are not particularly novel; several studies have shown end-game reversion towards the selfish equilibrium (for example, Fehr, Kirchsteiger & Riedl (1998b), Jordi Brandts & Charness (2004), Charness, Guillaume Frechette, & Kagel (2004), and Arno Riedl & Jean-Robert Tyran (2005)), while others report sessions that fail to generate significant cooperation (including Michael Lynch *et al.* (2001), Dirk Engelmann & Andreas Ortmann (2002), and Mary L. Rigdon (2002)).[2] The reputation-building repeated game theory in this paper helps to explain when such end-game reversion and failures of cooperation are likely to occur.

The assumption that firms believe workers' preferences (or *types*) are correlated can be justified on two grounds. First, if firms are uncertain about the underlying percentage of other-regarding workers in the economy, then correlation naturally emerges since data about an individual worker provides some information about the entire population of workers. Second, even without this underlying uncertainty, it is well established in the social psychology literature that beliefs are frequently stereotypical in nature, leading to more correlation than is warranted by Bayes's Law.[3] Regardless of the underlying cause, the existence of correlated beliefs (and the existence of other-regarding preferences) is well documented and is therefore natural to include in a descriptive game-theoretic model.

[2] In some studies, end-game reversion is not obvious when studying group average behavior, but is apparent at the individual level. In some papers, individual data is available only in the appendix.

[3] See the online appendix for a brief review of this literature.

The formal model is developed, piece-by-piece, in Section I and extended to the larger environment of interest in Section II. We describe our experiments in Section III and examine the results in Section IV. To check the robustness of our results, we compare the model's predictions to data from several previous experiments in Section V. A brief summary and possible directions for future work appear in Section VI.

I A Simple Repeated Labor Market

Our goal is to develop a model of rational cooperation in a finitely repeated labor market (which is isomorphic to a sequential prisoners' dilemma) in the absence of individual reputation effects. We generalize the sequential equilibrium reputation-building theory of Kreps *et al.* (n.d.) to include perceived type correlation and consider only the *full reputation equilibrium (FRE)* in which selfish workers imitate the reciprocal type with certainty in every period except the last.[4] To help communicate the key ideas, the theory is described in increasing levels of complexity, starting with complete information and publicly observed actions, then adding uncertainty about types, making actions private, and finally assuming stereotypical beliefs.

Assume there are n workers and m firms, with $n \geq m$.[5] In each period $t \in \{1, 2, \ldots, T\}$, each firm is randomly matched with one worker. Matched firms offer a wage $w_t \in \{\underline{w}, \overline{w}\}$ to their worker, who then responds with effort level $e_t \in \{\underline{e}, \overline{e}\}$, where $\underline{w} < \overline{w}$ and $\underline{e} < \overline{e}$ [***NOTE TO PUBLISHER: $\underline{w}$ and $\underline{e}$ are underscored, not italicized***]. Period t payoffs to the firm and worker are denoted by $\pi(w_t, e_t)$ and $u(w_t, e_t)$, respectively, where π is decreasing in w_t and increasing in e_t and u is increasing in w_t and decreasing in e_t. We as-

[4]The Kreps *et al.* theory was also generalized in Fudenberg & Maskin (1986, Theorem 4), which allows for arbitrary behavioral types but does not incorporate correlated beliefs.

[5]This is only for ease of exposition; the derived equilibrium with $n < m$ is identical to that with $n = m$. This is true since firms aren't facing any temptations to defect as the game nears its end, and therefore will not change their behavior when it becomes less likely that they will participate in future periods.

sume that $(\overline{w},\overline{e})$ Pareto dominates $(\underline{w},\underline{e})$. Finally, assume that unmatched workers receive no payoff for the period. The stage game for a matched firm-worker pair (with normalized payoffs) is shown in Panel A of Figure I. The assumptions on π and u give this game the standard sequential prisoners' dilemma structure.

The only Nash equilibrium outcome of the game is $(\underline{w},\underline{e})$.[6] Since $(\overline{w},\overline{e})$ Pareto dominates the equilibrium outcome, we refer to it as a *cooperative* outcome. If the firm believes the worker is not rational, but instead committed to playing the 'reciprocal' strategy (playing $\overline{e}$ when $\overline{w}$ is chosen and $\underline{e}$ when $\underline{w}$ is chosen), the firm's optimal strategy would be $\overline{w}$. If the firm is unsure about the worker's preferences, the optimal wage offer of the firm depends on his belief about the likelihood that the worker is 'selfish' (she has the payoffs and strategies shown in Panel A of Figure I) versus 'reciprocal' (she always plays the reciprocal strategy).[7]

Assume for now that the stage game is played only once and each firm believes its worker is reciprocal with probability p and selfish with probability $1-p$. This game of incomplete information is shown in Panel B of Figure I. If the firm offers $\underline{w}$ it will receive $\underline{e}$ from either type of worker. If it offers $\overline{w}$, it faces a lottery; with probability p it will receive $\overline{e}$ and with probability $1-p$ it will receive $\underline{e}$. This lottery is preferred to offering $\underline{w}$ if and only if $p \geq p^*$, where

$$p^* = \frac{\pi\left(\underline{w},\underline{e}\right) - \pi\left(\overline{w},\underline{e}\right)}{\pi\left(\overline{w},\overline{e}\right) - \pi\left(\overline{w},\underline{e}\right)}. \tag{1}$$

If the same firm and worker were matched in every period, there can exist a full rep-

[6]In equilibrium, the firm must offer $\underline{w}$ with probability one. The worker must respond to $\underline{w}$ with $\underline{e}$, but can respond to $\overline{w}$ with any $\Pr\left[\overline{e}|\overline{w}\right] \leq b/\left(1+b\right)$ since $\overline{w}$ is never observed. Thus, there is a continuum of equilibria, but $\Pr\left[\overline{e}|\overline{w}\right] = 0$ is the only one that is subgame perfect.

[7]We could instead assume that the reciprocal type receives payoffs of one if her observed action is consistent with reciprocation and zero otherwise. Doing so introduces other Nash equilibria into the game that are not subgame perfect. It also complicates the specification of beliefs in the sequential equilibrium of the repeated game. The current assumption is equivalent to restricting attention to sequential equilibria in which the reciprocal type plays the reciprocal strategy with probability one.

utation equilibrium in which the firm offers $\overline{w}$ in every period (as long as the worker has always played $\overline{e}$ in the past) and the selfish worker chooses $\overline{e}$ in response to $\overline{w}$ in every period except the last, at which point she plays $\underline{e}$ regardless of w_T. The firm's belief in any period is p_1 (his initial belief) if the worker has always played $\overline{e}$ in response to $\overline{w}$, and 0 otherwise. This equilibrium exists if (and only if) the firm's prior belief is at least p^*. The argument is relatively simple: In such an equilibrium, the firm's beliefs do not change from period to period since the selfish worker behaves exactly the same as the reciprocal worker until the final period. Letting p_t be the firm's belief that the worker is reciprocal at the beginning of period t, we have $p_t = p_1 \geq p^*$ along the equilibrium path. In the final period, $p_T \geq p^*$ implies that the firm offers $\overline{w}$. The selfish worker clearly chooses $\underline{e}$. In the penultimate period, the selfish worker who is offered $\overline{w}$ and knows $p_{T-1} \geq p^*$ can choose to deviate by playing $\underline{e}$, but this would cause $p_T = 0$ and $w_T = \underline{w}$.[8] With a discount factor of δ, conforming to the equilibrium is preferred to this deviation if and only if $\delta \geq \delta^*$, where

(2) $$\delta^* = \frac{u\left(\overline{w}, \underline{e}\right) - u\left(\overline{w}, \overline{e}\right)}{u\left(\overline{w}, \underline{e}\right) - u\left(\underline{w}, \underline{e}\right)}.$$

Note that $\delta^* \leq 1$. In the sequel, we assume $\delta = 1$ so that $\delta \geq \delta^*$ always holds.

The firm in period $T-1$ with belief $p_{T-1} \geq p^*$ knows that he will receive $\overline{e}$ if he offers $\overline{w}$ and $\underline{e}$ if he offers $\underline{w}$, and neither option will affect his beliefs or optimal strategies in the final period. Thus, the firm maximizes his current-period payoff by choosing $w_{T-1} = \overline{w}$. The argument is identical for all previous periods, so, by induction, an FRE exists if and only if $p_1 \geq p^*$.

Full reputation equilibria are clearly not the only sequential equilibria of this game in which the cooperative outcome can be realized for some number of periods. For example,

[8] The fact that the reciprocal type cannot play $\underline{e}$ in response to $\overline{w}$ means that the firm's belief must update to $p_T = 0$ upon observing $\underline{e}$.

if $T = 2$, there is a $p^{**} < p^*$ such that if $p_1 \in [p^{**}, p^*)$, the firm offers $\overline{w}$ in the first period and the selfish worker plays $\overline{e}$ with probability just low enough so that $p_2 = p^*$ if $\overline{e}$ occurs.[9] With positive probability, however, the worker chooses $\underline{e}$, causing the firm to choose $\underline{w}$ in the final period. This argument can be extended for any finite T, with the lower bound on p_1 decreasing in T. While such equilibria can be observationally equivalent to an FRE if $\overline{e}$ happens to occur in every period except the last, we focus only on the equilibrium in which $\overline{e}$ is chosen as a pure strategy in all but the last period. This equilibrium exists only when $p_1 \geq p^*$.

To generalize the above argument to the case where multiple firms are matched with multiple workers, it becomes necessary first to specify whether the random matching of workers to firms is publicly observed or not. Ultimately, we will demonstrate that, with sufficient stereotyping, anonymous matching will have no effect on FRE behavior.

A Publicly Observed Matching

If the actions and identities of each pairing are publicly observable and firms have common beliefs, then the firms share a belief p_{it} about each worker i in each period t, and each selfish worker knows that deviating from the FRE will guarantee that she receives $\underline{w}$ in all future periods. Again, an FRE exists (for worker i) only if $p_{i1} \geq p^*$. There is one added wrinkle: Workers face a probability $1 - m/n$ that they will not be employed in the next period. The quantity m/n now acts as a one-time discount on workers' future payoffs. A risk-neutral selfish worker will choose $\overline{e}$ given $\overline{w}$ if and only if this discount factor (m/n) is greater than δ^* from equation 2. Note that if the worker is willing to choose $\overline{e}$ given $\overline{w}$ in period $T - 1$, then she has an even stronger incentive to choose $\overline{e}$ in any previous period.

[9] With the normalized payoffs of Figure I, $p^* = b/(1+b)$, $p^{**} = b/(2+b+1/b)$ and $\Pr[e_1 = \overline{e} | w_1 = \overline{w}] = (1/b)(p_1/(1-p_1))$, which is strictly less than 1 when $p_1 < p^*$.

This proves the following proposition.[10]

Proposition 1 *Assume there are n workers and m firms. In the T-period repeated labor market with publicly observed random matching and public wage and effort choices, there is a full reputation equilibrium ($\overline{w}$ in every period and $\overline{e}$ in every period but the last) if and only if (1) firms' common prior belief about each worker's type is at least p^*, and (2) $m/n \geq \delta^*$.*

B Completely Anonymous Matching

We now assume that firms do not know the identity of the workers. Instead, firms hire workers from a particular population and cannot observe the past behavior of any one worker. This assumption, which matches the experimental environment of interest, minimizes the incentive for individuals to build reputations. We continue to assume that actions are public information; if a worker defects, the defection becomes common knowledge, but the identity of the defector is veiled.

