DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

FAIRNESS, OR JUST GAMBLING ON IT? AN EXPERIMENTAL ANALYSIS OF THE GIFT EXCHANGE GAME

Paul J. Healy

SOCIAL SCIENCE WORKING PAPER 1183

February 2004

Fairness, or Just Gambling on It? An Experimental Analysis of the Gift Exchange Game

Paul J. Healy

Abstract

Fehr, Kirchsteiger and Riedl [12] experimentally test a labor market in which worker effort levels are chosen after wages are set. They observe high wages and effort levels in the repeated game, contrary to the equilibrium prediction. In a similar experimental test of lemons markets, Lynch Miller, Plott and Porter [23] find support for the equilibrium prediction. The current paper finds more evidence of repeated game effects than in previous studies. In a model of incomplete information regarding the reciprocal nature of other players, the FKR design is shown to be conducive to reputation effects while the LMPP design is not.

JEL classification numbers: C72, C91, D83, H41

Key words: Gift exchange game, fairness, experiment, trust, reputation.

Fairness, or Just Gambling on It? An Experimental Analysis of the Gift Exchange Game*

Paul J. Healy†

1 Introduction

Evidence from game theory experiments indicates the presence other-regarding behavior in many laboratory games. In market environments, the existence and observable consequences of such behavior has remained largely an open question, though evidence from labor markets indicates that wages and effort levels are unexpectedly large, perhaps as a result of fairness concerns. Akerlof [2] and Akerlof and Yellen [4] present a neoclassical model of a labor market in which workers, having sentiment for their employer, "acquire utility from an exchange of 'gifts' with the firm (pg 543)" that results in workers choosing higher effort in response to higher wages (as in Solow [27],) so that the optimal wage does not clear the labor market. Fehr, Kirchsteiger, and Riedl [12] (hereafter FKR) test a similar labor market model in the laboratory using a two-stage posted-price labor market followed by an observable and costly effort decision made by the workers. High wages are consistently observed (and accepted) in the market and high effort levels are chosen by workers, leading to the rejection of selfish, rational behavior in favor of a "fair wage-effort" hypothesis in which workers receive utility from the exchange of gifts in the market. Thus, workers deviate from dominant strategies in order to reciprocate. Although the experimental results of FKR are highly consistent with this fairness hypothesis, some later experiments raise questions about the robustness of the result (see, for example, Rigdon [26], Charness, Frechette and Kagel [8] and Engelmann and Ortmann [9].)

Though rarely linked to this literature, Lynch, Miller, Porter and Plott [23] (hereafter LMPP) study a market for "lemons" (see Akerlof [1]) in an environment strikingly similar to that of the labor market experiments. Here, sellers who are able to choose

*The author wishes to thank Charlie Plott and the Laboratory for Experimental Economics and Political Science at Caltech for financial and intellectual support. Jon Eguia, Isa Hafalir, Brian Rogers, Ming Hsu, Basit Kahn, and John Ledyard all provided useful assistance and commentary.

†Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125. Email: pj@hss.caltech.edu. Telephone: (626) 395-4052 Fax: (626) 405-9841

the quality of the good *after* the price is determined consistently deliver the low cost, low quality "lemon" product when no binding quality guarantee has been made. This setting is isomorphic to the incomplete labor contract setting where the quality of labor is determined by the worker after the wage is set. FKR even note that their "experiment can, however, also be interpreted as a stylized version of certain kinds of goods or service markets. In many markets, the price of the good or service is fixed before the good is produced or the service is rendered. If the quality of this good or service cannot be completely specified in the contract, or if the quality is not verifiable by third parties, a similar problem arises as in the labor market (p. 438.)" Furthermore, to avoid framing effects, FKR's subjects are given the generic names of "buyers" and "sellers" rather than "firms" and "workers," making the two experiments nearly identical from the perspective of the subject.

The similarity of these two environments and the difference in observed results warrants further investigation. The goal of the current paper is to determine whether these differences can be explained as sensitivities to the design and payoffs of the game. Three environments are tested in the laboratory, each detailed in Section 2. The first is an exact replication of the FKR design. Although cooperation is observed in most periods, behavior clearly reverts to the stage game prediction in the final period. This is in contrast to the results of FKR, where high wages and effort are maintained in the final period. FKR acknowledge that in the beginning of their experiment, "employers do not know to what extent workers will exhibit reciprocal behavior (p. 444)," implying that a model of incomplete information may be appropriate. In the second environment, buyer and seller ID numbers and decisions were made publicly available to allow for increased opportunities for targeted reciprocation, signalling, and reputation building. Although the observed wage and effort levels were even higher than the first session in early periods, the decisions clearly move toward the stage game equilibrium in the final period. In the third and final environment, the "no-loss" condition on the profit function of the firms imposed by the FKR design is removed in favor of the profit function used in the LMPP design. In the latter design there is no interaction between wages and efforts in the firms' payoffs, while in the former design the marginal cost of wages is an increasing function of effort. Subjects first face the LMPP profit functions and the stage game equilibrium is observed in every period, as it was in the LMPP experiment. The same subjects then participate in a replication of the FKR design and cooperation re-emerges, with a collapse in the final period. Thus, the structure of the payoffs appears to have a significant effect on repeated game behavior in these environments. The experimental design is detailed in Section 3 and results are presented in Section 4.

A simple model of incomplete information is developed in Section 5. There exist sequential equilibria of this repeated game model in which firms offer high wages in every period and both selfish and reciprocal workers respond with high effort levels. This is similar to the model of Kreps, Milgrom, Roberts and Wilson [19]. In the final period, if workers are indeed reciprocal, the results of the FKR experiments obtain. If instead the workers have selfish preferences, workers respond with low effort in the final period, as in the results from the replication of the FKR experiment. The existence of these "reputation" equilibria is sensitive to the payoff parameters in the game. In particular, the payoff structure of the third experiment (motivated by the LMPP study) does not satisfy these conditions. Thus, the experimental data are consistent with the prediction that the stage game equilibrium will obtain in every period.

2 The Gift-Exchange Game

2.1 Description

A single play of a gift exchange game (GEG) has the following structure. A set J of firms and a set I of workers participate in a labor market in which workers' effort levels are observable, but not directly punishable. In the current setting, $|I| > |J|$, so there exists a surplus of available labor. Workers' preferences are strictly increasing in wages and strictly decreasing in effort, while firms' profits are increasing in effort and decreasing in wages.

In the first stage of the game, each firm $j \in J$ posts at most one wage offer w_j in the market that can be updated at any time to any offer greater than all outstanding wage offers in the market. The set of allowable offers is given by $W \cup \{\emptyset\}$, where $W \subseteq \mathbb{R}^+$ is a set of non-negative wage offers and $\{\emptyset\}$ represents no wage offer. For any $c < \sup W$, assume that $\inf \{w \in W : w \geq c\} \in W$. Workers may accept any outstanding offers w_j in the market at any time. After w_j has been accepted by some worker, firm j may not post any further wage offers, so each firm may hire at most one worker. After a fixed amount of time (or after all firms have had a wage offer accepted,) the first stage ends and unmatched agents earn zero profit.

In the second stage, each worker that accepted a wage offer selects an effort level e_i from a linearly ordered set E such that $\inf E \in E$. Formally, a strategy for firm j is $s_j = w_j$ and the strategy for worker i is given by the function $s_i : W \rightarrow \{0, 1\} \times E$ that determines whether to accept w_j and what level of effort to provide given acceptance of w_j. Let $\chi_i \in \{0, 1\}$ indicate whether or not worker i accepts a given wage offer and let $e_i \in E$ denote the effort level chosen. If, for example, E is a subset of the real line and $s_i(20) = (1, 10)$, then worker i chooses to accept any wage offer of 20 and subsequently selects an effort level of 10. Note that this strategy specification is consistent with Solow [27] in the sense that efforts may be a function of wage offers. This fact is a consequence of the extensive form of the game; by making workers the second mover, effort choices are quite naturally dependant on wages.

The monetary payoffs realized by each firm j and worker i are given by π and u, respectively, each taking strategy pairs from $(W \cup \{\emptyset\}) \times (\{0, 1\} \times E)$ and returning a payoff in $\mathbb{R}$. The functions π and u are identical across agents and are common knowledge. The function u is monotone decreasing in e and increasing in w and π is monotone increasing in e and decreasing in w. It is assumed $\pi(\emptyset, (\chi, e)) = u(\emptyset, (\chi, e)) = 0$ for all $e \in E$ and $\chi \in \{0, 1\}$, so that agents not participating in the period receive zero profit. The functions π and u are chosen so that there exists a range of wage-effort pairs that Pareto dominate the equilibrium wage-effort pair, resulting in a prisoners' dilemma-like structure on the payoffs.

2.2 Three Specifications

Three specifications of the GEG are tested experimentally. Each varies in the functional form of agent payoffs and in the information feedback conditions, though the game forms are identical across specifications. In all three games, the sets of allowable effort levels and wage offers are $E = \{1, 2, \ldots, 10\}$ and $W = \{5, 10, 15, \ldots\}$, respectively. Furthermore, $|I| = 9$ and $|J| = 6$, guaranteeing a surplus of at least three workers in every period.

