A Reexamination of “The Hidden Return to Incentives”

Jing Davis¹, Steven Schwartz²*, Richard Young³

¹Argyros School of Business and Economics, Chapman University, Orange, CA, USA
²School of Management, Binghamton University, Binghamton, NY, USA
³Fisher College of Business, The Ohio State University, Columbus, OH, USA

Email: *sschwart@binghamton.edu

Abstract

Prior literature has observed a “hidden return to incentives” where principals receive more cooperation from agents when formal incentives are available but not used than when not available. Previous experiments are replicated using a gift-exchange rather than a trust game. Hidden returns to incentives are not observed, and in fact the results show the opposite. Suggestions for future research are provided.

Keywords

Incentives, Trust Game, Gift-Exchange Game

1. Introduction

Several studies have identified a hidden cost of incentives [1] [2] [3]. A hidden cost of incentives occurs when intrinsically motivated behaviors such as reciprocity, honesty and fairness that otherwise would be present in the absence of formal incentives are withheld in their presence. These costs may be accentuated when the choice to use formal incentives is made by an individual seeking to gain from their use and the individual who is the subject of the incentives is aware of the deliberate choice.

Also identified in the literature is a hidden return to incentives (HRTI). HRTI occurs when the decision maker has a choice to use formal incentives, but declines. Fehr & List [2] and Fehr & Rockenbach [3] report on experiments using the trust game [4] and find HRTI. In comparing the condition where incentives are available but not chosen by the trustee to the condition where no incentives are available, reciprocity by the trustee is higher in the former.

In a more recent experiment, [5] use variations on the gift-exchange game to study incentives, but fail to find HRTI. Even more surprising, the introduction
of an option to use incentives reduces reciprocity when the option is not taken compared to when it is unavailable. A potential explanation for this failure to replicate is the inherent differences in the gift-exchange and trust games and how the parties view their relationship.\(^1\) That is, the difference in games may alter how the individuals frame their decision.

We conduct a gift-exchange experiment that investigates HRTI using design choices that more closely follow the trust games experiments than [5].\(^2\) Despite the remaining differences in the games, our results are similar to those of [5], in that the introduction of an option to use incentives reduces reciprocity when not chosen relative to when not available. Our results appear to call into question the robustness of the HRTI. We conclude by suggesting several follow-up experiments.

### 2. Background

Fehr & List [2] and Fehr & Rockenbach [3] report on largely identical experiments based on the trust game. In their Trust game both the trustor and trustee receive the same endowment, \(w\). The trustor chooses an integer \(x \leq w\) to transfer to the trustee. The transferred amount is tripled, becoming \(3x\). The trustor requests that an amount \(y'\) be returned to her; \(y'\) is payoff irrelevant. The trustee learns of \(y'\), and then decides an integer amount \(y \leq 3x\), to be returned to the trustor. The trustor’s and trustee’s earnings are \((w - x + y)\) and \((w + 3x - y)\), respectively. In both studies \(w\) is set to 10 experimental units.

In their Trust with Punishment variation of the game (TWP), the trustor may choose to play the Trust game with the trustee, or alternatively, may choose a formal contract with the trustee wherein the trustee must pay a fine of 4 to the experimenter if \(y < y'\). Before making his decision, the trustee knows which option the trustor has chosen. In TWP, \(y'\) becomes payoff relevant only if the punishment option is taken. Note further the sub-game if the principal refrains from punishment in TWP and the Trust game are identical. The basic result that emerges is that both \(x\) and \(y\) are greater in TWP when the trustor refrains from punishment than in Trust where punishment is not available. Fehr and List label this HRTI.

Kuang & Moser [5] employ a gift-exchange game very similar to that found in [6]. A gift-exchange game differs from a trust game in two primary ways. While a gift-exchange has a transfer from the trustor to the trustee and a return from the trustee to the trustor (in the form or “effort”), it is the choice of the trustee that increases social welfare, not the trustor. Perhaps more importantly, the game is usually framed as an employment relationship rather than just two individuals interacting.

Kuang & Moser [5] have a Gift-Exchange treatment, which uses a basic gift-exchange game without the possibility of punishment that is analogous to Trust in the experiment of [2] and [3]. They also have a Choice treatment, wherein the

---

\(^1\)A detailed description of the gift-exchange game used in our experiment is in the Method section.

\(^2\)This study was approved by the Ohio State behavioral IRB, study number 2015B0085.
trustor chooses between (1) a basic gift-exchange game and (2) a formal "forcing contract", wherein it is incentive compatible to choose the efficient effort level. Note the difference between [5] and trust experiments: in TWP only a return of 4 is incentive compatible, which is not efficient, whereas in [5] social efficient effort is incentive compatible. In contrast to HRTI, wages, effort and trustor earnings are lower when formal incentives are declined than when unavailable.

3. Method

In order to further examine the robustness of HRTI we administer a gift-exchange game with the same parameters as [5], but otherwise more closely following the protocols of the aforementioned trust games. Specifically, socially efficient levels of effort are not incentive compatible and the principal makes an effort level request analogous to the “return” requests in [2] and [3].

The principal chooses to pay an agent a wage $w$, where $w \in \{20, 21, \ldots, 120\}$. After learning the wage, the agent selects an effort, $e$, where $e \in \{0.1, 0.2, \ldots, 1\}$. The principal’s earnings are $e(120 - w)$. The agent’s earnings are $w - c(e)$, where $c(e)$ is the cost of effort to the agent. The relationship between effort and costs is shown below:

<table>
<thead>
<tr>
<th>$e$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c(e)$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>15</td>
<td>18</td>
</tr>
</tbody>
</table>

Two treatments were administered. In both treatments the principal requests a level of effort from the agent, $e'$. In the Gift-Exchange treatment the game is played as described above, and $e'$ is not payoff relevant. The equilibrium is for the principal to set $w = 20$ and for the agent to set $e = 0.1$.