Let the firms' common belief in period t that their randomly assigned worker is reciprocal be p_t. We refer to this as the *group reputation* of the workers because, by anonymity, p_t completely describes the firms' beliefs about the pool of workers. On the FRE path, $p_t = p_1$ for all t since both types of workers behave identically. If one worker deviates in some period $t < T$, then all firms know there is one worker that is selfish with certainty and $n - 1$ workers about which no more information has been revealed.[11] The firms' posterior then becomes $p_t\,(n-1)\,/n$. In this environment, one deviation slightly damages the group reputation, but the size of the effect is relatively small and decreases quickly in n.

Along the equilibrium path we know that $p_T \geq p^*$ (and thus $p_1 \geq p^*$) is necessary for

[10] Formal proofs are available in the online appendix.

[11] This is true even though deviations are a zero-probability event because reciprocal types are unable to deviate.

the firms to offer $\overline{w}$ in period T. But now suppose that $p_1 \geq p^* n / (n-1)$. In period $T-1$, if a single worker defects, the group reputation becomes $p_T = p_1 (n-1)/n \geq p^*$, so firms in the final period still believe it sufficiently likely that they will encounter a reciprocal worker and will therefore offer $\overline{w}$ in the final period. Thus, at least one selfish worker will defect in period $T-1$. In order for a full reputation equilibrium to exist, p_1 must lie between p^* and $p^* n / (n-1)$. This range is quite small for even moderate values of n.

As in the case of public matching, we still have the added wrinkle that a worker may be unemployed in the final period. Again, the probability of being employed (m/n) must be sufficiently large to induce the worker to cooperate (by playing $\overline{e}$ in response to $\overline{w}$) in period $T-1$. Combining this with the restriction on p_1 gives the following proposition.

Proposition 2 *In the T-period repeated labor market with completely anonymous random matching and public wage and effort choices, there is a full reputation equilibrium ($\overline{w}$ in every period and $\overline{e}$ in every period but the last) if and only if (1) the firms' common prior belief (p_1) satisfies*

(3)
$$p_1 \in \left[p^*, \frac{n}{n-1} p^* \right),$$

and (2) $m/n \geq \delta^*$.

C Stereotypes

Proposition 2 places a tight restriction on the range of allowable priors. The anonymity of the labor interaction makes the effect of a single worker's defection on the group's reputation relatively small. This occurs because firms believe that the existence of one selfish worker implies nothing about the types of the remaining workers. Suppose instead that firms believe types are correlated. In this case, the defection of a single worker signals not only that there is one selfish worker in the group, but that the other group members are more likely to be selfish as well. If a single worker were to defect, the group reputation would be more severely damaged, making it more likely that firms will switch to offering $\underline{w}$ in subsequent periods.

Formally, we model stereotyping by assuming that the workers' types are binary random variables whose correlation matrix has off-diagonal elements all equal to $\gamma \in [0, 1]$. Let p_1 be the prior marginal probability that any given worker is reciprocal. Upon observing that one worker i is in fact selfish, the firms' conditional probability that worker $k \neq i$ is reciprocal becomes $(1 - \gamma)\, p_1$.[12] If $\gamma = 0$, types are believed to be uncorrelated. If $\gamma = 1$, firms believe workers' types are perfectly correlated.

Perceived correlation may or may not be consistent with the actual distribution of types. For example, the firms may be initially uncertain about the base rate of reciprocal types in the economy, and observing a selfish type results in a downward shift in the estimated probability that another worker is reciprocal. This rational updating story seems appropriate for a newly established firm hiring from an unfamiliar population of workers, or for an experimental subject matched with a small group of other subjects drawn from a large population. It is perhaps inappropriate for firms with long histories of working with a stable population of potential employees. Regardless of the prior information about the group's characteristics, we can always motivate the perceived correlation as an irrational stereotyping phenomenon. Managers within the firm may use data from individual workers to make (possibly incorrect) inferences about the entire group. In the most extreme case ($\gamma = 1$), a single selfish worker causes the managers to conclude that all workers in this population are in fact selfish. Consequently, we refer to γ as the *stereotyping parameter*.

Now reconsider the completely anonymous matching case from above. If a single worker defects, firms know that one worker is selfish with certainty and believe each of

[12]This conditional probability is derived in the online appendix.

the remaining $n-1$ workers to be reciprocal with probability $(1-\gamma)\,p_1$. Thus, the workers' group reputation becomes $(1-\gamma)\,p_1\,(n-1)\,/n$. When $\gamma > 0$, the effect of a single defection on the group reputation becomes more severe. The following proposition formalizes how the stereotyping assumption widens the range of parameters on which an FRE can exist.

Proposition 3 *In the T-period repeated labor market with completely anonymous random matching, public wage and effort choices, and a common knowledge stereotyping parameter γ, there is a full reputation equilibrium ($\overline{w}$ in every period and $\overline{e}$ in every period but the last) if and only if (1) the firms' common prior belief (p_1) satisfies*

$$p_1 \in \left[p^*, \frac{1}{1-\gamma}\frac{n}{n-1}p^*\right), \tag{4}$$

and (2) $m/n \geq \delta^*$.[13]

When $\gamma > 1 - p^* n/\,(n-1)$ the upper bound in equation 4 exceeds one and the two conditions of Proposition 3 become identical to those of Proposition 1. Thus, with sufficient stereotyping, an FRE exists under completely anonymous matching if and only if it exists under public matching. It is also worth noting that all workers act identically along the equilibrium path until the final period, so firms do not observe data that contradicts their belief of type correlation until the last move of the game. Even perfect correlation ($\gamma = 1$) is consistent with observed play until the end.

Proposition 3 merges two key concepts: Reputation-building sequential equilibrium and stereotypical thinking. Both concepts have been independently studied and past liter-

[13] If wages are publicly observed but efforts are not, the proposition remains valid under mild assumptions. If condition (1) holds, then a single defection in period $T-1$ makes the matched firm's belief drop below p^*. That firm will offer $\underline{w}$ in period T. If wage choices are not truly simultaneous and $\underline{w}$ is observed before other firms are matched with workers, other firms will know that a selfish worker exists and will then choose $\underline{w}$ in period T as well. If the firm who was defected upon does *not* move first, the information will disseminate only after that firm makes his wage offer.

ature suggests that both are relevant phenomena. Several experimental studies (including Reinhard J. R. Selten & Rolf Stoecker (1986), Colin F. Camerer & Keith Weigelt (1988), Richard D. McKelvey & Thomas R. Palfrey (1992), John Neral & John Ochs (1992), and James Andreoni & John H. Miller (1993)) support the conclusion that players often follow reputation-building sequential equilibria in those games where long-run players can develop meaningful reputations. To confirm the existence of correlated beliefs, William McEvily *et al.* (2007) show that if a person belongs to a group whose members have been untrustworthy, people from other groups will expect the person to be untrustworthy as well, even when it is common knowledge that group membership boundaries were chosen arbitrarily. A review of the social psychology literature reveals that stereotyping is often observed in controlled settings,that awareness of heterogeneity does not eliminate the tendency to stereotype, and that stereotypes are strengthened in competitive situations and in situations that are cognitively demanding.[14]

II The Larger Environment

The goal through the remainder of this paper is to develop a new set of experiments that test the distinct implications of the full reputation equilibrium (FRE) with stereotyping and to analyze previous experimental results through the lens of the FRE model. This means scaling up the simplified version of the labor market described in Section I to one that matches existing experimental environments. In particular, we use as our environment the experimental design from the seminal paper of Fehr, Kirchsteiger & Riedl (1993) (hereafter FKR,) which is very similar to that of many subsequent studies.

Six firms ($m = 6$) and nine workers ($n = 9$) repeatedly participate in a market in which firms offer wages and then workers choose an effort level. The set of allowable wages is

[14] See the online appendix for details.

expanded to $\{5, 10, 15, \ldots\}$ and the set of allowable efforts is $\{1, \ldots, 10\}$. Firms post their wage offers for all to see and workers choose which wage to accept, if any. Workers who accept a wage become matched with the offering firm and the pair exit the market. The timing of moves is unrestricted; when the market is open any unmatched firm can post a wage and any unmatched worker can accept any posted wage.[15] The firms' per-period payoff function $\pi(w, e)$ is decreasing in w and increasing in e, while the workers' per-period payoff function $u(w, e)$ increases in w and decreases in e. We assume $u(25, 1) < 0 < u(30, 1)$ so that workers prefer to remain unmatched over accepting a wage below the reservation wage of 30. The market remains open for three minutes, after which all unmatched agents receive zero payoff for the period. Each three-minute market constitutes a period and twelve periods are played in total. The number of periods is common knowledge.

A Predictions

Let $\xi(w)$ denote a worker's effort choice in response to a wage offer w. If it is common knowledge that all agents aim to maximize a discounted sum of their per-period payoffs, then the unique Nash equilibrium of the stage game is unconditionally low effort by workers ($\xi(w) = 1$ for all w) and reservation wages offered by firms ($w = 30$).[16] We denote this equilibrium wage-effort pair by $(\underline{w}, \underline{e})$ and use it as a benchmark prediction against which we can compare our experimental results.

Static Equilibrium Prediction Firms offer the reservation wage ($\underline{w}$) in every period and workers always choose minimum effort ($\underline{e}$) regardless of the wage offers.

[15] Firms can revise their existing wage offer by submitting a new offer, as long as it improves on the best outstanding offer in the market.

[16] We exclude all no-trade equilibria, which only exist when agents can choose to opt out of the market. As in FKR, we set $\pi(\underline{w}, \underline{e}) > 0$ and $u(\underline{w}, \underline{e}) > 0$ so that subjects strictly prefer the equilibrium with trade over the no-trade equilibria. This introduces subgame perfect equilibria of the repeated game in which cooperation is maintained in early periods by the threat of a no-trade equilibrium following a deviation. However, we do not consider such equilibria because it cannot explain the observed cooperation in previous experiments with exogenous matching protocols that don't allow subjects a no-trade strategy.