The first variant of this game, denoted GEG1, is the original gift exchange game studied by Fehr, Kirchsteiger and Riedl. Here, the payoffs are given by

$$\pi_1(w, (\chi, e)) = \begin{cases} (126 - w)\ b_1(e) & \text{if} \quad w \neq \emptyset \ \&\ \chi \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where

$$b_1(e) = \frac{1}{10}e$$

and

$$u_1(w, (\chi, e)) = \begin{cases} w - 26 - c_1(e) & \text{if} \quad w \neq \emptyset \ \&\ \chi \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where

$$c_1(e) = \begin{cases} -1 + e & \text{if} \quad e \in \{1, 2, 3\} \\ -4 + 2e & \text{if} \quad e \in \{4, 5, 6, 7\} \\ -12 + 3e & \text{if} \quad e \in \{8, 9, 10\} \end{cases}$$

The functions b_1 and c_1 are the benefit and cost of effort, respectively, for GEG1. This payoff structure has a few interesting features. Note that the cost of effort to the worker is piecewise linear with marginal costs that weakly increase in effort. Also, the marginal benefit of effort to the firm is a linear function of the wage level. By increasing w, the firm decreases its marginal benefit of effort received. Similarly, for changes in effort, the marginal cost of wages varies. This payoff specification is arguably an unnatural model of profits and differs substantially from traditional labor economic theory since the marginal profit of effort is here a function of the wage rate. This design is chosen in the FKR experiments to prevent trusting firms from earning negative profits due to shirking workers. The payoffs π and u are denoted in francs, which are converted to dollars at a rate of 12 francs per dollar.

Another consequence of this specification is that wage payments do not represent a direct transfer of utility since the worker's utility is quasilinear in wages while the firm's profit level is not. If e is not chosen to be 10, then increased wages necessarily result in an increase in total utility. A graph of the level curves of π and u in the space $E \times W$ is given in Figure 1. These level curves are nearly perpendicular near the equilibrium, indicating that both parties have significant control over the profits of the other at little cost to themselves. As strategies move farther from equilibrium, the cost of affecting the other agent's profit increases.

[Figure 1 about here.]

In GEG1, wage offers are displayed for all agents to see, but the identity of the firm offering each wage is known only to the firms. Similarly, the acceptance of wage offers is public information, but the identity of the accepting worker is known only among the workers. Finally, the effort level decision of each matched worker is made after the market is closed and revealed only to the hiring firm. No other firms or workers observe this decision.

The second variant, GEG2, alters the information structure and payoff conversion rates of GEG1. First, all agents observe the player ID number associated with each wage offer and with each worker accepting any given wage offer. Second, effort level decisions are made immediately after a worker accepts a wage offer and this decision is posted (along with the worker's ID number) for all agents to observe. This not only provides information for the formation of reputations across periods, but also allows all agents to observe the realization of strategies chosen by each worker given the accepted wage offer before the market closes. Finally, 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 so that consequences of strategy choices have increased saliency. The payoff functions of GEG2 are identical to GEG1, so that $\pi_2 \equiv \pi_1$ and $u_2 \equiv u_1$. Thus, the fact that wages are not direct transfers between firms and workers and that isoprofit lines are nearly perpendicular near equilibrium are also true in GEG2.

The third variant, GEG3, alters GEG2 to make the payoffs of both agents quasilinear in wages. Furthermore, the cost of effort function is tripled ($c_3 = 3c_1$) to reduce the disparity between the effect of a change in effort on workers and firms. At these higher cost levels, workers cannot dramatically affect the payoff of the firms without significant penalty to their own payoffs. Finally, a linear rescaling of the benefit of effort function is used to adjust payoffs for the experimental environment. Formally, the payoffs for GEG3 are given by

$$\pi_3 (w, (\chi, e)) = \begin{cases} 126\, b_3 (e) - w & \text{if} \quad w \neq \emptyset \;\&\; \chi \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where

$$b_3 (e) \equiv 0.275 + 0.0725\, e$$

is the benefit of the effort realized by the firm[1] and

$$u_3(w,(\chi,e)) = \begin{cases} w - 26 - c_3(e) & \text{if} \quad w \neq \emptyset \ \& \ \chi \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where

$$c_3(e) \equiv \begin{cases} -3 + 3e & \text{if} \quad e \in \{1,2,3\} \\ -12 + 6e & \text{if} \quad e \in \{4,5,6,7\} \\ -36 + 9e & \text{if} \quad e \in \{8,9,10\} \end{cases}$$

This payoff structure reflects a quasilinear environment where wages represent balanced transfers from the firm to the worker, regardless of the effort level chosen. The level curves of the payoff functions are given in Figure 2. In this graph, the level curves are nearly parallel. This indicates that unilateral changes in either the wage or effort have roughly equal effects on both agents' payoffs.

[Figure 2 about here.]

2.3 Stage Game Equilibrium

Define $e_{\min} = \inf E$ and $w^* = \inf\{w \in W : u(w,(1,e_{\min})) \geq 0\}$. Here, $e_{\min}$ is the minimal effort choice and w^* is the minimal wage choice acceptable to a worker choosing $e_{\min}$. In the second stage of the one-shot game, consider the set of outcome-equivalent strategies for each worker i given by

$$S_i^* = \left\{ s_i^*(w) = \begin{cases} (1,e_{\min}) & \text{if} \quad w \geq w^* \\ (0,e) & \text{if} \quad w < w^* \end{cases} : e \in E \right\},$$

where $e_{\min}$ is chosen in response to any acceptable wage offer. The set S_i^* represents the set of undominated worker strategies since u is decreasing in e and $u(w,(1,e)) < u(w,(0,e))$ for all $e \in E$ if $w < w^*$. Note that although the choice of whether to accept a given wage offer is dependant on w, the choice of $e_{\min}$ is not.

In the current set of environments, $\pi(w^*,(1,e_{\min})) > 0$ and π is strictly decreasing in w. The unique subgame perfect equilibrium strategy for each firm j is to offer $w_j = w^*$. Every worker i will accept w^* and choose effort level $e_{\min}$. Note that since $|I| > |J|$, involuntary unemployment will still result since firms may hire at most one worker[2].

[1] In the experiment, $b_3(e)$ was rounded to two decimal places before calculating π. This was done for simplicity, so that the values could be easily placed in a table. However, since $b_3(e)$ is multiplied by 126 in the π formula, these rounding errors maginify to as large as 0.63. There appears to be no difference in analysis between the specifications. Further analysis uses the rounded values.

[2] In equilibrium, workers know that w^* will be the only wage offer in the market and will therefore accept w^* immediately. Allocation of firms to workers is assumed to be random in the situation of multiple simultaneous acceptances, so that the set of unemployed workers will be randomly selected.

There also exists a no-trade equilibrium to this game. If all players refuse to trade, no one player can be made better off by unilaterally offering to trade since participation by both a firm and a worker is necessary to realize non-zero payoffs. However, this equilibrium is Pareto dominated by the more natural equilibrium and also fails the subgame perfection criterion. Unless otherwise indicated, further discussion of the stage game equilibrium will refer only to the equilibrium with trade.

3 Experimental Design

The three designs were tested experimentally at the California Institute of Technology Laboratory for Experimental Economics and Polical Science (EEPS) using Caltech undergraduate students recruited via email. Subjects were randomly divided into two groups of 6 firms and 9 workers, with each group separated into different rooms. The instructions did not make reference to firms, workers, wages, or effort levels. As in the FKR design, subjects were instead called buyers and sellers and their task was to post prices for a good in a market and choose a "conversion rate" (rather than an effort level) that affected payoffs. The term "conversion rate" is used to emphasize that sellers are choosing what percentage of $(126 - w)$ their buyer will actually be paid. In GEG3, the effort level choice can no longer be thought of as a conversion rate on firms' profits, so the generic name "X" was instead used in the instructions to identify this choice variable.

Communication by telephone between two experimenters was used to transmit information between rooms during the market stage of each period. When effort levels were not publicly viewable, the worker wrote his effort decision on an index card that was delivered to the appropriate subject in the other room. The first session (S1) consisted of GEG1 repeated over twelve periods. The second (S2) ran GEG2 for twelve periods. The third session (S3) was divided into two parts. First, GEG3 was played for six periods. Immediately following, the same subjects read instructions and participated in GEG1 for six periods. This treatment-switching design in S3 tests whether or not social norms developed in GEG3 affect behavior in GEG1, which can be compared to behavior in S1. If behavior is substantially different between GEG1 and GEG3 within S3, then differences in the structure of the two games apparently cause differences in behavior. Each session lasted between 90 minutes and two hours. In S1 and S3, subjects earned an average of twenty dollars, while earnings in S2 were as high as \$130 due to the reduced exchange rate.

GEG1 was an exact replication of the FKR experiment using the same set of instructions, procedures, and dollar payoffs. Differences between behavior in the GEG1 in this study and in the previous experiment must be linked to subject pool differences. GEG2 is a slight modification of GEG1 in which agent ID numbers are disclosed publicly so that all agents have perfect recall of all previous actions by all agents, effort choices are made immediately following the wage acceptance decision, and payoff conversion rates are scaled up dramatically. GEG3 modifies GEG2 by making firms' profits quasilinear in wages and adjusting cost schedules appropriately. This is detailed in the previous section.

4 Experimental Results

See Figures 9, 10 and 11 for a complete representation of the data from the three experimental sessions. These results show that effort does appear to be an increasing function of wages. In GEG1 and GEG2, wage-effort pairs are well above equilibrium, but play converges toward the stage game equilibrium in the final period. In GEG3, players are unable to coordinate on high wage-effort pairs and the stage game equilibrium is observed across all periods. In GEG1 played after GEG3, subjects are able to coordinate on high wage-effort pairs, indicating coordination is easier in the latter environment. All of these results strongly favor the repeated game explanation provided above.

Result 1 *In all treatments in all sessions, wages and efforts of matched firms and workers are positively correlated.*

Support. Spearman rank correlation coefficients between wages and effort are calculated for each treatment. For each, the coefficient is estimated to be at or greater than 0.499 and significantly positive at the 0.5% level, indicating a significant positive correlation between wages and effort levels. The values for the correlation coefficient are given in Table B.2.

[Table 1 about here.]

■

The reciprocity of high effort levels for high wages is consistent with both fairness and repeated game hypotheses since this behavior is both "fair" on the part of the worker and supportive of a selfish worker imitating a reciprocal agent in the repeated game. The following set of results will provide evidence for the repated game explanation over the fairness hypothesis.

