

TRUST, RECIPROCITY AND PRINCIPAL-AGENT GAMES

Trust Game

- Whereas the Dictator Game, the Ultimatum Game and the Two-Stage Bargaining Game are concerned purely with allocation of a fixed pie, the Trust Game is concerned with pie creation as well.
- In this game, which is sometimes also called an Investment Game or a Sequential Prisoners' Dilemma, there are two players: trustor (investor) and trustee (agent). The trustor is endowed with a budget X and can send a part of the money $Y \leq X$ to the trustee. The money gets multiplied on the way to the trustee by a factor of α . This can be thought of as the investor entrusting his money to the agent who invests the money and the investment has a gross return of α . The trustee may send any part $Z \leq \alpha Y$ of it (including none and all) back to the trustor. This can be thought of as the agent returning the investment with its returns, minus some commission fee, back to the investor, or, if returning nothing, simply stealing the investor's money. The ultimate payoff to the trustor is $X - Y + Z$ and the ultimate payoff to the trustee is $\alpha Y - Z$.
- Assuming self-regarding preferences only and applying the concept of *subgame-perfect equilibrium* to this game, the prediction is that in the second stage, the trustee will send back nothing (i.e., he will be untrustworthy). Knowing this, the trustor will not send any money to the trustee in the first stage (i.e., he will not trust). As a result, there is no trust and no economic surplus is created. Contrary to that, maximization of economic surplus would require the trustor to send all of his money to the trustee (how much the trustee returns is irrelevant for economic surplus).
- This game was first experimentally implemented by **Berg, Dickhaut and McCabe (1995)** with $X = \$10$ and $\alpha = 3$. Their results:
- Again, what we observe is that the actual behavior differs significantly from predictions of subgame-perfect equilibrium with self-regarding preferences.

Reciprocity (Gift Exchange) Game

- This game is similar to the Trust Game, but with the difference that the extent of pie creation now depends on the action of the second-mover rather than the first-mover.
- In this game, there are two players who can be framed as a firm and a worker. In the first stage, the firm offers the worker a wage $w \in [\underline{w}, \bar{w}]$, where \underline{w} is some minimum wage. In the second stage, the worker chooses a costly effort $e \in [\underline{e}, \bar{e}]$, where \underline{e} corresponds to complete shirking. The marginal productivity of this effort is $v > 0$, and hence the firm receives ve . On the other hand, the cost to the employee is given by a strictly increasing and convex function $c(e)$ with $c(\underline{e}) = c'(\underline{e}) = 0$. Hence the ultimate payoff is $ve - w$ for the firm and $w - c(e)$ for the worker.