As noted, the benchmark prediction is a poor description of behavior in many previous studies where agents realize wage-effort pairs that Pareto dominate $(\underline{w}, \underline{e})$). Several authors have proposed models in which some agents' preferences are extended to include the payoffs and/or actions of their opponents. Among the most well-known is the (linear) inequality aversion model of Fehr & Klaus M. Schmidt (1999), where agents maximize their own payoff minus β_i times the difference in players' payoffs.[17] If β_i is large enough, ξ will be an increasing function of the wage. Firms know that high wages will be met with high effort and, depending on the shape of ξ, can increase their per-period payoff by offering higher wages. Since this will be an equilibrium outcome in the final period, there is a subgame perfect equilibrium of the repeated game with high wages and effort in every period, including the last. Similarly, Gary E. Bolton & Axel Ockenfels (2000) show how cooperation can be maintained if a sufficient fraction of workers aim only to minimize $|\pi(w,e) - u(w,e)|$, which is similar in spirit to the last period of the FRE where firms offer high wages if they believe a sufficient fraction of workers are reciprocal.[18]

These 'outcome-based' models can be compared to 'intentions-based' models such as that of Matthew Rabin (1993) in which a worker would prefer to match a 'generous' wage offer with a 'generous' effort choice.[19] In this specification, $\xi(w) = 1$ for all w below some cutoff and is increasing above the cutoff. Martin Dufwenberg & Georg Kirchsteiger (2004) refine Rabin's model for extensive form games and show the existence of an equilibrium in the sequential prisoners' dilemma (Panel A of Figure I) in which the worker behaves reciprocally. Kevin A. McCabe & Vernon L. Smith (2000), Charness & Rabin (2002),

[17]The full model distinguishes between inequality that favors i and inequality that favors his opponent. In the labor market game, the latter never occurs to workers in equilibrium. See Fehr & Schmidt (1999, p. 849).

[18]James C. Cox and Vjollca Sadiraj (2005) analyze a model of altruism featuring CES utility functions over own payoff and others' payoffs (rather than assuming inequality aversion) that can predict high effort in response to high wages.

[19]There is no restriction on the cutoff between 'stingy' and 'generous', but it should depend on the worker's beliefs about the firm's belief about ξ; a wage is only generous if the worker believes the firm thought it was generous. Since beliefs enter into payoffs, this is an example of a 'psychological game'. See John D. Geanakoplos, David A. Pearce & Ennio Stacchetti (1989) for details.

Falk & Urs Fischbacher (2006), and Cox, Daniel Friedman & Steven D. Gjerstad (2007a) provide other examples of intentions-based models.[20]

In what follows, we take a reduced-form approach to fairness and assume only that ξ is an increasing function of the wage. Given any ξ, we can define $\overline{w} = \arg\max_{w\geq 30} \pi\left(w, \xi\left(w\right)\right)$ and $\overline{e} = \xi\left(\overline{w}\right)$ to be the outcome predicted by the model. Both outcome-based and intentions-based models can predict $\overline{w} > \underline{w}$ and $\overline{e} > \underline{e}$; however, these fairness theories include neither the updating of firms' beliefs through time nor the possibility that selfish workers may imitate fair-minded types. Instead, they predict the cooperative outcome in every period, including the last. Thus we have our second prediction:

Static Fairness Prediction In every period *including the last*, workers' efforts are an increasing function of the wage. Firms offer high wages in every period ($\overline{w}$) and all workers respond with high effort ($\overline{e}$).

The FRE concept extends the static fairness prediction by allowing beliefs to evolve and selfish workers to imitate. In the simplified version of the labor market game in Section I, a reciprocal type was clearly defined as one who chose $\overline{e}$ in response to $\overline{w}$ and $\underline{e}$ in response to $\underline{w}$. The appropriate notion of a reciprocal type in the larger game of interest is more ambiguous; any of the fairness models discussed above could serve this purpose. Since each model provides a particular response function ξ, we proceed by simply assuming that ξ is an increasing function of the wage offer such that $\overline{w} > \underline{w}$ and $\overline{e} > \underline{e}$.[21] In this way, the FRE concept adds a dynamic component to *any* given fairness model. If firms think it sufficiently likely that enough workers will act according to the fairness model, then firms

[20]Recent experiments (including Fehr, Falk & Fischbacher (2000), McCabe, Rigdon & Smith (2003), Charness (2004), and Cox, Klarita Sadiraj & Vjollca Sadiraj (2007b)) show that second movers are less reciprocal when first movers' actions are chosen randomly by the experimenter, providing support for intentions-based models over outcome-based models of fairness.

[21]In the language of Fudenberg & Maskin (1986), each model of fairness generates a different possible *behavioral type*. We fix one behavioral type as our 'reciprocal' type and proceed.

will offer high wages and selfish workers will (rationally) imitate reciprocal workers until the final period.

Full Reputation Equilibrium (FRE) Prediction #1 In every period *except the last*, workers' efforts are an increasing function of the wage, firms offer high wages ($\overline{w}$), and workers respond with high effort ($\overline{e}$). In the last period, firms offer high wages ($\overline{w}$), a proportion of workers choose high effort ($\overline{e}$), and the remaining workers choose unconditionally low effort ($\underline{e}$).

The difference between FRE Prediction #1 and the Static Fairness Prediction lies entirely in final period behavior. As a result, experimental tests comparing the two predictions will have relatively little power since only a handful of data points are relevant from each session. By changing the parameters of the game, however, we can generate a prediction that strongly separates the two. Recall from Proposition 3 that FRE Prediction #1 can only occur if firms' prior beliefs are above the threshold p^* and the number of firms relative to the number of workers is above the threshold δ^*. If we were to change the payoff functions so that p^* is increased, existence of the FRE becomes less likely, in the sense that only very high prior beliefs (p_1) could support the FRE. If we simultaneously increase δ^* to exceed m/n, then existence becomes impossible. This provides a second, more powerful prediction for testing the FRE.

FRE Prediction #2 If the payoff functions (π and u) are changed so that p^* and δ^* (from equations 1 and 2) are sufficiently increased, then the Static Equilibrium Prediction obtains (firms offer the reservation wage ($\underline{w}$) and all workers respond with low effort ($\underline{e}$) in every period).

In the Static Fairness Prediction, changing the payoff functions may change the increasing function (ξ) that relates effort to wages and, consequently, the values of $\overline{w}$ and $\overline{e}$, but the comparison against FRE Prediction #2 remains powerful as long as $\overline{w} > \underline{w}$ and $\overline{e} > \underline{e}$.

Note that beliefs about others' types (hence, identities) play no rule in the static equilibrium or, by assumption, in the static fairness model. Comparing Propositions 1 and 3, identities matter in the FRE model only when the stereotyping parameter is sufficiently small. Assuming this is not the case, none of the three models predict that behavior will change when moving from anonymous matching to publicly observed matching. This fact will be useful in motivating part of our experimental design.

Public Matching Prediction Behavior does not change between anonymous matching and publicly observed matching.

III Experimental Design

We proceed by taking the design of FKR's well-known experiment on repeated labor markets (Fehr, Kirchsteiger & Riedl (1993)) and modifying it to test our various predictions. In the first treatment, we replicate the FKR experiment exactly, changing only the participants, location, and experimenters. In the second, we allow matchings to be public information to test the Public Matching Prediction. We keep public matching in the third treatment and change the payoff functions (π and u) to evaluate FRE Prediction #2. Finally, we run a session in which the same subjects participate in both the first and third treatments, providing a within-subjects comparison of the two treatment effects. The details of each treatment are outlined below.

Sessions were run at the Laboratory for Experimental Economics and Political Science (EEPS) at the California Institute of Technology using undergraduate students recruited via E-mail.[22] Subjects were randomly divided into two groups of 6 firms and 9 workers. The groups were separated into different rooms, instructions were provided to subjects and read

[22] All individuals who had previously indicated an interest in participating in experiments through the EEPS lab or the Social Science Experimental Laboratory (SSEL) at Caltech were recruited. Subjects were considered eligible if and only if they had not participated in another session related to this project.

aloud.[23] In following FKR's design, the instructions do not use a labor market framing: Subjects are referred to as buyers and sellers and their task is to post prices for a generic good in a market and choose a 'conversion rate' (rather than an effort level) that affects payoffs.[24]

Experimenters transmitted subjects' decisions between rooms via telephone and posted those decisions on the blackboard in each room.[25] When matchings were not anonymous, subject ID numbers were displayed with their decisions. When effort levels were not public information, the worker wrote her effort decision on an index card that was delivered by an experimenter to the appropriate firm in the other room. The payoff functions, available strategy choices, numbers of firms and workers, and the number of periods were all publicly announced in every session.

A Treatment 1: Low Thresholds, Anonymous Matching (LA)

The first treatment, denoted **LA**, is an exact replication of FKR's (1993) experiment, including the use of subject instructions published in that study. Wage offers are public information, but effort choices are private and matching is anonymous.[26] The payoffs are given by

$$\pi_l(w,e) = (126 - w)\,(e/10) \tag{5}$$

[23]Copies of these instructions appear in the online appendix.

[24]FKR use the term 'conversion rate' to emphasize that sellers, by their choice of e, are choosing the percentage of $(126 - w)$ their buyer will be paid. In treatments where the effort level choice can no longer be thought of as a conversion rate on firms' profits, the generic name 'X' was instead used in the instructions to identify this choice variable.

[25]In later sessions, the market information was projected on a screen (using a popular spreadsheet program) instead of being written on the board. This did not affect the subjects' procedures or available information in any way.

[26]Recall from footnote 13 that Proposition 3 is valid (under a mild assumption) as long as wages are public information.

and

$$u_l\left(w,e\right)=w-26-c\left(e\right), \tag{6}$$

where $c\left(e\right)$ (the cost of effort) is given in Table 1.[27] In the experiment, π_l and u_l are denoted in francs, which are then converted to dollars at a rate of 12 francs per dollar. The stage game equilibrium payoffs are 9.6 and 4 for the firm and worker, respectively.

We argue that this treatment is highly conducive to cooperation under both the fairness and FRE models since the payoff functions give the players substantial leverage over their partners' payoffs. For example, moving from the stage-game equilibrium wage-effort pair $(30,1)$ to the pair $(40,1)$ costs the firm one franc (8.3 cents), but benefits the worker by *ten* francs. Similarly, moving from $(30,1)$ to $(30,2)$ costs the worker one franc but benefits the firm by 9.6 francs. This results in a large set of strategy pairs that Pareto dominate the stage-game equilibrium on which players can coordinate, as can be seen from the graph of indifference curves in Panel A of Figure II.

B Treatment 2: Low Thresholds, Public Matching (LP)

The second treatment, **LP**, is identical to the **LA** treatment, except for the following changes: First, agents observe the player ID numbers associated with all decisions. Second, all effort choices are made public information and are chosen immediately upon accepting a wage offer instead of being chosen privately at the end of each period. Finally, to increase saliency of decisions, the conversion rate between experimental currency and actual payoffs is increased to 4 francs per dollar for the workers and 9 francs per dollar for the firms.[28]

[27] As in FKR, workers actually chose $e/10$ instead of e. We scale by 10 in this manuscript for clarity of exposition.

[28] Subjects were not aware of the conversion rate difference between firms and workers during the experiment.

Since the payoff functions are the same as in the **LA** treatment, so too are the thresholds p^* and δ^*.

C Treatment 3: High Thresholds, Public Matching (HP)

The third treatment, **HP**, alters **LP** by changing the payoff functions to generate higher values of the thresholds p^* and δ^*. Specifically, the payoff for firms is

$$\pi_h(w, e) = 126v(e) - w,$$

where $v(e)$ is given in Table 1.[29] The payoff for workers is

$$u_h(w, e) = w - 26 - 3c(e).$$

The conversion rate of 12 francs per dollar is used for all subjects.

This treatment is less conducive to cooperation, relative to the **LA** treatment. Subjects have much less leverage over the payoffs of their opponents; moving from $(30, 1)$ to $(40, 1)$ transfers 10 francs from the firm to the worker, while moving from $(30, 1)$ to $(30, 2)$ costs the worker 3 francs and benefits the firm by 8.8 francs. Consequently, the set of strategy pairs that Pareto dominate the stage-game equilibrium is strictly smaller, as seen in Panel B of Figure II.