Result 2 *In a replication of the Fehr, Kirchsteiger and Riedl gift exchange game (GEG1), strategies converge to the stage-game equilibrium in the final period.*

Support. Effort levels for all six transactions in the final period exactly correspond to the dominant strategy prediction of $e = 1$. The observed average wage is actually *below* the dominant strategy prediction. This is due to one worker who, in his rush to accept a wage offer in order to avoid being left out of the market, accidentally accepted a wage offer well below his opportunity cost. Of the remaining five wages in the final period, four are at the subgame-perfect prediction of 30 and the fifth is at 35, which is the smallest possible wage offer above 30. Clearly, the hypothesis that the mean wage is greater than the market-clearing wage cannot be rejected for these six observations. ■

The convergence to the stage game equilibrium in this data provides strong evidence that subjects in this experiment are not completely fairness-minded and appear to be playing a repeated game equilibrium. For any model of fairness to accommodate these results, agents' preference for fairness must decline relative to preferences for their own pecuniary reward in the final period.

Recall that FKR find no convergence to the stage game equilibrium in the final periods of this game. Since GEG1 represents an exact replication of the FKR experiment using different subjects and different experiments, the difference in behavior is most likely due to a subject pool effect, where the percentage of fair-minded agents is significantly lower in the population used for the current study.

Result 3 *In the first 11 periods of the 12-period GEG1, the average wage increases in time, while average effort does not significantly change.*

Support. Each transaction is numbered chronologically from 1 to 72, with 6 transactions per period. The rank correlations between transaction numbers and wages and between transaction numbers and efforts are estimated for the first 66 transactions (11 periods) using Spearman rank correlation coefficients. Rank correlation between wages and transaction number is estimated at 0.6905 with a 2-sided p-value of 1.4×10^{-10}, while correlationg between effort and transaction number is estimated at 0.1711 with a p-value of 0.1696. Rank correlations are also estimated using period number as a proxy for time instead of transaction number, yielding very similar results. ■

Results 1 and 3 imply an interesting correlation structure on time, wages, and effort. In particular, higher-than-average wages are linked to higher-than-average effort, but the upward drift in wages does not result in an upward drift in effort. Thus, reciprocity only appears to be relative to the current average wages and not affected by the absolute magnitude of wage offers. It is argued in the next section that the upward drift in wages is consistent with the repeated game model. If wage offers are too near the stage game equilibrium, rational workers will have an incentive to "defect" earlier since future benefits to cooperation are small. On the other hand, higher wage offers increase the relative benefit of reciprocating and therefore allow the cooperative outcome to persist until the final period.

The following result indicates that the ability of players to achieve outcomes that Pareto dominate the stage game equilibrium is not robust to the payoff specifications.

Result 4 *In GEG3, the minimum effort level is played more often than all other strategies combined and the effort level regresses to the stage game equilibrium strategy in the final two periods.*

Support. Of 35 effort decisions, 23 are at $e = 1$, so the lowest effort level choice represents a majority of observed decisions. The last three effort levels of period 5 and all six effort levels of period 6 are at the minimum of $e = 1$. ■

Result 5 *In GEG3, the firm's subgame perfect equilibrium strategy of the stage game is the modal observation and the frequency of this strategy increases with time.*

Support. The subgame perfect strategy of the stage game for the firm is a wage offer of 30. This occurs in 15 of 35 wage offers in treatment GEG3, which is more often than any other strategy. The second most frequently occurring strategy is a wage offer of 35, which is observed in 6 of the 35 contracts. Of the 15 stage game equilibrium strategies, 10 occur in the final three periods of the treatment. ■

These observations are in contrast to predictions of a pure theory of fairness. If fairness considerations are independent of payoff specifications, then fair-minded workers will always be willing to reciprocate in all gift exchange environments. This is not observed in the GEG3 experiment. The following result indicates that the difference in game parameters is responsible for the the differences in behavior.

Result 6 *In session S3, average wages and effort levels increase after switching from GEG3 to GEG1.*

Support. To avoid problems with non-stationarities in the time series, each wage and effort from GEG3 is compared to the wage and effort from GEG1 with the same time identifier (or "bid number".) These differences are analyzed using a Wilcoxon signed rank sum test. For wages, the GEG1 values are significantly greater than those from GEG3, with an estimated z-statistic of 4.693. Similarly, GEG1 effort choices are significantly greater with a z-statistic of 3.652. Thus, significance in both cases is better than 0.015%. ■

This result is obvious from Figure 11. It is clear that the change in behavior is immediate following the change of the game in the seventh period. The fact that the same group of subjects generates two very different sets of data in two similar games played consecutively implies that the difference in behavior is due differences in the two games. Therefore, the GEG1 appears much more condusive to higher wage-effort pairs than is GEG3.

Result 7 *In GEG2, wages and effort are significantly greater than in GEG1.*

Support. Again, wages and effort were paired between sessions according to their time identifier and a Wilcoxon signed rank sum test was performed in the differences. Wages are significantly higher in GEG2, with a z-statistic of 5.925. Effort is also significantly higher in GEG2, with a z-statistic of 5.401[3]. ■

[3] That this difference is more significant than that documented in the previous result is a consequence of the larger sample size.

Recall that the only differences between GEG2 and GEG1 are that the worker ID numbers and effort levels are publicly displayed with each transaction and the payoffs are increased to 4 francs/dollar for the workers and 9 francs/dollar for the firms. Therefore, some combination of the existence of worker reputations, the common knowledge of all effort levels, and the increased payoffs causes wages and effort levels to increase away from the stage game prediction even more so than in the GEG1 treatment. This result indicates that the underlying reasons for deviations from the stage game prediction are amplified in the high-payoff setting with reputations. This is consistent with a reputation equilibrium hypothesis in which firms are more likely to engage in trusting behavior if the relative payoff is greater. Given that the data converge toward the stage game equilibrium in the final period of this session indicates that fairness is an unlikely explanation of behavior.

5 A Model of Incomplete Information

In this section, a multi-stage game of incomplete information is developed with the same structure as the original gift exchange game. The various outcomes observed across experiments are explainable as equilibria of this game for some profile of agent types. This model is chosen to provide sufficient flexibility to explain the variety of experimental results, while still maintaing tractability by working with standard game-theoretic constructs. Given that a model of pure selfishness and complete information fails to predict the high wages and effort observed in the laboratory while simple models of fairness fail to predict the reversion to stage-game equilibrium in the final period, a mixed model of incomplete information with reasonable behavioral assumptions may prove flexible enough to explain all types of observed behavior. Furthermore, it will be shown that the fragility of the cooperative outcome is explainable as a sensitivity of reputation-building equilibria to payoff parameter choices.

The model proceeds as follows. A set J of firms plays the gift exchange game with a set I of workers. In each period t of the multi-stage game, firms choose among a set of two wage offers, given by $W = \{\underline{w}, \overline{w}\}$, and a particular wage offer of firm j in period t is denoted $w_{j,t}$. In each period t, a fraction $\alpha \equiv |J| / |I|$ of the $|I|$ workers are randomly chosen with equal probability and each is matched with one wage offer. Let $j(i,t)$ denote the firm matched to worker i (if any) in period t and $j(i,t)$ denote the worker matched to firm j. The set of possible actions (or, pure strategies) $e_{i,t}$ for the worker matched with a high wage offer $(w_{j(i,t),t} = \overline{w})$ is the set $E(\overline{w}) = \{\underline{e}, \overline{e}\}$. Workers matched with a low wage offer must choose from the singleton set $E(\underline{w}) = \{\underline{e}\}$. Thus, workers matched with high wages have a choice between high and low effort levels, while workers matched with a low wage are only able to choose low efforts[4]. Unmatched workers have no choice

[4]The restriction of $E(\underline{w}) = \{\underline{e}\}$ is for simplicity of exposition. As will be apparent, allowing $E(\underline{w}) = \{\underline{e}, \overline{e}\}$ does not alter the equilibrium analysis since selecting a high effort in response to a low wage can be assumed to perfectly signal that the agent is *not* reciprocal, causing reduced current-period payoffs and reduced continuation payoffs.

$(E(\emptyset) = \emptyset.)$ Since the strategy space is significantly reduced compared to the full GEG specification, this model is referred to as the mini-GEG.

Workers' types θ_i are selected from the set $\Theta = \left\{\underline{\theta}, \overline{\theta}\right\}$, where $\underline{\theta}$ represents a "selfish" worker and $\overline{\theta}$ represents a "reciprocal" or "fair" worker. Worker payoffs for type $\underline{\theta}$ in each stage t are given by the function $u\left(w_{i(j,t),t}, e_{i,t}|\underline{\theta}\right)$ whose values exactly match the monetary payoff specifications of the game given $\chi = 1$. Type $\overline{\theta}$ workers instead receive payoffs given by

$$u\left(w_{j(i,t),t}, e_{i,t}|\overline{\theta}\right) = \begin{cases} -1 & \text{if} \quad \left(w_{j(i,t),t}, e_{i,t}\right) = \left(\overline{w}, \underline{e}\right) \\ 0 & \text{otherwise} \end{cases}$$

Thus, selfish workers (type $\underline{\theta}$) receive utility for only their monetary payoffs and reciprocal workers (type $\overline{\theta}$) receive a fixed, negative utility in each period in which they do not choose $e_{i,t} = \overline{e}$ given $w_{j(i,t),t} = \overline{w}$. Note that both types of workers receive 0 payoff when they are not selected to participate in period t.

[Figure 3 about here.]

[Figure 4 about here.]