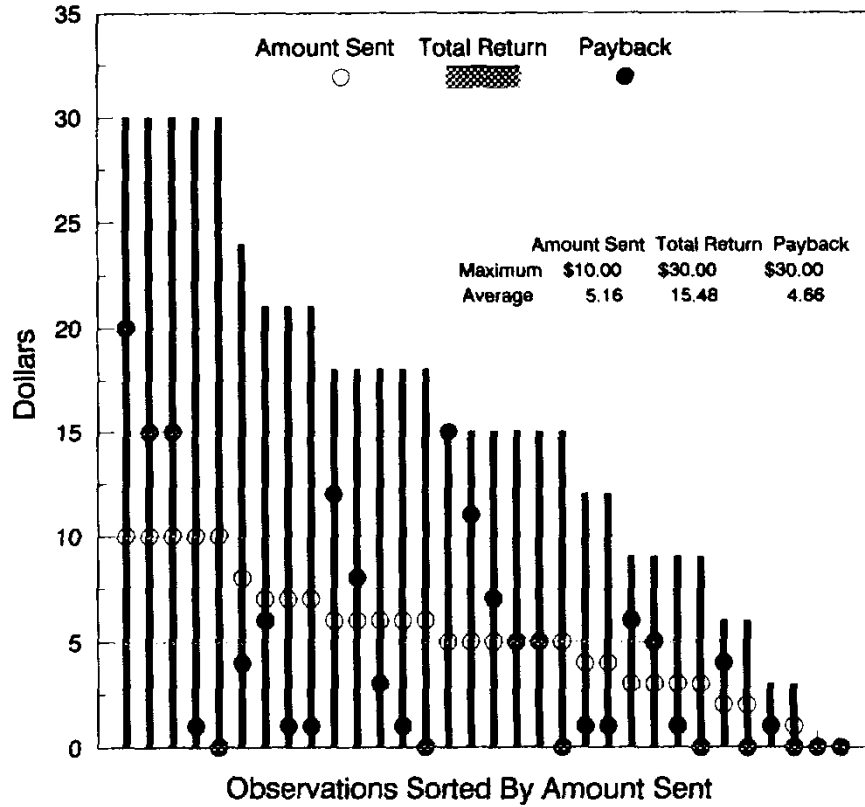


FIG. 2. Trust experiment results showing amount sent (○), total return (■), and payback (●). No history was provided to the subjects.

- Assuming self-regarding preferences only and applying the concept of *subgame-perfect equilibrium* to this game, the prediction is that, in the second stage, the worker will shirk and set $e = \underline{e}$ (note that this is a one-shot game with no reputational concerns). Then in the first stage, knowing that wage-setting makes no difference to employee effort, the firm simply sets $w = \underline{w}$. As a result, no economic surplus is created. Contrary to that, maximization of economic surplus would require setting e where $c'(e) = v$, which definitely implies $e > \underline{e}$.
- This game was first implemented by **Fehr, Kirchsteiger and Riedl (1993)**. Here are their results:
- Note that the slope of effort in the wage rate is so high, that it does indeed pay off to the employer to offer a high wage. Hence the result provides an empirical underpinning of the *efficiency-wage theory* of unemployment.
- Again, what we observe is that the actual behavior differs significantly from predictions of subgame-perfect equilibrium with self-regarding preferences.
- This finding is challenged by later literature, though. For example, **Fehr, Klein and Schmidt (2007)** find that even though effort does indeed increase with the wage, the slope is so small that the employer's payoff decreases with wage. This result is obtained in the environment with different parametrization than in Fehr, Kirchsteiger and Riedl and with an endogenous contract form by the employers, and it is not clear whether a part of the finding could not be attributed to that.