D Treatment Predictions: An Example

The following example illustrates how the predictions of Section A vary across the three treatments. Assuming reciprocal workers exhibit linear inequality-averse preferences (Fehr

[29] The function $v(e)$ can be approximated by $11/40 + 2.9e/40$. To make the decision similar to that of the FKR design, subjects actually chose values of $v(e)$ from the table, which listed the appropriate value of $c(e)$ for each possible $v(e)$.

& Schmidt (1999)) with $\beta_i = 0.4$, the response function ξ in the **LA** and **LP** treatments is flat ($\xi(w) = 1$) for for $w \leq 40$, increases to a maximum of eight as w increases to 80, and then falls sharply beyond $w = 95$. Taken this ξ as given, firms' profit-maximizing wage offer is $\overline{w} = 80$, which results in $\overline{e} = 8$. In the **HP** treatment, ξ increases more slowly (and in larger discrete jumps) to a maximum of eight for $w \geq 90$. In this case, firms' profit-maximizing offer is $\overline{w} = 90$, which gives $\overline{e} = 8$.

If the Static Equilibrium Prediction is correct, we expect wage-effort pairs of $(30, 1)$ in every period, regardless of the treatment. If the Static Fairness Prediction (with the above specification of ξ) is correct, we expect $(80, 8)$ in every period of the low-threshold treatments (**LA** and **LP**) and $(90, 8)$ in every period of the high-threshold treatment (**HP**). Note that we do not expect any final period changes in behavior under either the Static Equilibrium or Static Fairness Predictions.

According to Proposition 3, the FRE exists in the anonymous matching treatment (**LA**) if $m/n \geq \delta^*$ and firms' common prior (p_1) lies between p^* and $(1/(1-\gamma))(n/(n-1))p^*$. Using $(\overline{w}, \overline{e}) = (80, 8)$ (from above) and assuming $\gamma = 2/3$, we find that $\delta^* = 0.24$, which is less than $m/n = 6/9$, and that $p^* = 0.155$, which means prior beliefs must lie in $[0.155, 0.524)$. In other words, if firms believe that the percentage of reciprocal workers is between 15.5 and 52.4 percent, then the FRE exists. In this case, FRE Prediction #1 predicts $(80, 8)$ in periods one through eleven and wage offers of 80 followed by a mix of high and low efforts (eight and one) in the final period.

With public matching, Proposition 1 shows that there is no upper bound on firms' prior beliefs (p_1) for the existence of the FRE. Thus, the FRE exists in the **LP** treatment as long as p_1 is at least 15.5 percent. If this is the case, we expect the same behavior as in the **LA** treatment: high wages and efforts until the final period, in which high wages are met by a mix of high and low efforts.

In the **HP** treatment, the calculated thresholds δ^* and p^* are increased to 0.65 and 0.856, respectively. Since δ^* is (slightly) less than m/n, the FRE can still exist, but it now requires that firms believe at least 85.6 percent of workers are reciprocal. If this condition is not met, we apply FRE Prediction #2 and expect minimal wage-effort pairs $(30, 1)$ in every period.[30]

All of the predictions of this example are summarized in Table 2

Although this example predicts higher wage-effort levels under **HP** than either **LA** or **LP**, it is possible to construct models that predict cooperation in **LA** and **LP**, but not in **HP**.[31] The only way to test such a model against the FRE prediction is by examining final-period behavior.

E Treatment Predictions: A General Approach

Since each possible ξ maps into a particular choice of $(\overline{w}, \overline{e})$, we can alternatively characterize the reciprocal type by the stage game equilibrium outcomes rather than by the response function. Then, for each possible equilibrium pair $(\overline{w}, \overline{e})$, we can calculate the thresholds p^* and δ^* and evaluate the size of the parameter set on which the FRE exists. This gives a rough measure of the 'likelihood' of existence (denoted $L(\overline{w}, \overline{e})$) which we can compare across different experimental treatments.

Specifically, we set $L(\overline{w}, \overline{e}) = 0$ if $m/n < \delta^*$ at $(\overline{w}, \overline{e})$ and set $L(\overline{w}, \overline{e})$ equal to the Lebesgue measure of the set of parameters on which the FRE exists, which, from Proposition 3, is defined by $p_1 \geq p^*$ and $\gamma > 1 - (n/(n-1))(p^*/p_1)$. Since increasing p^* tightens the constraint on p_1 while slackening the constraint on γ, it is impossible for existence to occur for all $(p_1, \gamma) \in [0, 1]^2$. In fact, the upper bound for $L(\overline{w}, \overline{e})$ is $\exp\{-(n-1)/n\}$.[32]

Figure III compares the graph of $L(\overline{w}, \overline{e})$ for both the **LA** and **LP** treatments against

[30] In fact, reciprocal workers may not accept a wage of 30 if their weight on disadvantageous inequity, α_i, is at least 0.397, which is true if we require $\alpha_i \geq \beta_i$. In that case, firms will be forced to offer $\underline{w} = 35$ if p_1 lies in $[0.052, 0.856)$. Regardless, we always predict minimal effort.

[31] For example, by scaling up the non-pecuniary term in Rabin's (1993) model, we can predict values of w^* for **LA**, **LP**, and **HP** of 45, 35, and 30, respectively.

[32] This upper bound is derived in the online appendix.

the graph for the **HP** treatment. Since $n = 9$, the function's upper bound is 0.41 in both cases. It is clear that under low threshold treatments (**LA** and **LP**) there are many $(\overline{w}, \overline{e})$ pairs on which FRE existence is possible, and many of them have likelihood values approaching the maximum. In the high threshold treatment (**HP**), existence occurs for only a few $(\overline{w}, \overline{e})$ pairs and all have low likelihood values. We can conclude that we can expect to see FRE behavior in the low threshold treatments, but in the high threshold treatments this prediction is both unlikely and highly sensitive to changes in beliefs or in the parameters of the particular fairness model (which determine $(\overline{w}, \overline{e})$).

F Experimental Sessions

Five sessions were run. In the first session (S1), subjects participated in the **LA** treatment for twelve periods. In the second session (S2), subjects participated in **LP** for twelve periods, while in the third and fourth session (S3 and S4), subjects participated in **HP** for twelve periods. The fifth session (S5) was divided into two parts: First, **HP** was played for six periods. Immediately following, the same subjects read instructions and participated in **LA** for six periods.[33] The treatment-switching design in S5 tests whether or not social norms or reputations developed in **HP** affect behavior in **LA**, which can then be compared to behavior in S1. Note that if only low wages and effort are observed in the first six periods then firms have gained little information about worker's types. This permits the FRE to develop in the second half of the session. If, on the other hand, cooperation emerges in the first five periods and disappears in the sixth, then the 'bubble' is burst and a FRE cannot develop in the second half.

Each session lasted between 90 minutes and two hours. In sessions S1 and S5, subjects earned an average of \$35, while earnings in S2 averaged \$62 due to the reduced exchange

[33] Although subjects were informed that they would participate in two different experiments, they were not given specific information or instructions about the second treatment until the conclusion of the first.

rate. In S3 and S4, average earnings were around $25 because cooperation rates were lower. Individual subjects did not participate in more than one session.

IV Results

All data from all five sessions are presented in Figures IV, V and VI.[34] The general pattern of the data conform to the two predictions of the FRE model: High wages and effort emerge in the **LA** and **LP** treatments, but revert to the stage game equilibrium in the final period. Cooperation is drastically reduced in the **HP** treatments and the stage-game equilibrium is the modal outcome.

A The Wage-Effort Relationship

The most robust result across previous experiments is the positive correlation between wages and efforts at the aggregate level. That a large proportion of workers' response functions (denoted ξ above) are increasing in the wage represents a clear failure of the Static Equilibrium Prediction. Although many authors have taken this correlation to be supportive of the Static Fairness Prediction, the correlation is consistent with the FRE predictions as well. Even in the **HP** treatment where the FRE is unlikely to exist, the fact that a percentage of the workers are truly reciprocal implies that a weak positive correlation should still exist as wages vary slightly due to noise.

[34] In session S4, two subjects acting as workers had not been matched with many wage offers in the first several periods and consequently had accumulated very little earnings by the 7^{th} and 8^{th} periods. These subjects, informed that they would not have to pay their losses to the experimenter, began to accept the smallest possible wages and offer the highest possible effort in an attempt to create maximal wealth for the (anonymous) firms. After 4 such actions, one worker was removed from the experiment and the other immediately (and voluntarily) stopped participating. Interviews with subjects revealed that they were frustrated by the open-outcry, first-come, first-served nature of the market, which was perceived as unfair because louder, faster subjects were more likely to get matched with a firm. These 4 data points are removed from analysis, but likely affected beliefs in the market for the remainder of the session.

We examine the wage-effort relationship in each treatment by comparing Pearson correlation coefficients between wages and efforts. The estimated correlations for the **LA**, **LP**, and **HP** treatments are 0.56, 0.64, and 0.61, respectively, and are all significant at the 0.001 level.[35] Since strategies may be history dependent, statistics that aggregate across periods may be misleading and result in biased tests. To avoid this problem we instead aggregate across treatments and estimate correlations for each period separately to find significant correlation (p-values below 0.001) for every period except the first.[36] Although the correlations are significantly positive, it is clear from Figure VII that the estimated slope of the wage-effort relationship is significantly lower in the **HP** treatment (compared to **LA**) and significantly higher in the **LP** treatment.

Positive correlations suggest the existence of reciprocal workers. Although this experiment has little power to distinguish between fairness models, we can say that the correct model should predict flatter response functions (ξ) in the high-ratio treatment (where reciprocity is more costly) and steeper response functions under the public matching treatment. Most existing models predict a flatter response when reciprocity is more costly (see Subsection D,) but few capture the relevance of anonymity.[37]

B Low Threshold Treatments: Cooperative Bubbles

Cooperation clearly develops early and persists at least until the penultimate period in low-threshold (**LA** and **LP**) sessions. Across these two treatments there are 161 accepted wage offers in periods 1 through 11, and only 7 of those are less than 40. The average offer

[35]Non-parametric Spearman rank-order coefficients are 0.4848, 0.6417, and 0.6076 with p-values all less than 0.001.

[36]The reader should be careful to note that inter-period dependencies also introduces autocorrelation among period-by-period hypothesis test results. In other words, that correlation is significantly different in period 10 is likely related to the fact that the difference was also significant in period 9.

[37]One notable exception is the model of social identity by George A. Akerlof and Rachel E. Kranton (2000).

is 68.76 with a standard error of only 1.25. The average of the corresponding 161 effort choices is 4.88 with a standard error of 0.21. Only 23 effort choices are at the minimum, and 18 of these are in response to the lowest wage offer of the period.