Figure 3 shows the extensive form of one period of the mini-GEG with $W = \{30, 100\}$ and $E = \{1, 10\}$ and payoffs (for firms and selfish workers) taken from the GEG1 parameters. Figure 4 shows the same for the GEG3 payoffs. It is clear that $w_j = \underline{w}$ and $e_i = \underline{e}$ is the unique Nash equilibrium of either stage game. In particular, $\underline{e}$ is a dominant strategy for the workers and $\underline{w}$ is a best response for the firms given $\underline{e}$. However, if the firm believes that $\overline{e}$ will obtain with sufficiently high probability, then $\overline{w}$ is the best response. This depends on the firm's subjective probability that the worker is a reciprocating agent and the relative payoffs for cooperation ($\overline{w}$ and $\overline{e}$), defection ($\overline{w}$ and $\underline{e}$), and equilibrium ($\underline{w}$ and $\underline{e}$)[5].

The extensive form of the repeated game consists of a random move by nature that first selects the type of workers. Each branch of this move is followed by T periods of the mini-GEG, differing only in their payoffs to the workers. Information sets link the firms' decision nodes between these subgames[6]. A sequential equilibrium of this game is defined as a pairing of a strategy profile and a system of beliefs such that each agent is playing optimally at every information set given the strategies of others and the system

[5]The terms cooperation, defection, and equilibrium are taken from the prisoners' dilemma literature since this game is effectively a sequential prisoners' dilemma.

[6]The use of the term subgame is a bit loose since the only proper subgame of a game of incomplete information is the entire game.

of beliefs, and the beliefs at each information set are derived from previous beliefs and action probabilities in accordance with Bayes' Law.

In this model, a "reputation equilibrium" may develop in which firms in the last period, still unsure about the type of the workers, take a gamble by offering a high wage. If workers are selfish, $\underline{e}$ will be chosen, and if workers are reciprocal, $\overline{e}$ is observed. In the penultimate period, if both selfish and reciprocal workers choose high effort with probability 1, then firms who are willing to offer high wages in the final period (given current beliefs) also offer high wages in the current period. Since selfish workers prefer cooperation to equilibrium, they will prefer this high effort strategy. By induction, it is possible that firms offer high wages in every period and workers respond with high effort in every period except the last, at which point low effort is observed by selfish workers and high effort is observed by reciprocal workers. However, the existence of such an equilibrium is sensitive to the parameters of the game.

For notational simplicity, let $A = \pi(\underline{w}, \underline{e}) - \pi(\overline{w}, \underline{e})$ and let $B = \pi(\overline{w}, \overline{e}) - \pi(\overline{w}, \underline{e})$. Here, A/B is a ratio between zero and one indicating the "excess" benefit of equilibrium (compared to defection) as a fraction of the "excess" benefit of cooperation. This serves as an important parameter in measuring the how tempted the firms may be to gamble on the worker types by offering a high wage. If A/B is near zero, then the high wage gamble is more appealing since the equilibrium payoff (which can be guaranteed by offering a low wage) pays only slightly more than the defection outcome that may obtain by gambling. Similarly, define $C = u(\overline{w}, \underline{e}) - u(\overline{w}, \overline{e})$ and $D = u(\overline{w}, \underline{e}) - u(\underline{w}, \underline{e})$, so that C/D represents the worker's *forgone* benefit from cooperation (compared to defection) as a precentage of the forgone benefit from stage-game equilibrium. This ratio measures the selfish workers' relative benefit for maintaining a reputation; if C/D is near zero, cooperation pays nearly as much as defection (which is the best possible outcome for the worker) and much more than equilibrium, so a one-period defection followed by a string of equilibrium payoffs is a significantly worse than a string of cooperative payoffs[7].

First assume that worker types are drawn independently. Let p_1 be the common knowledge prior probability that $\theta_i = \overline{\theta}$ for each i. In the GEG1 information environment, firms cannot distinguish workers and only observe one effort decision per period. Thus, each firm j has an individual probability estimate that a randomly drawn worker will be of the high type in each period t, denoted $p_{j,t}$. In the GEG2 and GEG3 information structure, all firms observe all effort decisions, so beliefs are shared among firms but unique across workers. Thus, beliefs are denoted by $p_{i,t}$.

Consider the GEG1 setting. Since firms are randomly matched with workers, offering a high wage in the final period has an expected payoff of

$$p_{j,T}\pi(\overline{w}, \overline{e}) + (1 - p_{j,T})\pi(\overline{w}, \underline{e}),$$

[7]Time discounting is ignored since subjects are paid off for all periods simultaneously at the end of the experiment.

which is greater than the low-wage payoff of $\pi(\underline{w}, \underline{e})$ if and only if

$$p_{j,T} \geq \frac{A}{B}.$$

Thus, firms are willing to take the high wage gamble in the final period if and only if their average belief is above the A/B threshold.

However, workers in this setting have little incentive to maintain reputations. Since each worker i in period t affects only the beliefs of firm $j(i,t)$, a high-wage, high-effort reputation equilibrium may not be sustainable. For instance, consider the case where all selfish workers are acting reciprocally with probability one in every period. One worker $i \in I$ will deviate from this strategy by playing $e_{i,t} = \underline{e}$ if reducing $p_{j,t} = 0$ does not affect the future decisions of the firms. Thus, only under very restrictive conditions on $p_{j,T-1}$ will all selfish workers prefer to imitate reciprocal workers. Furthermore, additional restrictions on C and D are needed to ensure that selfish workers do in fact prefer to maintain a reputation, since ruining the high belief held by one firm only marginally reduces future payoffs. This is formalzed by the following proposition.

Proposition 8 *If worker decisions are anonymous and private, worker types are independent, and*

$$\frac{C}{D} \geq \frac{1}{|I|},$$

then there does not exist a reputation equilibrium in of the T-period repeated mini-GEG in which all workers choose $\overline{e}$ with probability 1 in all periods $t < T$ and all firms choose $\overline{w}$ in every period.

The proof is provided in Appendix B.1. The condition to ensure non-existence of a pure strategy equilibrium is fairly weak. Letting $\overline{w} = 100$, $\underline{w} = 30$, $\overline{e} = 10$, and $\underline{e} = 1$, this condition is satisfied for both the GEG1 and GEG3 payoffs, so no pure strategy reputation equilibrium of these two mini-games exists. Thus, for the experimental data to be explained by this model, either there exists a mixed strategy equilibrium that predicts frequent high effort levels or the assumptions of the model can be altered to regain the existence of a pure strategy reputation equilibrium.

Attempting to solve for a mixed strategy reputation equilibrium in a setting of independent types where not all workers participate in every period is quite complex. In the GEG3 information setting, given $p_{i,t}$ for some worker i, there are three possibilities for $p_{i,t+1}$. Namely, if i is assigned to a firm and his strategy $q_{i,t} = \Pr[e_{i,t} = \overline{e}]$ results in $\overline{e}$ being chosen, then $p_{i,t+1} > p_{i,t}$. If $\underline{e}$ results, then $p_{i,t+1} = 0$. Finally, if i is unmatched in period t, then $p_{i,t+1} = p_{i,t}$. Now consider the optimal behavior of selfish workers in period $T-1$ and define $\bar{p}_t = (1/|I|)\sum_{i \in I} p_{i,t}$ for each t.

Selfish workers i such that $p_{i,T-1} = 0$ have no incentive to choose $\overline{e}$ with positive probability since their strategy cannot affect future play. Similarly, if $\bar{p}_T \geq A/B$ or $\bar{p}_T < A/B$ regardless of the strategy choice of agent i in period $T-1$, then i will choose

$\underline{e}$ with probability one. The only workers that may have an incentive to offer high wages with positive probability are those workers i such that their strategy choice may affect whether or not $\bar{p}_T \geq A/B$. If worker i such that $p_{i,T-1} > 0$ chooses a strategy of $q_{i,t}$ in each period t, then $\bar{p}_T$ can be decomposed into two parts, given by

$$\bar{p}_T = \bar{p}_{-i,T} + \frac{1}{|I|} \frac{p_1}{p_1 + (1 - p_1) \prod_{s=1}^{T-1} q_{i,s}},$$

where

$$\bar{p}_{-i,T} = \frac{1}{|I|} \sum_{i' : p_{i',T-1} > 0} \frac{p_1}{p_1 + (1 - p_1) \prod_{s=1}^{T-1} q_{i',s}}.$$

For consistency, assume that if $i' \in I$ was not matched with a high wage offer in some period $s < T$, then $q_{i',s} = 1$, even if this is not the strategy i' would have chosen had i' participated. In order for there to exist a strategy $q_{i,T-1} \in (0, 1]$ such that i is a pivotal player, it must be true that

$$\frac{A}{B} - \frac{1}{|I|} < \bar{p}_{-i,T} < \frac{A}{B}.$$

Specifically, if $\bar{p}_{-i,T}$ is on the low end of this range, i must choose a low value of $q_{i,T-1}$ in order to be pivotal, and if $\bar{p}_{-i,T}$ is nearly A/B, then i may choose larger values of $q_{i,T-1}$ and still be pivotal.

However, this condition places a tight range on the possible values of $\bar{p}_{-i,T}$. Note that $\bar{p}_{-i,T}$ is a function of several random variables, such as which agents were selected to play in previous periods and whether or not $\bar{e}$ was realized by workers in previous periods playing mixed strategies. Furthermore, $\bar{p}_{-i,T}$ is not affected by the strategy choices of agent i. Therefore, i is pivotal only if the strategies of others and the randomness of nature line up to satisfy the above condition. Since other workers have only a weak incentive to ensure that agent i is pivotal given that randomness makes the task more difficult, workers will likely prefer to defect from any reputation-building equilibrium. If any reputation equilibrium in mixed strategies does exist, strategies will be path-dependent and outcomes will be highly sensitive to the random draws of nature.