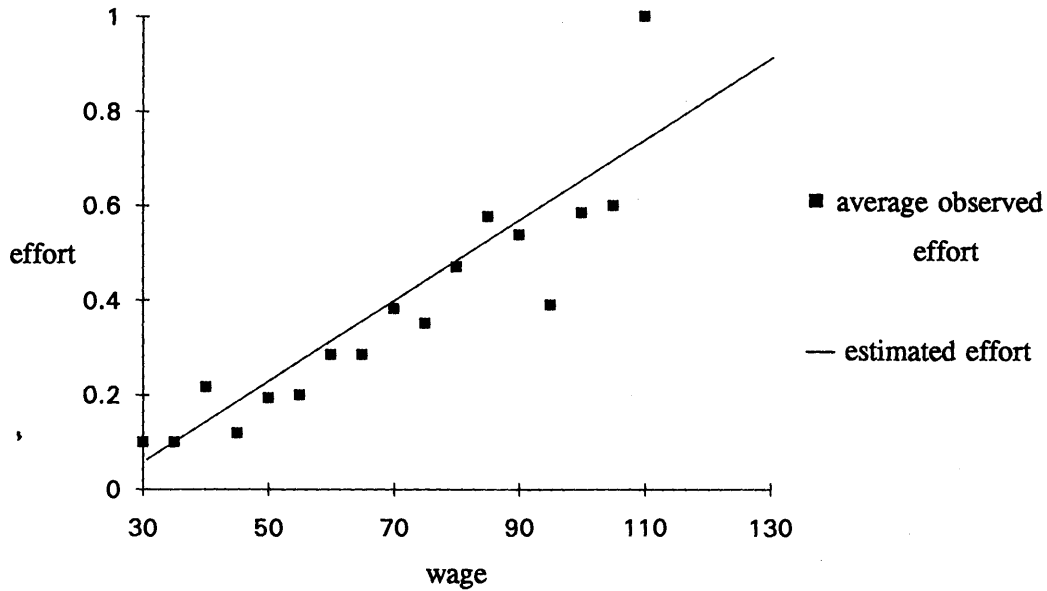


FIGURE I
The Wage-Effort Relation

Principal-Agent Game

- This game extends the Reciprocity Game by adding contractual features by which the firm attempts to implicitly or explicitly induce effort on the part of the worker. Implicit methods include suggested level of effort whereas explicit methods include bonus for a high level of effort (“carrot”) or a penalty for a low level of effort (“stick”).
- This game was implemented by **Fehr, Gächter and Kirchsteiger (1997)**. They compare the Gift Exchange Game (WRT) with the Principal-Agent Game (SRT). In both games, the firm postulates an effort level she would like to obtain from the worker. In the latter game, the employer promises a bonus/fine scheme if the worker sticks to or deviates from the postulated effort level. This bonus/fine is not enforceable *ex post*, though. Both the bonus and the fine are costly for the firm. Results:

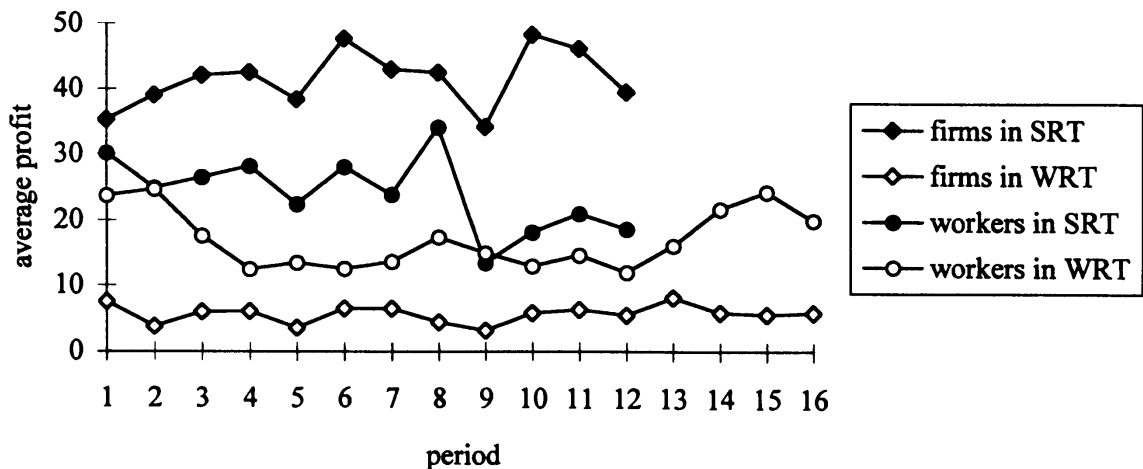


FIGURE 4.—Firms' and workers' gains per trade in the SRT and the WRT.

- So we observe that both postulated and realized effort are higher in the Principal-Agent Game. The authors also demonstrate that more economic surplus is realized in this game and payoffs of both players are higher.
- The results on this comparison are further extended by **Fehr, Klein and Schmidt (2007)**. They give firms an option to choose among three contracts:
 1. *incentive contract* with postulated effort: the firm pays a fixed fee to acquire a monitoring technology for effort; there is a fine for not exerting this effort level
 2. *trust contract* (analogous to the Gift Exchange Game) with postulated effort
 3. *bonus contract* (analogous to the Principal-Agent Game) with postulated effort
- Note that in this environment, the trust contract is a special case of the bonus contract with zero promised bonus.
- They show that in terms of effort and the creation of economic surplus, bonus contract dominates the incentive contract (if incentive compatible), which in turn dominates the trust contract. This is also reflected in payoffs of firms, but not very much for the workers. The firms learn this over time and hence toward the end of the timeline, most offered contracts converge to the more efficient ones from the allowed menu of choices.
- They also provide theoretical explanation for their results and well as why results are different in Fehr, Gächter and Kirchsteiger (1997) by using the inequality-aversion theory of **Fehr and Schmidt (1999)** (see below).