The pattern of early-period reciprocity is consistent with both the Static Fairness Prediction and FRE Prediction #1 and offers strong support against the Static Equilibrium Prediction. The only separation between the FRE and Static Fairness Predictions is that, under the FRE, selfish workers should defect in the final period. Aggregating over the final period of each session, 13 of 18 effort choices are at the minimal level. Seven of these are in response to low wage offers and therefore have little power in distinguishing between selfishness and fairness, but 6 out of the remaining 11 workers chose minimal effort despite receiving wages above the reservation wage. Using binomial tests on these data, we cannot reject the claim that the percentage of reciprocal workers in this population lies somewhere between 28 and 72 percent.[38] Although this estimate is imprecise (because we are restricted to using only small fraction of the data,) the existence of selfish workers and role of heterogeneity are apparent.

Finally, to test the claim that efforts 'crash' in the final period, for each period 1 through 11 we compare the effort choices to those of the last period using Wilcoxon rank-sum tests. The p-values of these tests (see Column two of Table 3) reveal that the final-period effort is significantly lower than each of the previous periods. The test of period 11 efforts against period 12 has a p-value of 0.0034. Clearly, the final-period crash in efforts is significant.[39]

One phenomenon not predicted by either the fairness or FRE models is a drop in final period *wage* offers. We offer two possible explanations for this: First, there is some

[38]Formally, we run one-tailed binomial tests on the null hypothesis that $\Pr\left[e_T = \underline{e} | w_T > \underline{w}\right] = q$ for each $q \in [0, 1]$. The null is not rejected when $q \in (0.272, 0.728)$.

[39]The results are less obvious when looking at various measures of the *ratio* of effort to wages. In general, the ratio is lower in the final *two* periods, but statistical significance only obtains for less than half of the period-by-period comparisons. This extra noise is consistent with a model of heterogeneous preferences, but the analysis is complicated by the fact that wages and effort seem to 'crash' slightly earlier than predicted. For example, the average ratio is lowest in period 11 of **LA** and period 12 of **LP**.

evidence that the crash actually occurs one period before the end. This can occur in a reputation equilibrium where selfish workers begin mixing between reciprocity and selfish behavior in later periods (see the discussion on page 7), or it may be a reaction to 'trembles' in which selfish workers inadvertently revealed their type. The second explanation is that workers aren't sure whether firms' prior beliefs are above or below p^*, inducing firms with low prior beliefs (call them 'doubters') to act as if they have high prior beliefs ('believers') for some length of time. Workers can't distinguish doubters from believers in early periods, and if they think enough firms are believers, workers will follow the FRE as specified above. In the final period, doubters must reveal their lack of trust and offer low wages before the selfish firms (whom doubters believe to be numerous) reveal their type and select low efforts.

A second phenomenon not well explained by either model is the apparently greater cooperation in **LP** over **LA**. Period-by-period Wilcoxon tests (see Columns three and four of Table 3) verify that wages under **LP** are significantly greater in each of periods 3 through 12 and efforts are significantly greater in periods 6 through 10. Thus, we reject the Public Matching Prediction of identical behavior between treatments. As mentioned above, it may be that the correct model of fairness incorporates some notion of identity or observability, or it may simply be that individual reputations are somehow stronger than group reputations and can sustain higher levels of cooperation.

C High Threshold Treatments: Cooperation Undone

The most powerful test of the FRE model against static models of fairness is in the switch from low-threshold to high-threshold treatments. If the FRE model is accurate, firms are unlikely to have the prior beliefs necessary to generate cooperation, so we should observe wage offers near the reservation wage and efforts responding in kind. If a static model of fairness is accurate, wages and effort should remain above the static equilibrium level. Under the assumptions of Subsection D, for example, we expect wages of 90 and efforts of 8.

It is clear from Figures IV through VI that wages and efforts are lower under **HP** than under either **LA** or **LP**. Since we are interested in the effect of high thresholds over low thresholds, we compare **HP** against **LP**. Period-by-period Wilcoxon tests (see Columns five and six of Table 3) confirm that wages are significantly lower in the **HP** treatment for all periods except period 2 and efforts are significantly lower in all periods except the first and last.[40]

Of the 169 effort choices in the **HP** treatment, 102 (60.4 percent) are at the minimum level and all 169 choices are below the average effort choice in the **LP** treatment. Minimal effort is observed in 12 of 17 transactions in the penultimate period and in *every* transaction in the final period.[41] There are 26 workers who receive at least one wage offer above the reservation wage ($w > 30$), and 20 of them (77 percent) respond to such an offer with minimal effort ($e = 1$) at least once, though only 5 (19 percent) respond with minimal effort on *every* occasion.[42]

Distinguishing between fair-minded and selfish firms is a more difficult task since the presence of worker heterogeneity makes it less obvious that the reservation wage ($w = 30$) is payoff maximizing. Taking the empirical distribution of workers' responses to each wage offer (aggregated across periods) as given, the reservation wage is in fact the offer with the highest expected payout, at 15.6 francs, but the expected loss of offering either 35 or 40 is

[40]If we aggregate **LA** and **LP** together, wages are significantly lower (at the 5 percent level) in all periods except the last and efforts are lower in all periods except the first.

[41]A regression of effort level on period number gives an estimated slope of -0.096 with a p-value of less than 0.001. A similar regression for wages gives a slope of -0.63 with a p-value of 0.016.

[42]This suggests that a more accurate model would allow workers' types to change through time, switching between selfishness and fairness, perhaps due to learning, mood changes, boredom, a preference for unpredictability, or a conscious search to discover which behavior 'feels right' in this setting. Irrespective of the particulars of the model, type heterogeneity and the presence of selfish behavior are clearly significant.

less than 5 francs (42 cents). Offering a wage above 40 has an expected loss of at least 8.5 francs (71 cents). We observed the reservation wage in 32 percent of the 169 transactions, while wage offers of $w \leq 40$ constitute 68 percent of all observations.[43] Thus, a majority of firms do not display significant other-regarding behavior.

D Switching Treatments: Cooperation Reborn

The most remarkable result comes from session five, where subjects participate in the **HP** treatment for six periods (with this endpoint being public information) and then discover that they will participate in six additional periods under the **LA** treatment. The results (see Figure VI) are clear: Cooperation is nearly absent under **HP**, but emerges quickly under **LA**. It must be the treatment parameters, and not the particular subjects, that determine the extent of cooperation. This result is fully in line with FRE Prediction #2.

Wilcoxon tests of period-by-period differences (comparing each period $t \in \{1, \ldots, 6\}$ to period $t + 6$; see Columns seven and eight of Table 3) show significant differences in wages in periods two through five and significant (or marginally insignificant) differences in efforts in periods two through six. Under the **HP** treatment, 16 of 35 accepted wage offers and 23 of 35 effort choices were at the stage game equilibrium. Under the **LA** treatment, these frequencies drop to 3 of 36 wage offers and 10 of 36 effort choices.

V Full Reputation Equilibrium in Previous Experiments

To test the robustness of the FRE prediction, we can look at the 'likelihood of existence' function (L) derived in Subsection E for various previous studies and see if FRE behavior

[43]Individual firms also display time-varying behavior that makes type classification difficult. For example, 17 of the 18 firms attempt a wage offer of 30 (or less) at least once, but 17 of 18 offer at least one wage of 50 or more. Thus, only 2 firms can be labeled cleanly as one type or the other. It appears that a more realistic model allows for time-varying type identifications.

occurs in those experiments where the likelihood measure is relatively high.

Many authors have employed 'no-loss' profit functions of the form $\pi(w, e) = (v - w)\ e$, where v is a fixed constant.[44] This generally creates a large set of wage-effort pairs that Pareto dominate the equilibrium and relatively high levels of the likelihood measure, as in the **LA** treatment above. High wages and efforts are commonly observed in these settings, with little or no reversion to the stage game equilibrium (for example, see Fehr et al. (1998a), Fehr & Falk (1999), Gächter and Falk (2002), Hannan, Kagel & Moser (2002), and Charness (2004)).[45] This indicates that most or all workers are indeed reciprocal-minded. On the other hand, the data provided by Fehr, Kirchsteiger & Riedl (1998b) show strong signs of a final-period crash under the 'no-loss' payoff specification. In particular, 16 out of 26 workers choose $\underline{e}$ in the final period after high wages and effort are observed in previous periods.[46]

Several experiments have removed the 'no-loss' condition by using quasi-linear profits of the form $\pi(w, e) = ve - w$. This does not necessarily imply that reputation equilibria are eliminated. For example, Panel A of Figure VIII shows the likelihood measure for the experiment of Brandts & Charness (2004), where $\pi(w, e) = 10 - w + 5e$, $u(w, e) = 10 - e + 5w$, and wages and efforts are chosen from $[0, 10]$. From the figure it is clear that the environment supports reputation equilibria, and in fact the data show that high average wages and effort move toward the stage-game equilibrium in the final period.[47]

[44]This functional form is often justified by the observation that subject behavior differs in the domain of losses. By picking $w \leq v$, firms can guarantee non-negative payoffs. See Fehr, Kirchsteiger & Riedl (1993, p. 441).

[45]Gächter & Falk (2002) use exogenous matching and private wages (see footnote 13). When players are matched with a single partner every period, they observe a sharp drop in final-period efforts. When partners change each period, effort is relatively low, but wages remain high. Engelmann & Ortmann (2002) also run a treatment with private wages and find behavior close to the selfish equilibrium.

[46]See the appendix of their paper for this data. Interestingly, wages remain high in one session despite frequent observations of $\underline{e}$ by one player. Although this is not a full reputation equilibrium, it can be supported as a repeated game equilibrium if only one worker is truly selfish, γ is low, and p_1 is accurate.

[47]Individual data are not presented, so it is unclear whether the group collectively chose slightly lower strategies or if the separation predicted by the group reputation model obtained.

Rigdon (2002) and Riedl & Tyran (2005) also use quasi-linear profits. The set of wage-effort pairs that can sustain a full reputation equilibria is smaller and the probabilities of existence are generally lower, as demonstrated by Panels B and C of Figure VIII. In Riedl & Tyran, average wages are constant around 45 in all periods with average efforts around 6, and several sessions feature crashes in effort in the final period.[48] The wage-effort pair $(45, 6)$ can be supported in a full reputation equilibrium, but it does require that firms initially believe that over 88 percent of workers are reciprocal. In Rigdon's experiment, effort decays to equilibrium early in the session, with wages following. Here, workers and firms are either unable to coordinate on a full reputation equilibrium or beliefs and stereotyping parameters are insufficient for such an equilibrium to obtain.

Lynch *et al.* (2001, session 21) use a quasi-linear environment with only two effort choices.[49] Their payoff parameters are closest to those of the **HP** treatment above. Graphing the likelihood measure (Panel D of Figure VIII) demonstrates that at the high effort level ($\overline{e} = 1$), a full reputation equilibrium can only exist for a very small number of wages and is very unlikely. As predicted, wages and effort converge early to the stage game equilibrium. Lynch *et al.* conclude from their data that "a seller's demand depends not only upon his/her own 'reputation' for delivering [high quality], but also upon the market 'reputation' (p. 276)." Thus, the authors acknowledge that group reputations play an important role in these settings.

VI Conclusion

In Section I we developed the full reputation equilibrium (FRE) concept for a model of the labor market. The theory uses a mixture of heterogeneous types (selfish and reciprocal)

[48] See the appendix of their paper for this data.