In general, the difficulty for reputation equilibria in this setting is that individual workers have only a marginal impact on the beliefs of the firms. Restrictive conditions are required for any one agent to be pivotal. However, the laboratory evidence implies that a reputation equilibrium in fact obtains. Therefore, it is conjectured that individual workers believe that their strategies have a larger effect on firms' beliefs than would be implied by Bayesian updating.

There are several behavioral explanations for why workers may have these beliefs. For example, perhaps within any population, worker types are in fact perfectly correlated. Subjects playing the role of firms know that types are perfectly correlated, but having never participated in a similar interaction before the experiment, their prior beliefs about the types are non-degenerate.

If selfish workers believe that firms will act as though worker types are perfectly correlated, then an individual worker playing $\underline{e}$ in any period can unravel the reputation equilibrium. This is true whether or not types are in fact perfectly correlated. Thus, individual workers may have a strong incentive to uphold the reputation equilibrium by choosing $\overline{e}$ in early periods. Even though workers in this setting know that their types are not perfectly correlated, the belief that firms believe types are correlated is sufficient to motivate reputation building. However, this argument implies that the subjects randomly selected to play the role of firms have different *a priori* beliefs about others' beliefs.

Another alternative is that firms do not use Bayesian updating to form their posterior beliefs after each period of play. In particular, firms are pessimistic and, upon seeing $\underline{e}$, believe that all workers are of type $\underline{\theta}$ with probability 1. Perhaps this pessimism is a type of "trigger strategy" in beliefs that implicitly rewards selfish agents for acting rationally and consequently increasing the firms' payoffs. If it is common knowledge that firms use this non-Bayesian updating, each worker will know that their decisions are pivotal. Such behavior is analytically equivalent to a model of perfectly correlated types and therefore simple to develop analytically.

Whether types are in fact perfectly correlated or firms are pessimistic non-Bayesians, the assumption that workers believe that any observation of $\underline{e}$ causes firms to believe that all agents are selfish is in fact sufficient to generate theoretical results highly consistent with the observed data. The following proposition formalizes this claim.

Proposition 9 *If $p_1 \geq A/B$, then there exists a pure strategy reputation equilibrium of the T-period repeated mini-GEG with perfectly correlated types if and only if $C/D \leq \alpha$. In this equilibrium, all firms offer $\overline{w}$ in every period, all selfish workers play $\overline{e}$ in every period $t < T$ and $\underline{e}$ in T, and all reciprocal workers play $\overline{e}$ in every period.*

See Appendix B.2 for a proof.

Let $\overline{w} = 100$, $\underline{w} = 30$, $\overline{e} = 10$, and $\underline{e} = 1$. In the GEG1 mini-game, $C/D \approx 0.257$ and $\alpha = 2/3$. Thus, a pure strategy reputation equilibrium exists in this game. Since $A/B \approx 0.299$, if firms have at least a 3/10 prior probability that the workers are of type $\overline{\theta}$, then cooperation should be observed by the firm in every period and by the worker in every period except the last. In the GEG3 mini-game, $C/D \approx 0.771$, which is greater than $\alpha = 2/3$, so no pure strategy reputation equilibrium will exist in this setting. Also, $A/B \approx 0.855$, so much higher prior beliefs would be needed for firms to prefer the high wage gamble in a one-shot game.

These results are in line with the observed experimental data. In GEG1 (and GEG2, which has identical parameters,) reputation equilibria are observed with selfish workers defecting in the final period, while in GEG3, behavior reverts to the stage game equilibrium in the initial periods. Furthermore, results of previous experiments with the GEG1 parameters in which reciprocity is observed in every period are explained as a reputation equilibria with truly fair workers. This model of perfectly correlated types therefore appears to be an accurate predictor of overall behavior in the gift exchange game.

These results extend to a larger set of possible values for $\overline{w}$ and $\overline{e}$. While the appropriate values of $\underline{w}$ and $\underline{e}$ are clearly the stage game equilibrium values w^* and $e_{\min}$, the process of selecting of $\overline{w}$ and $\overline{e}$ is less transparent. If the existence of the reputation equilibrium is robust to this choice for a particular specification, then the result makes more general predictions about behavior under that specification. By Proposition 9, the pure strategy reputation equilibrium exists for some prior p_1 if and only if $A/B < 1$ and $C/D \leq \alpha$. Figures 5 and 6 show the values of $\overline{w}$ and $\overline{e}$ such that $A/B < 1$ and $C/D \leq 2/3$ for the GEG1 and GEG3 specifications, respectively. Figures 7 and 8 display the value of $1 - A/B$ for each of those wage-effort pairs. This value represents the meausure of the interval $[A/B, 1]$ on which prior beliefs support the reputation equilibrium. Larger values of $1 - A/B$ suggest that a wider range of beliefs will generate a reputation equilibrium. These figures demonstrate that the GEG1 specification has many more wage-effort pairs capable of supporting a pure strategy reputation equilibrium than the GEG3 specification, and the condition on prior beliefs is, in general, much less restrictive in GEG1. Therefore, existence of reputation equilibria in the GEG1 are significantly more robust to perturbations of initial beliefs and the choice of $\overline{w}$ and $\overline{e}$.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

In generalizing the analysis of the mini-GEG to the full GEG specification, the concept of a reciprocal worker is less concrete given the larger strategy space. In the previous literature, a positive correlation between effort and wages is taken as an indication of reciprocity. This can be operationalized assuming that reciprocal workers play some pure strategy that is monotone increasing in w. For example, when $E = \{1, 2, \ldots, 10\}$ and $W = \{5, 10, 15, \ldots\}$, assume that reciprocal workers play a roughly linear strategy $\tilde{s}(w)$ given by

$$\tilde{s}(w) = \begin{cases} (0, 1) & \text{if} \quad w < 30 \\ \left(1, \frac{w-20}{10}\right) & \text{if} \quad w \in \{30, 40, \ldots\} \\ \left(1, \frac{w-15}{10}\right) & \text{if} \quad w \in \{35, 45, \ldots\} \end{cases} .$$

If a worker is known to be reciprocal, $\pi_1(w, \tilde{s}(w))$ is maximized at $\tilde{w} = 75$ in GEG1 and at $\tilde{w} = 35$ in GEG3. If the firm is uncertain about the worker's type, $\tilde{w}$ is weakly increasing in his belief. For example, if the firm believes the worker is reciprocal with probability 0.2, then $\tilde{w}$ is 55 in GEG1 and 30 in GEG3. In a pure strategy reputation equilibrium, the firms' beliefs do not update until after the final period, so the value of $\tilde{w}$ associated with the prior belief p_1 can be sustained as a choice for $\overline{w}$ in the repeated game. Thus, $\overline{w} = \tilde{w}$ is an appropriate choice for the mini-GEG analysis. For this particular example of $\tilde{s}$, the GEG1 again supports a reputation equilibrium while the GEG3 does not[8][9].

Gachter and Falk [16] examine reputation formation in the two-player gift exchange game. In sessions where subjects are matched with different partners in every period and decisions are private information, high wages and effort are observed in every period. In sessions where subjects are matched with the same partner every period, behavior is similar until wages (not effort) crash toward the stage game prediction in the final period. Individual subjects are classified as either "reciprocal", an "egoist", or an "imitator." The authors conclude that roughly half of the subjects are reciprocal, 20% are egoists, and 20% are imitators who build false reputations. Although this provides further evidence that both reciprocal types and repeated game effects appear to exist in the repeated gift exchange game, it does not support the hypothesis of truly correlated types among workers.

6 Conclusion

The well-known observation of Fehr, Kirchsteiger and Riedl [12] that fairness may be observed in labor market settings is compared to earlier results of Lynch, Miller, Plott and Porter [23] in which a seemingly identical market with incomplete contracting produces significantly less evidence of fairness. Examination of the payoff structure of the game reveals differences in the incentives of firms between the experiments. In particular, the "no-loss" condition of FKR appears alter the payoff structure in such a way that cooperation through high wages and effort is especially valuable to the agents.

Given that a replication of the FKR experiment resulted in reversion to the stage-game equilibrium in the final period, models of purely selfish agents and models of purely reciprocal agents are inadequate. This end-game effect is observed in a treatment similar to the FKR study with more information feedback. Furthermore, in a treatment with payoffs similar to those of LMPP, strategies converge quickly to the stage game prediction. A simplification of the gift exchange game with incomplete information, correlated types,

[8]If $\tilde{s}(w)$ has a relatively steep positive slope and W is unbounded above, then $\tilde{w}$ may diverge to infinity. In this case, it is necessary to bound W by some $\overline{w}$.

[9]One possibility not explored in this analysis is that there exist some $w' > \tilde{w}$ such that $u(w', \tilde{s}(w'))$ is very negative. By offering a very high w', the firms can induce selfish workers to reject w' and receive zero payoff, thus revealing their type. The optimality of this strategy depends on $\pi(w', \tilde{s}(w')) - \pi(\tilde{w}, \tilde{s}(\tilde{w}))$ and the number of remaining periods.

and only two strategies per agent reveals that these results are explainable by such a model. In particular, a pure strategy reputation equilibrium is expected to exist in the FKR design, but not in the LMPP-type design. Thus, the predictions of the model appear to match the laboratory data.

It is important to note that this model is a generalization of both the standard selfish model and reciprocal or fairness models. The current experiments, when matched with the predictions of this model, find evidence that subjects at Caltech are apparently of the selfish type, while past experiments (including FKR and others) indicate that subjects in Vienna and elsewhere are better modeled as reciprocal or fair-minded agents. However, this model suggests that behavior in gift exchange games is highly sensitive to payoff specifications and whether or not a selfish agent is willing to "gamble" on the fairness of others.

Appendix A Data

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

Appendix B Proofs

B.1 Independent Types & Pure Strategies

Proposition 10 *If worker decisions are anonymous and private, worker types are independent, and*

$$\frac{C}{D} \geq \frac{1}{|I|},$$

then there does not exist a reputation equilibrium in of the T-period repeated mini-GEG in which all workers choose $\overline{e}$ with probability 1 in all periods $t < T$ and all firms choose $\overline{w}$ in every period.