In the Choice treatment, the principal can choose to either play a gift-exchange game as described above, or a gift-exchange game with punishment. The punishment is in the form of a fine of 6 paid by the agent to the experimenter if $e < e'$. The principal also surrenders 2 to the experimenter if the fine is administered. If the punishment is chosen, $e'$ becomes payoff relevant. Assuming individuals only care about their own payoff, the equilibrium is for the principal to choose $e' = 0.5$ and $w = 20$ and the agent’s best response is to choose $e = 0.5$.

The Gift-Exchange treatment consisted of two sessions with 40 total participants. The Choice treatment consisted of three sessions with 46 total participants. All participants were undergraduate volunteers from The Ohio State University. Agents and principals are re-matched after each period and ten periods were administered. Experiments were computerized, using z-Tree [7].

Our focus is on a comparison of Gift-Exchange treatment (hereafter EX-GE) and endogenous gift-exchange (hereafter EN-GE) where the principal made a choice of gift-exchange in the Choice treatment. Greater reciprocity in EN-GE would be further evidence of the HRTI, while greater reciprocity in EX-GE would be a further contradiction, analogous to the result in [5].

---

3No demographic data were collected.
4. Results

Result 1: *Agents’ effort and principals’ payoffs are lower when principals choose not to use the punishment (EN-GE) than when no punishment option is available (EX-GE).*

Result 1 is opposite HRTI. Summary statistics are found in Table 1. Average effort in EN-GE is 0.25, while average effort in EX-GE is 0.43 (Wilcoxon-Mann-Whitney test, \( p < 0.01 \)). Additionally the principals’ earnings in EN-GE are on average 15 and are significantly lower than that in EX-GE, with average earnings of 25 (Wilcoxon-Mann-Whitney test, \( p < 0.01 \)).

**Figure 1** presents agents’ effort partitioned by wage offers. As might be expected, the difference between effort provision in EX-GE versus EN-GE is mainly found for high wage offers. For low wage offers there is no expectation of reciprocity in either condition.

Result 2: *The difference between agents’ effort when principals choose not to use the punishment (EN-GE) versus when there is no punishment option (EX-GE) mainly occurs in the later periods.*

**Figure 2** presents agents’ effort over time. The figure reveals that in the first four periods agents’ effort when principals refrain from using the punishment option is not significantly different from when the punishment option is not available. Moreover, in the latter six periods agents’ effort is significantly lower

![Figure 1. Agent’s effort across treatments.](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage</td>
<td>54 (18)</td>
<td>50 (17)</td>
<td>57 (18)</td>
</tr>
<tr>
<td>Agent’s Effort</td>
<td>0.43 (0.25)</td>
<td>0.37 (0.22)</td>
<td>0.25 (0.23)</td>
</tr>
<tr>
<td>Principal’s Payoffs</td>
<td>25 (13)</td>
<td>24 (16)</td>
<td>15 (14)</td>
</tr>
<tr>
<td>Agent’s Payoffs</td>
<td>49 (15)</td>
<td>43 (16)</td>
<td>55 (18)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>200</td>
<td>182</td>
<td>48</td>
</tr>
</tbody>
</table>
Figure 2. Agent’s effort by periods.

in EN-GE than EX-GE (Wilcoxon-Mann-Whitney test, $p = 0.03$ in period 5, $p = 0.03$ in period 6, $p = 0.01$ in period 7, $p = 0.03$ in period 8, $p = 0.02$ in period 9 and $p = 0.12$ in period 10). One interpretation is that as agents have experience with punishment they view the relationship with the principal more antagonistically, and withdraw effort even when the principal does not punish.

5. Further Research

The hidden returns of incentives have been found in two studies comprising three different experiments using the trust game, but have been contradicted in two experiments using a gift-exchange game. The most salient difference between the two is that a gift-exchange game is framed as an employment relationship while a trust game is not. It is possible that the introduction of penalties, even if not used, is suggestive of an adversarial relationship in gift-exchange that is not present in a trust game. Cardinaels & Yin [8], using a bargaining game that is also framed as an employment relationship but is somewhat different than trust and gift-exchange games, hypothesize that when agents see a punishment regime being used it signals that principals expect that agents might cheat and hence makes cheating appear to be “the norm”.

Two future experiments might be helpful. The first is bringing the trust and gift-exchange games into a single experimental design. This would control for potential nuisance factors such as differences in the participant pool. The second is to expose participants to the idea of a punishment treatment, without actually giving principals the choice to use it, and compare that to a choice treatment. This would allow a discernment between agents’ reframing the game after experiencing a punishment versus agent’s reframing the game after only becoming aware of a potential punishment regime but not actually experiencing it.

Finally, an idea that would further explore the robustness of the hidden returns to incentives is to use the trust game as in [2] and [3], but have the participants play multiple periods in a “strangers” matching protocol. It is possible that once the idea of punishment is made more salient through repetition, reci-
procation will decline even if the punishment option is not used.

**Acknowledgements**

Thanks to Anil Arya, Katie Coffman, Jeremy Douthit, John Kagel, Doug Schroeder and Di Yang for their comments and suggestions. This research was sponsored by The Ohio State University Department of Accounting and MIS.

**References**