[49] See also Cason & Gangadharan (2002), who add costly quality certification to the experimental design.

and repeated-game arguments to show that selfish workers will prefer to build a false reputation of being reciprocal in early periods if the future benefit is sufficiently high. In fact, this can persist until the penultimate period. When interactions are anonymous, we must additionally assume that firms ‘stereotype’ the group of workers, believing their types to be positively correlated. If a group of workers knows they are being stereotyped, each becomes responsible for the entire group’s future reputation; if one defects, none are trusted. The veil of anonymity does not hide the individual from future punishments enacted upon the entire group.

The experimental data from this study indicate the existence of both type heterogeneity and repeated game effects. Cooperation developed in early periods is virtually eliminated in the final period. As predicted by the FRE theory, the development of cooperation is sensitive to the parameters of the game. When we weaken the future benefit of early cooperation, little to no cooperation develops. Surprisingly, cooperation can be ‘switched on’ when the game parameters unexpectedly change from the latter design to the former. Thus, cooperation appears to be conditional on the game’s parameters in a way that is predicted by the FRE theory.

This model introduces further testable hypotheses that warrant investigation. Empirical studies of consumer behavior may confirm the existence of stereotypes. For example, do customers who have had bad experiences with one mechanic show reduced demand for auto repairs in general? The stereotype formation process could be studied more directly via belief elicitation experiments or perhaps using fMRI technology. A variety of tests could be constructed to further examine the validity and limits of the stereotyping assumption and help to predict which values of γ are likely for a given environment. On the theoretical front, the introduction of perceived type correlation into the standard repeated game model could be applied to a wide range of domains with incomplete information and anonymity, providing new explanations for observed cooperative behavior in repeated interactions.

*Department of Economics, The Ohio State University, 1945 North High Street, Columbus, OH 43210 and Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213. Email: healy.52@osu.edu. The author is grateful for conversations with and comments from Colin Camerer, Gary Charness, Ernst Fehr, John Ledyard, Charlie Plott, Steve Spear, seminar participants at The Ohio State University and Purdue University, and two anonymous referees. This research was generously funded by Charlie Plott and the Caltech Laboratory for Experimental Economics and Political Science (EEPS). Isa Hafalir helped design and run the early experimental sessions. Additional assistance in the laboratory was provided by Joel Grus and Basit Kahn.

References

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," The Quarterly Journal of Economics, 2000, 115 (3), 715–753.

Andreoni, James and John H. Miller, "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence," The Economic Journal, 1993, 103 (418), 570–585.

Bolton, Gary E. and Axel Ockenfels, "ERC: A Theory of Equity, Reciprocity, and Competition," The American Economic Review, 2000, 90 (1), 166–193.

Brandts, Jordi and Gary Charness, "Do Labour Market Conditions Affect Gift Exchange? Some Experimental Evidence," The Economic Journal, July 2004, 114, 684–708.

Brewer, Marilynn B., Joseph G. Weber, and Barbara Carini, "Person memory in intergroup contexts: Categorization versus individuation," Journal of Personality and Social Psychology, 1995, 69 (1), 29–40.

Camerer, Colin F. and Keith Weigelt, "Experimental Tests of a Sequential Equilibrium Reputation Model," Econometrica, 1988, 56 (1), 1–36.

Cason, Timothy N. and Lata Gangadharan, "Environmental Labeling and Incomplete Consumer Information in Laboratory Markets," Journal of Environmental Economics and Management, 2002, 43 (1), 113–134.

Charness, Gary B., "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," Journal of Labor Economics, 2004, 22 (3), 665–688.

__ and Ernan Haruvy, "Altruism, Equity, and Reciprocity in a Gift Exchange Experiment: An Encompassing Approach," Games and Economic Behavior, 2002, 40 (2), 203–231.

__ and Matthew Rabin, "Understanding Social Preferences With Simple Tests," Quarterly Journal of Economics, 2002, 117 (3), 817–869.

__ , Guillaume Frechette, and John Kagel, "How Robust is Laboratory Gift Exchange?," Experimental Economics, 2004, 7 (2), 189–205.

Corneille, Olivier and Vincent Y. Yzerbyt, "Dependence and the formation of stereotyped beliefs about groups: from interpersonal to intergroup perception," in Craig McGarty, Vincent Y. Yzerbyt, and Russel Spears, eds., Stereotypes as Explanations: The formation of meaningful beliefs about social groups, Cambridge: Cambridge University Press, 2002.

Cox, James C. and Vjollca Sadiraj, "Direct Tests of Models of Social Preferences and Introduction of a New Model," 2005. University of Arizona Working Paper.

__ , Daniel Friedman, and Steven D. Gjerstad, "A Tractable Model of Reciprocity and Fairness," 2007. *Games and Economic Behavior*, forthcoming.

__ , Klarita Sadiraj, and Vjollca Sadiraj, "Implications of Trust, Fear, and Reciprocity for Modeling Economic Behavior," 2007. *Experimental Economics*, forthcoming.

Dufwenberg, Martin and Georg Kirchsteiger, "A Theory of Sequential Reciprocity," Games and Economic Behavior, 2004, 47 (2), 268–298.

Engelmann, Dirk and Andreas Ortmann, "The Robustness of Laboratory Gift Exchange: A Reconsideration," 2002. CERN-EI Working Paper.

Erber, Ralph and Susan T. Fiske, "Outcome dependency and attention to inconsistent information," Journal of Personality and Social Psychology, 1984, 47 (4), 709–726.

Falk, Armin and Urs Fischbacher, "A Theory of Reciprocity," Games and Economic Behavior, 2006, 54 (2), 293–315.

__ , Ernst Fehr, and Urs Fischbacher, "Testing Theories of Fairness – Intentions Matter," 2000. University of Zurich Institute for Empirical Research in Economics Working Paper.

Fehr, Ernst and Armin Falk, "Wage Rigidity in a Competitive Incomplete Contract Market," Journal of Political Economy, 1999, 107 (1), 106–134.

__ and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," Quarterly Journal of Economics, 1999, 114 (3), 817–868.

__ , Erich Kirchler, Andreas Weichbold, and Simon Gachter, "When Social Norms Overpower Competition – Gift Exchange in Experimental Labor Markets," Journal of Labor Economics, 1998, 16 (2), 324–351.

__ , Georg Kirchsteiger, and Arno Riedl, "Does Fairness Prevent Market Clearing? An Experimental Investigation," Quarterly Journal of Economics, 1993, 108 (2), 437–459.

__ , __ , and __ , "Gift Exchange and Reciprocity in Competitive Experimental Markets," European Economic Review, 1998, 42 (1), 1–34.

Fudenberg, Drew and Eric S. Maskin, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," Econometrica, 1986, 54.

Gachter, Simon and Armin Falk, "Reputation and Reciprocity: Consequences for the Labour Relation," Scandanavian Journal of Economics, 2002, 104 (1), 1–27.

Geanakoplos, John D., David G. Pearce, and Ennio S. Stacchetti, "Psychological Games and Sequential Rationality," Games and Economic Behvaior, 1989, 1 (1), 60–79.

Hannan, R. Lynn, John H. Kagel, and Donald V. Moser, "Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity

Differences, and Effort Requests on Behavior," Journal of Labor Economics, 2002, 20 (4), 923–951.

Hayashi, Nahoko, Elinor Ostrom, James Walker, and Toshio Yamagishi, "Reciprocity, Trust, and the Sense of Control A Cross-Societal Study," Rationality and Society, 1999, 11 (1), 27–46.

Healy, Paul J., "Institutions, Incentives and Behavior: Essays in Public Economics and Mechanism Design." PhD dissertation, California Institute of Technology 2005.

Kandori, Michihiro, "Social Norms and Community Enforcement," Review of Economic Studies, 1992, 59 (1), 63–80.

Klein, Benjamin and Keith B. Leffler, "The Role of Market Forces in Assuring Contractual Performance," Journal of Political Economy, 1981, 89 (4).

Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," Journal of Economic Theory, 27 (2), year=1982, pages=245–252).

Lynch, Michael, Ross Miller, Charles R. Plott, and Russell Porter, "Product Quality, Informational Efficiency, and Regulations in Experimental Markets," in Charles R. Plott, ed., Information, Finance and General Equilibrium: Collected Papers on the Experimental Foundations of Economics and Political Sciences, Vol. 3, Northampton, MA: Elgar, 2001.

MacLeod, W. Bentley and James M. Malcomson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," Econometrica, 1989, 57 (2), 447–480.

Macrae, C. Neil and Galen V. Bodenhausen, "Social Cognition: Thinking Categorically about Others," Annual Review of Pscyhology, 2000, 51 (1), 93–120.

McCabe, Kevin A. and Vernon L. Smith, "Goodwill Accounting in Economic Exchange," in Gerd Gigerenzer and Reihard J.R. Selten, eds., Bounded Rationality: The Adaptive Toolbox, Cambridge, MA: MIT Press, 2000, pp. 319–340.

__ , Mary L. Rigdon, and Vernon L. Smith, "Positive Reciprocity and Intentions in Trust Games," Journal of Economic Behavior and Organization, 2003, 52 (2), 267–275.

McEvily, William, Roberto A. Weber, Cristina Bicchieri, and Violet Ho, "Can groups be trusted? An experimental study of collective trust," 2007. In *The Handbook of Trust*, forthcoming.

McKelvey, Richard D. and Thomas R. Palfrey, "An Experimental Study of the Centipede Game," Econometrica, 1992, 60 (4), 803–836.

Neral, John and Jack Ochs, "The Sequential Equilibrium Theory of Reputation Building: A Further Test," Econometrica, 1992, 60 (5), 1151–1169.

Neuberg, Steven L. and Susan T. Fiske, "Motivational Influences on Impression Formation: Outcome dependency, accuracy-driven attention, and individuating processes," Journal of Personality and Social Psychology, 1987, 53 (3), 431–444.

Pendry, Louise F. and C. Neil Macrae, "Stereotypes and Mental Life: The Case of the Motivated but Thwarted Tactician," Journal of Experimental Psychology, 1994, 30 (4), 303–325.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," The American Economic Review, 1993, 83 (5), 1281–1302.

Riedl, Arno and Jean-Robert Tyran, "Tax Liability Side Equivalence in Gift-Exchange Labor Markets," Journal of Public Economics, 2005, 89 (12), 2369–2382.

Rigdon, Mary L., "Efficiency Wages in an Experimental Labor Market," Proceedings of the National Academy of Sciences of the United States of America, 2002, 99 (20), 13348–13351.

Rothgerber, Hank, "External intergroup threat as an antecedent to perceptions of in-group and out-group homogeneity," Journal of Personality and Social Psychology, 1997, 73 (6), 1206–1211.

Ruscher, Janet B., Susan T. Fiske, Hiromi Miki, and Scott F. van Manen, "Individuating processes in competition: interpersonal versus intergroup," Personality and Social Psychology Bulletin, 1991, 17 (6), 595–605.

Selten, Reinhard J. R. and Rolf Stoecker, "End Behavior in Sequences of Finitely Repeated Prisioner's Dilemma Supergames," Journal of Economic Behavior and Organization, 1986, 7 (1), 47–70.

Tajfel, Henri, "Experiments in Intergroup Discrimination," Scientific American, 1970, 223 (5), 96–102.