Proof. It suffices to show that in some period, one worker will have an incentive to defect from this proposed equilibrium.

Period T

In the final period, selfish workers prefer to play $e_{i,T} = \underline{e}$ and all reciprocal workers play $e_{i,T} = \overline{w}$.

Any firm j is willing to offer high wages if and only if

$$p_{j,T}\pi\left(\overline{w}, \overline{e}\right) + \left(1 - p_{j,T}\right)\pi\left(\overline{w}, \underline{e}\right) \geq \pi\left(\underline{w}, \underline{e}\right).$$

This occurs if and only if

$$p_{j,T} \geq \frac{A}{B}.$$

If $p_1 \geq A/B$ and all agents follow the proposed equilibrium in all previous periods, then this condition will be satisfied for all j.

Period $T-1$

Assume that the proposed equilibrium has obtained in all previous periods, so $p_{j,T-1} = p_1$ for each j and this is common knowledge. A selfish worker i such that $w_{j(i,T-1),T-1} = \overline{w}$ expects to earn $u(\overline{w}, \overline{e}) + \alpha u(\overline{w}, \underline{e})$ over the final two periods by playing $\overline{e}$ with probability 1. If i instead plays $\overline{e}$ with some probability $q_{i,T-1} < 1$, then if $e_{i,T-1} = \underline{e}$, firm $j(i, T-1)$'s belief in the final period becomes

$$p_{j(i,T-1),T} = \frac{|I|-1}{|I|} p_1$$

and each firm $j \neq j(i, T-1)$'s average belief in the final period is $p_{j,T} = p_1$. If

$$p_1 \geq \frac{|I|}{|I|-1} \frac{A}{B},$$

then the defection does not affect any firm's decision in the final period (since average beliefs will stay above A/B,) so i prefers the deviation. If (and only if)

$$p_1 \in \left[\frac{A}{B}, \frac{|I|}{|I|-1} \frac{A}{B}\right),$$

then the deviating selfish worker i who plays $q_{i,T-1} < 1$ expects to earn

$$q_{i,T-1} \left[u(\overline{w}, \overline{e}) + \alpha u(\overline{w}, \underline{e})\right] + (1 - q_{i,T-1}) \left[u(\overline{w}, \underline{e}) + \alpha \left(\frac{|J|-1}{|J|} u(\overline{w}, \underline{e}) + \frac{1}{|J|} u(\underline{w}, \underline{e})\right)\right]$$

over the last two periods. This is larger than the expected two-period payoff of complying with the equilibrium if and only if

$$\frac{C}{D} \geq \frac{1}{|I|}.$$

This condition is satisfied by assumption, so each selfish worker does have an incentive to defect in period $T-1$ regardless of p_1. Thus, all selfish workers will choose $\underline{e}$ with probability 1 in $T-1$, and no worker has an incentive to deviate. ■

B.2 Perfectly Correlated Types

Proposition 11 *If $p_1 \geq A/B$, then there exists a pure strategy reputation equilibrium of the T-period repeated mini-GEG with perfectly correlated types if and only if $C/D \leq \alpha$.*

In this equilibrium, all firms offer $\overline{w}$ in every period, all selfish workers play $\overline{e}$ in every period $t < T$ and $\underline{e}$ in T, and all reciprocal workers play $\overline{e}$ in every period.

Proof. Assume $p_1 \geq A/B$.

Since types are correlated, if all workers play $q_t = \Pr\left[e_{i,t} = \overline{e} | w_{j(i,t),t} = \overline{w}\right]$ in period t, then

$$p_{t+1} = \frac{p_t}{p_t + (1 - p_t)\, q_t},$$

which is always weakly greater than p_t. If the realization of any worker's strategy is $e_{i,t} = \underline{e}$, then $p_{t+1} = 0$. Thus, firms' beliefs weakly increase until a worker reveals that he is perfectly rational, at which point, all firms know that all workers are selfish and play reverts to the fully selfish subgame perfect equilibrium[10]. In each period t, define $M_t = \left\{i \in I : w_{j(i,t),t} = \overline{w}\right\}$ and note that $\#M_t = |J|$ in every period.

Period T

All selfish workers play $e_{i,T} = \underline{e}$ and all reciprocal workers play $e_{i,T} = \overline{e}$.

If $p_T \geq A/B$, then each firm is willing to gamble on the workers by offering a high wage since

$$p_T \pi\left(\overline{w}, \overline{e}\right) + (1 - p_T)\, \pi\left(\overline{w}, \underline{e}\right) > \pi\left(\underline{w}, \underline{e}\right).$$

If $p_T = 0$, firms offer low wages.

Period $T - 1$

Selfish workers prefer to choose a mixed strategy $q_{T-1} = \Pr\left[e_{i,T-1} = \overline{e} | w_{j(i,T-1),T-1} = \overline{w}\right]$ such that $p_T \geq A/B$ whenever the realization of all $|J|$ workers' strategies in M_{T-1} is $e_{i,T-1} = \overline{e}$. This is acheived when

$$q_{T-1} \leq \left(\frac{p_{T-1}}{1 - p_{T-1}} \frac{1 - (A/B)}{A/B}\right)^{1/|J|}.$$

However, if $p_{T-1} \geq A/B$, then the right-hand side of this expression is weakly greater than 1, so the inequality is satisfied. Thus, regardless of the workers' strategies, firms will offer high wages in the final period whenever all workers in M_{T-1} provide high effort.

[10]Note that in the GEG1 specification, only one firm sees that effort choice of any given worker. However, if firm j observes $\underline{e}$ in period t, then he will offer $\underline{w}$ in period $t + 1$, instantly signalling to the other firms that $p_{t+1} = 0$. Thus, as long as firms such that $p_{t+1} = 0$ offer wages first in period $t + 1$, then the above result holds. If not, then other firms will update their beliefs to zero by period $t + 2$. The former assumption is used here.

Each worker $i \in M_{T-1}$ has an expected payoff over the final two periods given by

$$q_{i,T-1}\left[u\left(\overline{w},\overline{e}\right)+\alpha\left(\begin{array}{c}\prod_{i'\in M_{T-1}\backslash\{i\}} q_{i',T-1}u\left(\overline{w},\underline{e}\right)\\ +\left(1-\prod_{i'\in M_{T-1}\backslash\{i\}} q_{i',T-1}\right)u\left(\underline{w},\underline{e}\right)\end{array}\right)\right]$$
$$+\left(1-q_{i,T-1}\right)\left[u\left(\overline{w},\underline{e}\right)+\alpha u\left(\underline{w},\underline{e}\right)\right].$$

Note that this payoff is increasing in $q_{i,T-1}$ if and only if

$$u\left(\overline{w},\overline{e}\right)+\alpha\left(\begin{array}{c}\prod_{i'\in M_{T-1}\backslash\{i\}} q_{i',T-1}u\left(\overline{w},\underline{e}\right)\\ +\left(1-\prod_{i'\in M_{T-1}\backslash\{i\}} q_{i',T-1}\right)u\left(\underline{w},\underline{e}\right)\end{array}\right)-u\left(\overline{w},\underline{e}\right)-\alpha u\left(\underline{w},\underline{e}\right)\geq 0,$$

which is true if and only if

$$\alpha\left(\prod_{i'\in M_{T-1}\backslash\{i\}} q_{i',T-1}\right)\geq\frac{C}{D}.$$

Note that if $q_{i',T-1}=0$ for any $i'\in M_{T-1}\backslash\{i\}$, then $q_{i,T-1}=0$ is a best response regardless of C and D since i' is fully revealing the workers' type to the firms. If $C/D>\alpha$, then worker i's payoff is decreasing in $q_{i,T-1}$, so i will choose low effort with certainty. Thus, when $C/D>\alpha$, $q_{i,T-1}=0$ for all $i\in M_{T-1}$ must be true in any equilibrium. If $C/D\leq\alpha$, then there exists an equilibrium in which $q_{i,T-1}=1$ for all $i\in M_{T-1}$.

Assume now that $p_{T-1}\geq A/B$, $C/D\leq\alpha$, and firms know all types of workers will offer high effort with probability 1. Suppose each firm j offers a high wage with probability $r_{j,T-1}$, giving firm j and expected profit over the last two periods of

$$r_{j,T-1}\pi\left(\overline{w},\overline{e}\right)+\left(1-r_{j,T-1}\right)\pi\left(\underline{w},\underline{e}\right)+p_{T-1}\pi\left(\overline{w},\overline{e}\right)+\left(1-p_{T-1}\right)\pi\left(\overline{w},\underline{e}\right).$$

This is strictly increasing in $r_{j,T-1}$ (regardless of the strategies of the other firms,) so every firm will choose to offer high wages with probability 1.

Period $T-k$

Assume $p_{T-k}\geq A/B$. As before, regardless of strategies chosen by the workers, if the realization of all workers' strategies in M_{T-k} is high effort, then $p_{T-k+1}\geq A/B$, and high wages and effort will be realized in period $T-k+1$ with probability 1.

Thus, the expected payoff over the last $k+1$ periods to a worker $i\in M_{T-k}$ when each $i'\in M_{T-k}$ plays $q_{i',T-k}$ is

$$q_{i,T-k}\left[u\left(\overline{w},\overline{e}\right)+\alpha\left(\begin{array}{c}\prod_{i'\in M_{T-k}\backslash\{i\}} q_{i',T-2}\left[(k-1)\,u\left(\overline{w},\overline{e}\right)+u\left(\overline{w},\underline{e}\right)\right]\\ +\left(1-\prod_{i'\in M_{T-k}\backslash\{i\}} q_{i',T-2}\right)\left[ku\left(\underline{w},\underline{e}\right)\right]\end{array}\right)\right]$$
$$+\left(1-q_{i,T-k}\right)\left[u\left(\overline{w},\underline{e}\right)+\alpha ku\left(\underline{w},\underline{e}\right)\right].$$

This is increasing in $q_{i,T-k}$ if and only if

$$u(\overline{w},\overline{e})+\alpha\left(\begin{array}{c}\prod_{i'\in M_{T-k}\setminus\{i\}}q_{i',T-2}\left[(k-1)\,u(\overline{w},\overline{e})+u(\overline{w},\underline{e})\right]\\ +\left(1-\prod_{i'\in M_{T-k}\setminus\{i\}}q_{i',T-2}\right)\left[ku(\underline{w},\underline{e})\right]\end{array}\right)$$
$$-\,u(\overline{w},\underline{e})-\alpha ku(\underline{w},\underline{e})>0.$$

As in period $T-1$, this expression is positive if and only if

$$\alpha\left(\prod_{i'\in M_{T-k}\setminus\{i\}}q_{i',T-k}\right)>\frac{C}{D},$$

so when $C/D>\alpha$, only $q_{i,T-k}=0$ for all $i\in M_{T-k}$ can be an equilibrium strategy. If $C/D\le\alpha$, then there exist equilibria in which $q_{i,T-k}=1$ for all $i\in M_{T-k}$.

Assume now that $p_{T-k}\ge A/B$, $C/D\le\alpha$, and firms know all types of workers will offer high effort with probability 1. Suppose each firm j offers a high wage with probability $r_{j,T-k}$, giving firm j and expected profit over the last $k+1$ periods of

$$\begin{aligned}&r_{j,T-k}\pi(\overline{w},\overline{e})+(1-r_{j,T-k})\,\pi(\underline{w},\underline{e})\\&+(k-1)\,\pi(\overline{w},\overline{e})+p_{T-k}\pi(\overline{w},\overline{e})+(\underline{e}-p_{T-k})\,\pi(\overline{w},\underline{e}).\end{aligned}$$

This is strictly increasing in $r_{j,T-k}$ (regardless of the strategies of the other firms,) so every firm will choose to offer high wages with probability 1.

Thus, when $p_1\ge A/B$ and $C/D\le\alpha$, there exists a sequential equilibrium in which high wages and high effort obtain in every period before the last. If $C/D>\alpha$, then there exists no such equilibrium. ■

References

[1] Akerlof, G. (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84: 488-500.

[2] Akerlof, G. (1982). "Labor Contracts as Partial Gift Exchange." *Quarterly Journal of Economics* 97: 543-569.

[3] Akerlof, G. (1984). "Gift Exchange and Efficiency-Wage Theory: Four Views." In *AEA Papers and Proceedings* 74: 79-83.

[4] Akerlof, G. and J. Yellen (1988). "Fairness and Unemployment." In *AEA Papers and Proceedings* 78: 44-49.

[5] Benoit, J.-P. and V. Krishna (1985). "Finitely Repeated Games." *Econometrica* 53: 905-922.

[6] Benoit, J.-P. and V. Krishna (1987). “Nash Equilibria of Finitely Repeated Games.” *International Journal of Game Theory* 16.

[7] Camerer, C. and K. Weigelt (1988). “Experimental Tests of a Sequential Equilibrium Reputation Model.” *Econometrica* 56: 1-36.

[8] Charness, G., G. Frechette and J. Kagel (2001). ”How Robust is Laboratory Gift Exchange?” Mimeo.

[9] Engelmann, D. and A. Ortmann (2002). ”The Robustness of Laboratory Gift Exchange: A Reconsideration.” Mimeo.

[10] Fehr, E. and A. Falk (1999). ”Wage Rigidity in a Competitive Incomplete Contract Market.” *Journal of Political Economy* 107: 106-134.

[11] Fehr, E., S. Gachter and G. Kirchsteiger (1997). ”Reciprocity as a Contract Enforcement Device.” *Econometrica* 65: 833-860.

[12] Fehr, E., G. Kirchsteiger and A. Riedl (1993a). ”Does Fairness Prevent Market Clearing? An Experimental Investigation.” *Quarterly Journal of Economics* 108: 437-459.

[13] Fehr, E., G. Kirchsteiger and A. Riedl (1993b). ”Gift Exchange and Reciprocity in Competitive Experimental Markets.” *European Economic Review* 42: 1-34.

[14] Fehr, E., E. Kirchler, A. Weichbold and S. Gachter (1998). ”When Social Norms Overpower Competition – Gift Exchange in Experimental Labor Markets.” *Journal of Labor Economics* 16: 324-351.

[15] Fudenberg, D. and J. Tirole (1991). *Game Theory.* Cambridge, MA: MIT Press.

[16] Gachter, S. and A. Falk (2002). “Reputation and Reciprocity: Consequences for the Labour Relation.” *Scandanavian Journal of Economics* 104: 1-27.

[17] Gachter, S. and E. Fehr (2002). ”Fairness in the Labour Market: A Survey of Experimental Results.” Pg. 95-132 in *Survey of Experimental Economics, Bargaining, Cooperation, and Election Stock Marekets*, F. Bolle and M. Lehmann-Waffenschmidt, editors. Heidelberg: Physica Verlag.

[18] Healy, P. and C. Noussair (2003). ”Bidding Behavior in the Price is Right Game: An Experimental Study.” Forthcoming in *Journal of Economic Behavior & Organization.*

[19] Kreps, D., P. Milgrom, J. Roberts, and R. Wilson (1982). “Rational Cooperation in the Finitely Repeated Prisoners’ Dilemma.” *Journal of Economic Theory* 27: 245-252.

[20] Kreps, D. and R. Wilson (1982). “Reputation and Imperfect Information.” *Journal of Economic Theory* 27: 253-279.

[21] Kujal, P. and V. Smith (2002). "Fairness and Short Run Price Adjustment in Posted Offer Markets." Forthcoming in *The Handbook of Experimental Economics Results*, C.R. Plott and V. Smith, editors. Amsterdam: North-Holland Elsevier.

[22] Lawler, E. and P. O'Gara (1967). "The Effects of Inequity Produced by Underpayment on Work Output, Work Quality, and Attitudes Toward Work." *Journal of Applied Psychology* 51: 403-410.

[23] Lynch, M., R. Miller, C. R. Plott, and R. Porter (2001). "Product Quality, Informational Efficiency, and Regulations in Experimental Markets." Pg. 269-318 in *Information, Finance and General Equilibrium: Collected Papers on the Experimental Foundations of Economics and Political Science, Volume Three.* Northampton, MA: Elgar.

[24] Palfrey, T. and R. McKelvey (1992). "An Experimental Study of the Centiped Game." *Econometrica* 60: 803-836.

[25] Pritchard, R., M. Dunnetter and D. Jorgenson (1972). "Effects of Perceptions of Equity and Inequity on Worker Performance and Satisfaction." *Journal of Applied Psychology Monograph* 56: 75-94.

[26] Rigdon, M. (2002). "Efficiency wages in an experimental labor market." In *Proceedings of the National Academy of Sciences of the USA* 99(20): 13348-13351.

[27] Solow, R. (1979). "Another Possible Source of Wage Stickiness." *Journal of Macroeconomics.* Winter 1979, 1: 79-82.

[28] Yellen, J. (1984). "Efficiency Wage Models of Unemployment." In *AEA Papers and Proceedings* 74: 200-205.

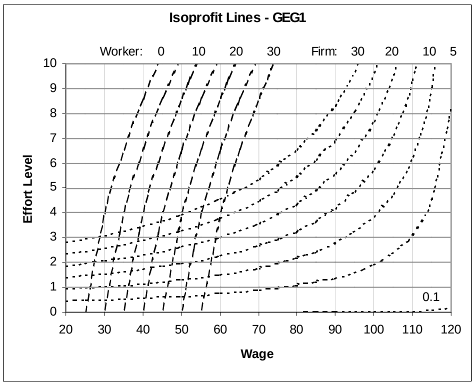

Figure 1: Isoprofit lines for workers and firms in GEG1 - the gift exchange game used by Fehr, Kirchsteiger & Riedl (1993).

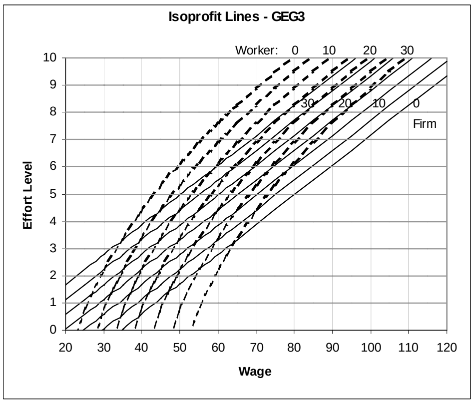

Figure 2: Isoprofit lines for GEG3 - the gift exchange game with quasilinear utilities and higher marginal costs of effort.

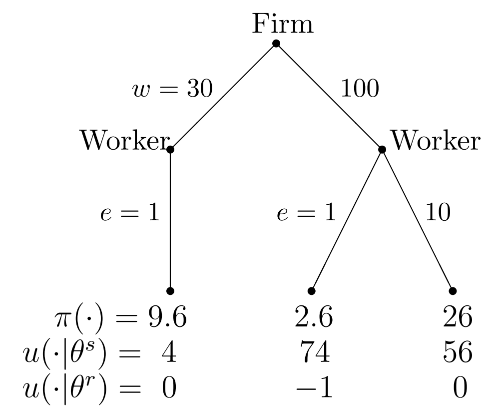

Figure 3: The GEG1 mini-game assuming $W = \{30, 100\}$ and $E = \{1, 10\}$.