Yzerbyt, Vincent Y., A. Coull, and S. J. Rocher, "Fencing off the deviant: the role of cognitive resources in the maintenance of stereotypes," Journal of Personality and Social Psychology, 1999, 77 (3), 449–462.

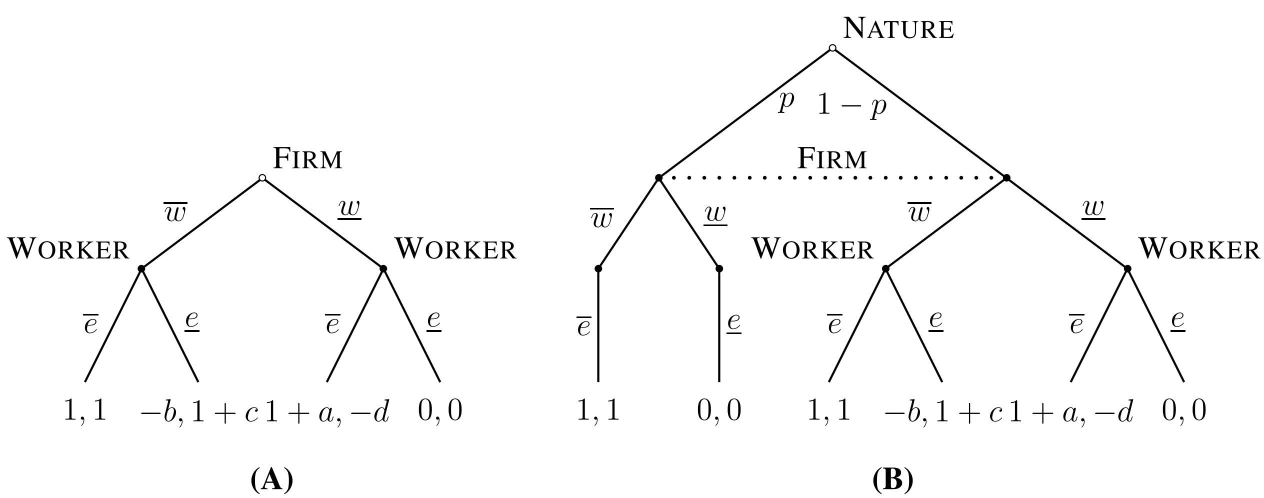
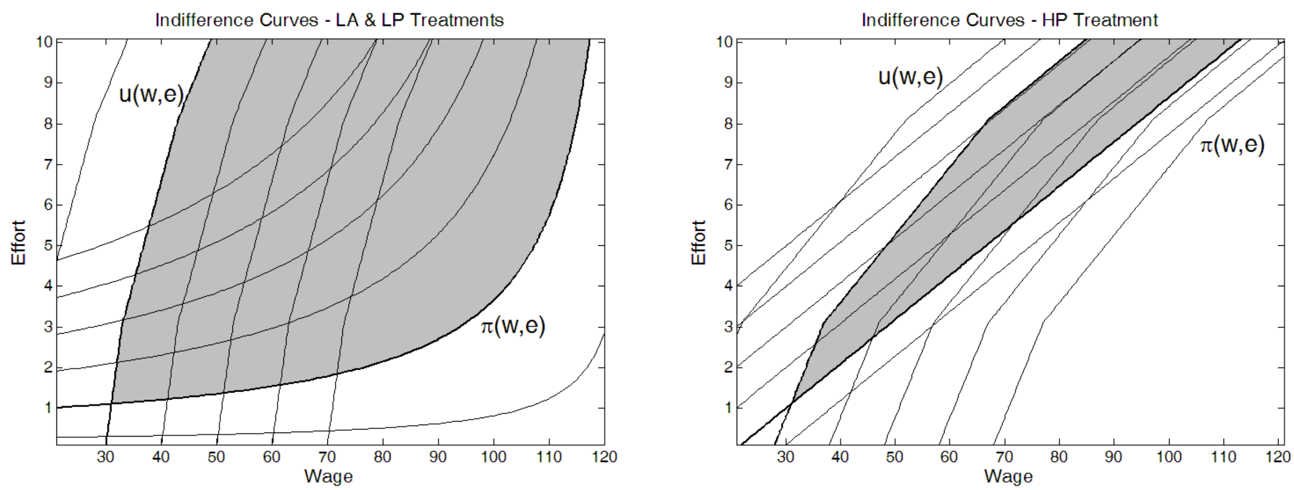
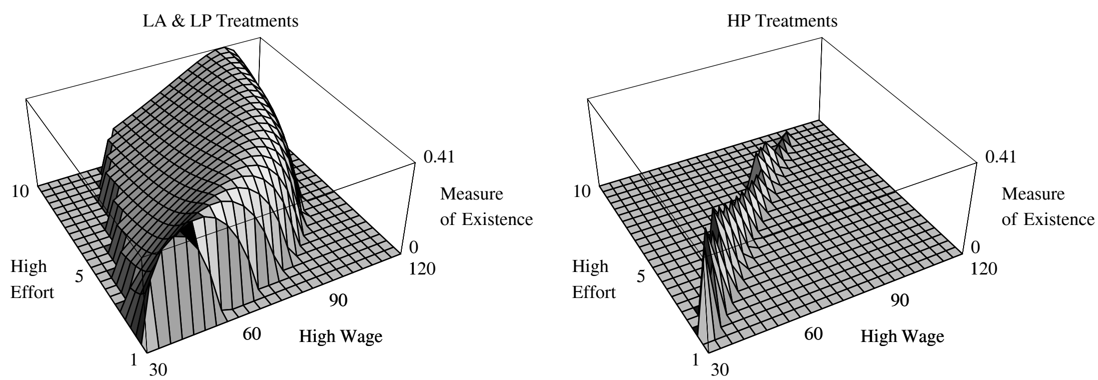


Figure I: A single period of the labor market with (A) a selfish worker, and (B) two possible worker types. Payoffs are normalized with a, b, c, and d strictly positive.



(A) **(B)**

Figure II: Indifference curves for workers and firms in (A) low-threshold treatments (including Fehr *et al.* (1993)) and (B) the high-threshold treatment. Shaded areas Pareto dominate the equilibrium outcome of $(30, 1)$.



(**A**) (**B**)

Figure III: The measure of parameters on which a full reputation equilibrium exists for each $(\overline{w}, \overline{e})$ pair in (A) the **LA** and **LP** treatments and (B) the **HP** treatment. Graphs are scaled to the maximum possible measure.

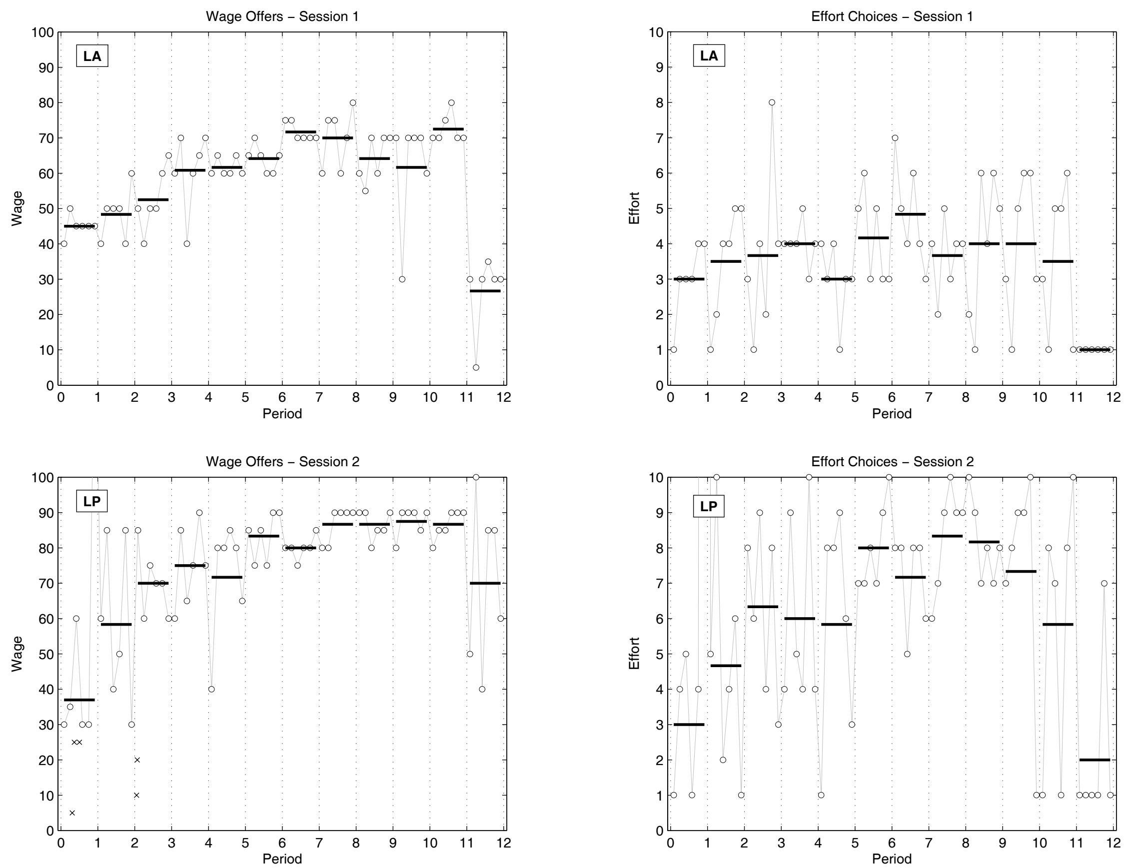


Figure IV: Wage and effort levels across time in sessions S1 (a replication of the Fehr *et al.* (1993) experiment,) and S2 (the same design with individual reputations added). Solid lines represent period averages and x's represent unaccepted bids.