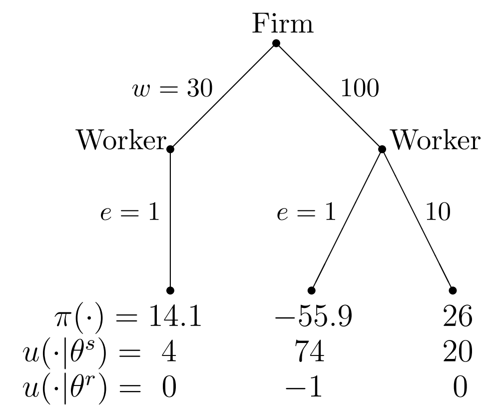

Figure 4: The GEG3 mini-game assuming $W = \{30, 100\}$ and $E = \{1, 10\}$.

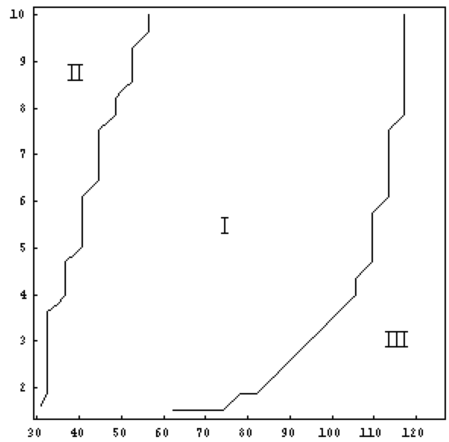

Figure 5: Existence of reputation equilibria with the GEG1 specification for various values of $\overline{w}$ and $\overline{e}$. A pure strategy reputation equilibrium exists in Region I. In Region II, $C/D < 2/3$, so the equilibrium cannot exist. In Region III, $A/B \geq 1$, so no prior beliefs can support the equilibrium.

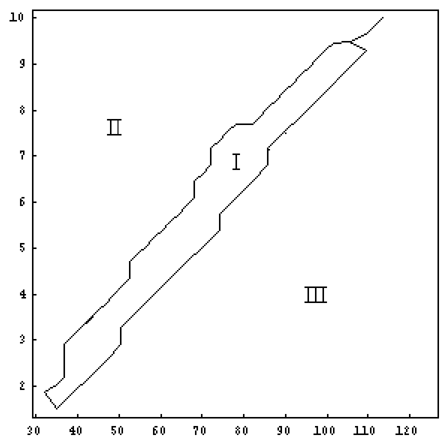

Figure 6: Existence of reputation equilibria with the GEG3 specification for various values of $\overline{w}$ and $\overline{e}$. A pure strategy reputation equilibrium exists in Region I. In Region II, $C/D < 2/3$, so the equilibrium cannot exist. In Region III, $A/B \geq 1$, so no prior beliefs can support the equilibrium.

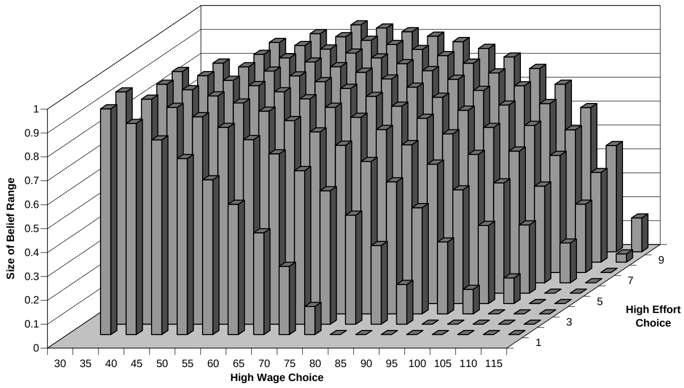

Figure 7: Size of the range of beliefs that support a pure-strategy reputation equilibrium in the GEG1 specification for various values of $\overline{w}$ and $\overline{e}$. This is equivalent to $1 - A/B$ when $A/B < 1$. A zero-sized belief range implies that no pure strategy reputation equilibrium exists.

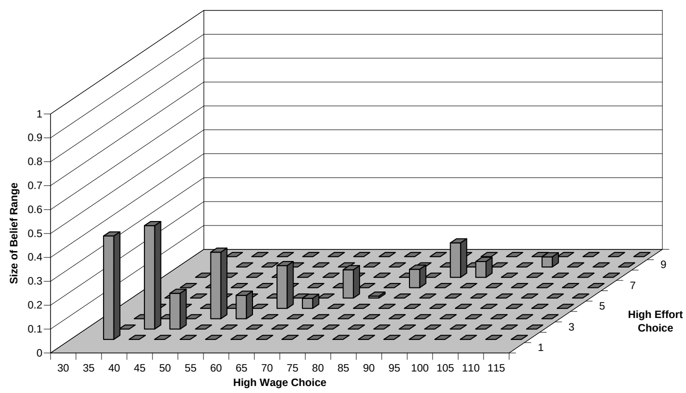

Figure 8: Size of the range of beliefs that support a pure-strategy reputation equilibrium in the GEG3 specification for various values of $\overline{w}$ and $\overline{e}$. This is equivalent to $1 - A/B$ when $A/B < 1$. A zero-sized belief range implies that no pure strategy reputation equilibrium exists.

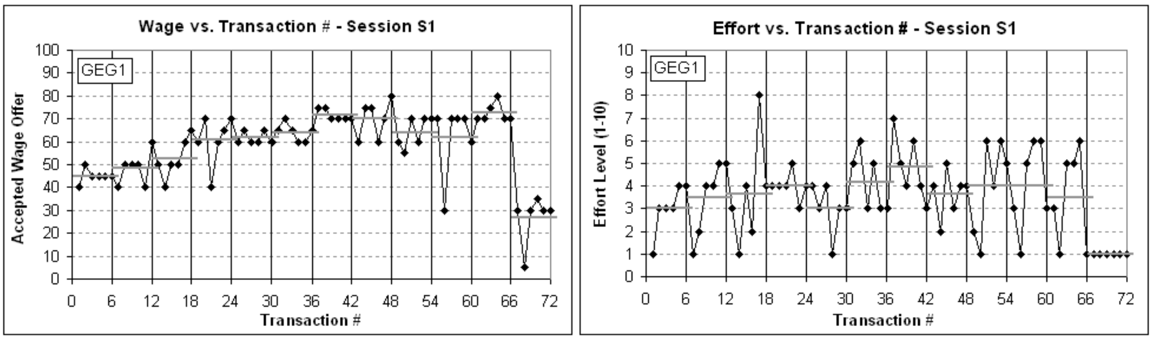

Figure 9: Wage and effort levels across time in session S1 – a replication of the Fehr, Kirchsteiger and Riedl (1993) experiment.

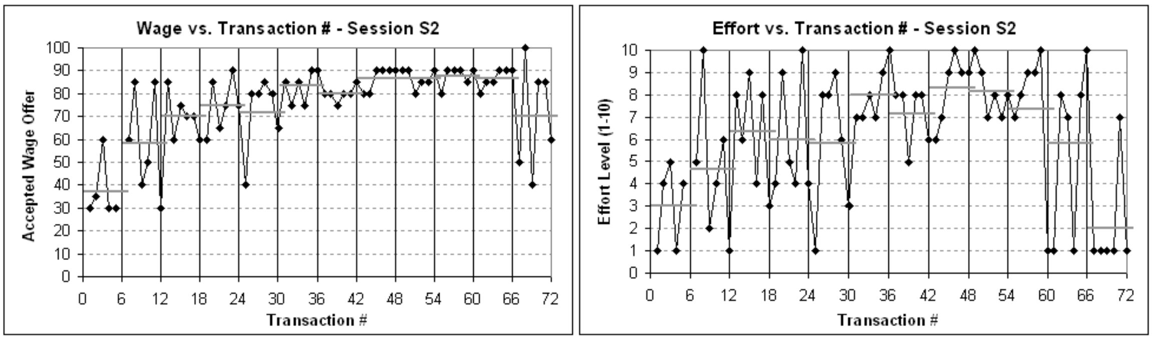

Figure 10: Wage and effort levels across time in session S2.

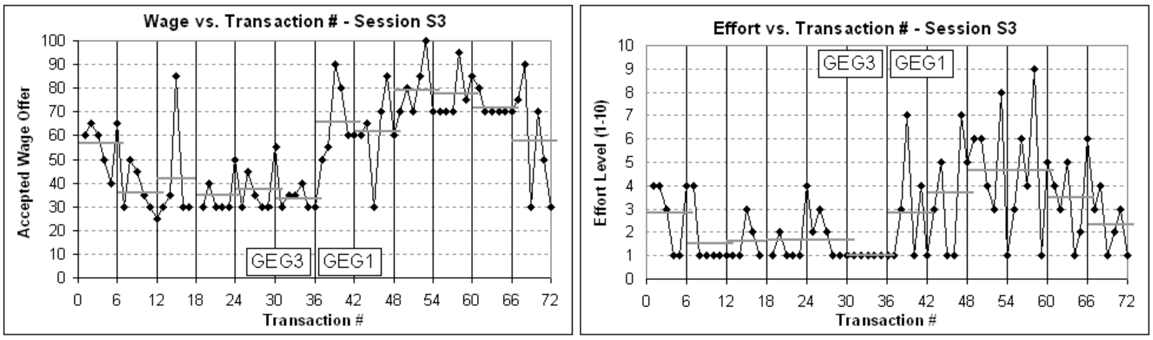

Figure 11: Wage and effort levels across time in session S3.

Session (Treatment)	**Corr**(w, e)	**2-Sided** p**-Value**
S1 (GEG1)	0.546	7.07×10^{-7}
S2 (GEG2)	0.641	1.64×10^{-9}
S3 (GEG3)	0.499	0.0023
S3 (GEG1)	0.511	0.0015

Table 1: Spearman rank correlation coefficients between effort and wages for each treatment.