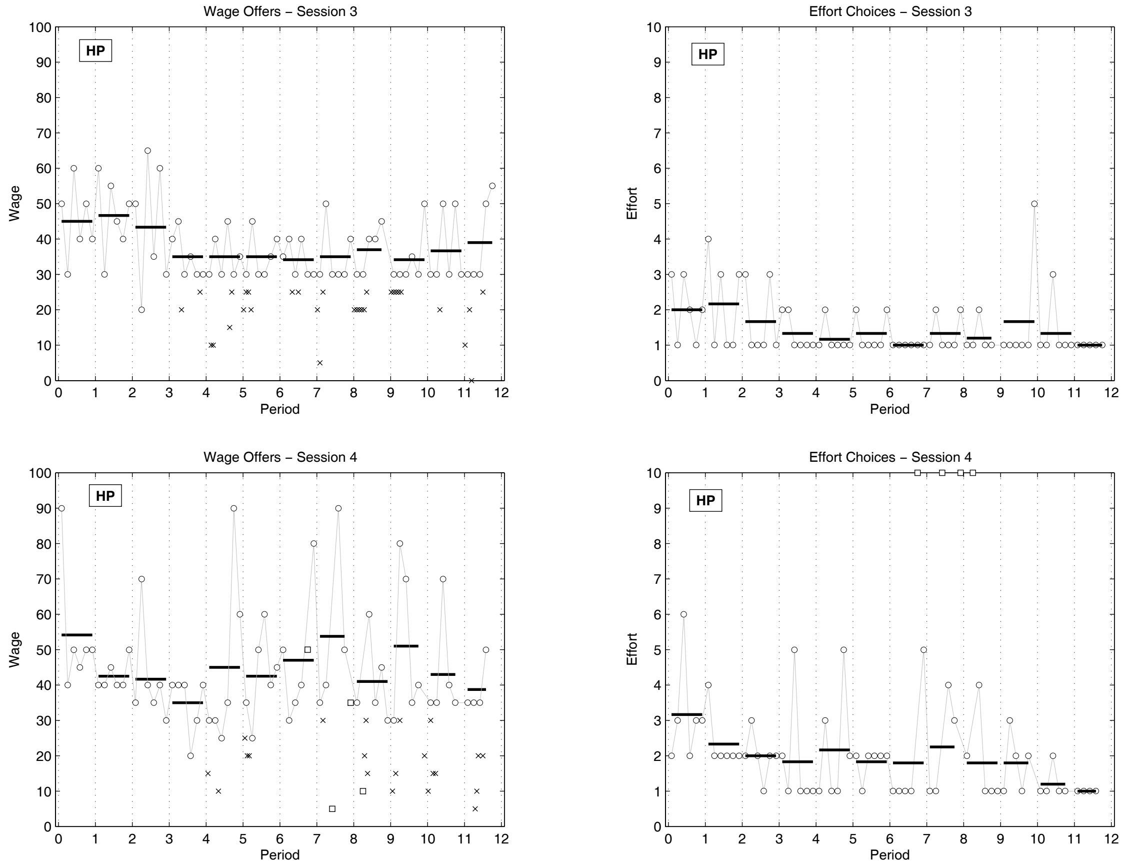


Figure V: Wage and effort levels across time in sessions S3 and S4 (with quasi-linear payoffs). Solid lines represent period averages and x's represent unaccepted bids. Four data points in S4 (represented by squares) are removed from analysis – see footnote 34.

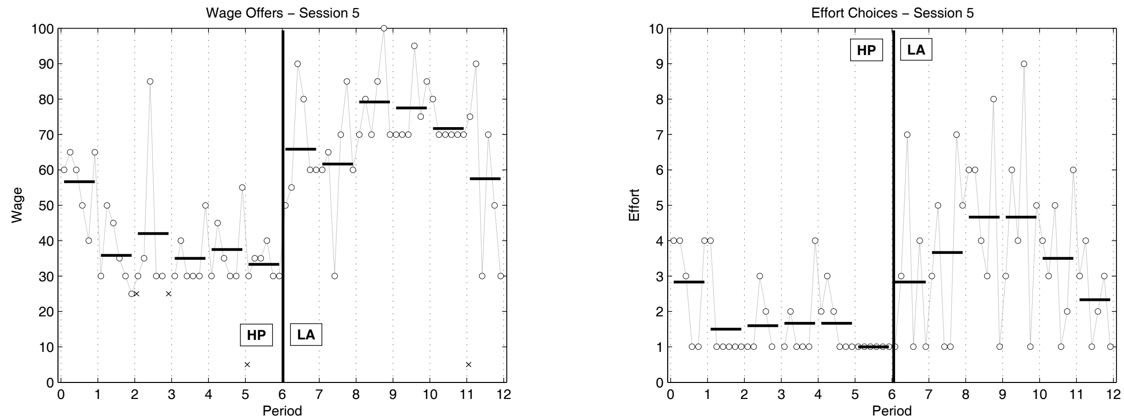


Figure VI: Wage and effort levels across time in session S5 – switching from **HP** to **LA** after period 6. Solid lines represent period averages and x's represent unaccepted bids.

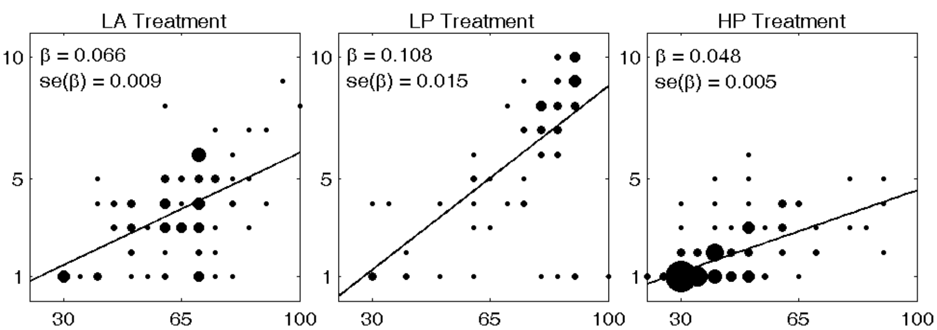


Figure VII: Bubble plots of efforts (ordinate) against wages (abscissa) for each treatment, including regression slope estimates (β) and standard errors ($se(\beta)$).

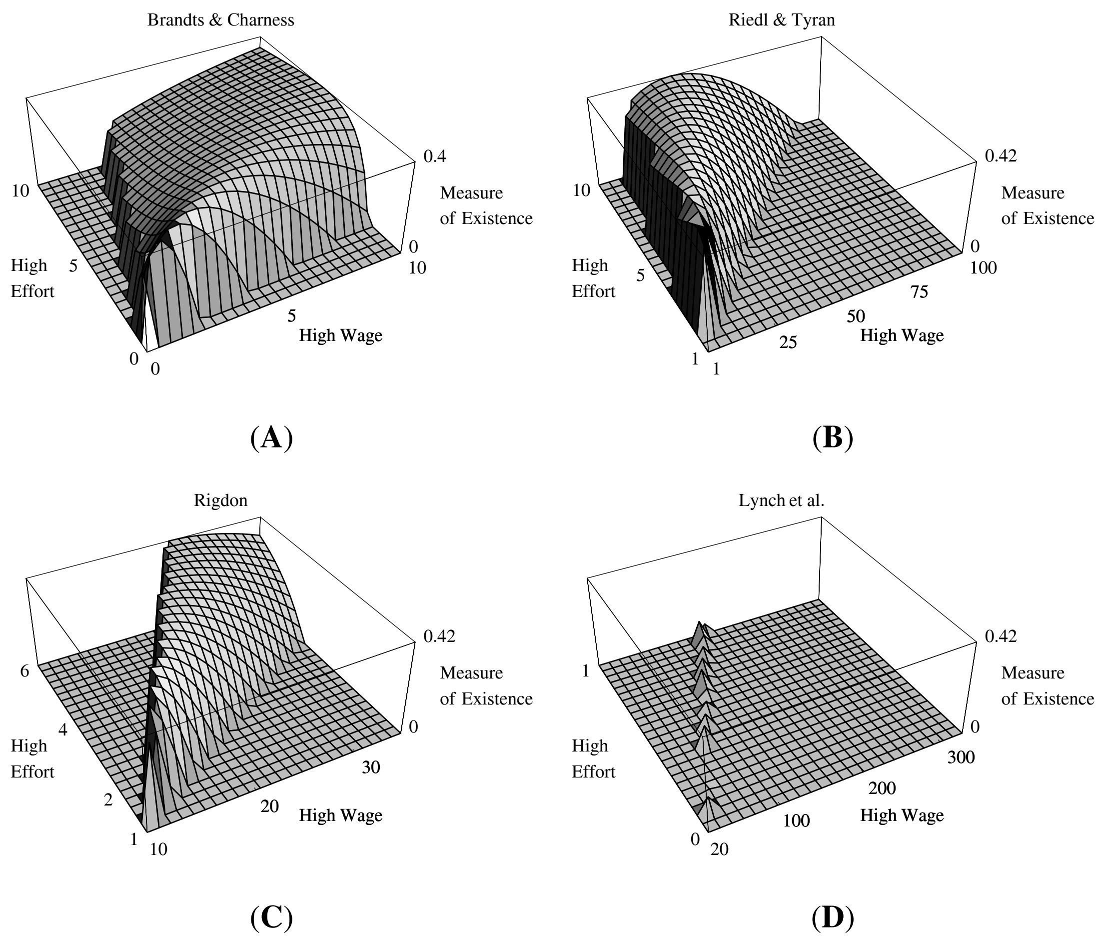


Figure VIII: The measure of parameters on which a full reputation equilibrium exists for each $(\overline{w}, \overline{e})$ pair in (A) the 'excess supply of labor' treatment of Brandts & Charness (2004), (B) Riedl & Tyran (2003), (C) Rigdon (2002), and (D) Lynch *et al.* (2001), where 'effort' is a binary choice. Graphs are scaled to the maximum possible measure.

e	1	2	3	4	5	6	7	8	9	10
$c(e)$	0	1	2	4	6	8	10	12	15	18
$v(e)$	.35	.42	.49	.57	.64	.71	.78	.86	.93	1.0

Table 1: The cost of effort ($c(e)$) and value of effort ($v(e)$).

		Static Equilibrium	Static Fairness	Full Reputation Equilibrium
LA	Periods 1–11	$(30, 1)$	$(80, 8)$	$(80, 8)$
	Period 12	$(30, 1)$	$(80, 8)$	$(80, 1\&8)$
LP	Periods 1–11	$(30, 1)$	$(80, 8)$	$(80, 8)$
	Period 12	$(30, 1)$	$(80, 8)$	$(80, 1\&8)$
HP	Periods 1–11	$(30, 1)$	$(90, 8)$	$(30, 1)$
	Period 12	$(30, 1)$	$(90, 8)$	$(30, 1)$

Table 2: Predictions of wage-effort pairs from an example with Fehr-Schmidt preferences.

	Period $t \geq$ Period 12	**LP** $\geq$ **LA**		**LP** $\geq$ **HP**		**LA** $\geq$ **HP** (S5 only)	
Period	**LA** & **LP** Effort	Wages	Efforts	Wages	Efforts	Wages	Efforts
1	**0.017**	0.130*	0.775	**0.031**	0.619	0.645	0.784
2	**0.002**	0.610	0.576	0.112	**0.035**	**0.024**	0.106
3	**<0.001**	**0.013**	0.113	**0.006**	**<0.001**	**0.039**	*0.0519*
4	**<0.001**	*0.082*	0.212	**<0.001**	**<0.001**	**0.002**	**0.046**
5	**0.002**	**0.048**	0.184	**0.003**	**0.005**	**0.002**	0.102
6	**<0.001**	**0.002**	**0.002**	**<0.001**	**<0.001**	0.156	*0.061*
7	**<0.001**	**0.015**	**0.002**	**<0.001**	**<0.001**	–	–
8	**<0.001**	**<0.001**	**<0.001**	**0.002**	**<0.001**	–	–
9	**<0.001**	**0.006**	**<0.001**	**<0.001**	**<0.001**	–	–
10	**<0.001**	**0.004**	**0.039**	**<0.001**	**0.006**	–	–
11	**0.003**	**<0.001**	0.167	**<0.001**	**0.020**	–	–
12	–	**0.025**	0.878	**0.002**	0.140	–	–

Table 3: p-values of period-by-period Wilcoxon rank-sum tests. Boldfaced values are ≤ 0.05 and italicized values are ≤ 0.10.

*In Period 1, average wages were higher under **LA** than under **LP**.