Is No News Bad News?
A Hostage Trust Game with Incomplete Information
and Fairness Considerations of the Trustee

Thomas Gautschi
University of Mannheim

New Developments in Signaling and Game Theory
Monte Verità, October 14–19, 2012
The Trust Game

- The Trust Game is a formal description of a specific class of cooperation problems based on dyadic and non-cooperative exchange.
- In a simple one-shot Trust Game, player 1 (trustor) moves first and decides between placing or withholding trust.
- The game ends if the trustor withholds trust.
- If trust is placed, player 2 (trustee) decides between honoring or abusing placed trust. The game ends.
- Given the actors preferences, the game has a single subgame perfect equilibrium such that the trustor withholds trust while the trustee will abuse trust if placed.
- The game is a formal description of a trust situation where the trustee has an incentive to abuse trust while the trustor looses when placed trust is abused.
Escaping from the Trap of Mutual Defection

- Both players would profit from mutual cooperation, compared to the equilibrium outcome of mutual defection.

- Several ways to escape from this Pareto-inefficient solution are conceivable:
  - infinitely repeated games (i.e., dyadic embeddedness)
  - network embeddedness and exploiting reputation effects
  - credible commitments (i.e., hostages)

- A hostage is a self-binding commitment to mitigate (or completely remove) the incentive to defect.

- Hostages ensure mutual cooperation (binding hostages) or facilitate cooperation (non-binding hostages).

- A hostage can thus be seen as a signaling device.
Signals of Trust or Signals of Distrust?

What happens if the hostage is not posted?
Signals of Trust or Signals of Distrust?

Trust Game

\[ T_j > R_j > P_j \text{ and } R_i > P_i > S_i. \]
Signals of Trust or Signals of Distrust?

Trust Game

\[ T_j > R_j > P_j \] and \[ R_i > P_i > S_i, \]

Hostage Trust Game

\[ H_i \geq 0 \] and \[ H_j \geq T_j - R_j. \]
Signals of Trust or Signals of Distrust?

Trust Game

$T_j > R_j > P_j$ and $R_i > P_i > S_i$.

Hostage Trust Game

$H_i \geq 0$ and $H_j \geq T_j - R_j$.
Empirical Evidence on the TG and HTG

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>p–Value</th>
<th>Marginal effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subgame HTG</td>
<td>-0.48</td>
<td>0.00</td>
<td>-0.10</td>
</tr>
<tr>
<td>Risk</td>
<td>-2.08</td>
<td>0.00</td>
<td>-0.44</td>
</tr>
<tr>
<td>Temptation</td>
<td>-0.02</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.62</td>
<td>0.33</td>
<td></td>
</tr>
</tbody>
</table>

$N = 1883$, Pseudo $R^2 = .29$, $L = -668.3$

† Only relevant predictors of Table 6.5 (Snijders 1996: 158) are reported.

Probit regression with robust standard errors that the trustor chooses to trust in the ‘no hostage’ subgame of the Hostage Trust Game.

Value of hostages: $H_i = 0$, $H_j > 0$ in Groningen, and $H_i = H_j > 0$ in Amsterdam.

$$\text{Risk} = \frac{(P_1 - S_1)}{(R_1 - S_1)}, \text{Temptation} = \frac{(T_2 - R_2)}{(T_2 - S_1)}$$

- Subjects in the role of the trustor behave differently in a one-shot Trust Game and the respective subgame of a Hostage Trust Game (reached after the trustee denies to post the hostage).
- Experimental subjects playing as the trustor were less likely to place trust in a Hostage Trust Game with no hostage previously posted, compared to their decision in the Trust Game.
Puzzeling or not Puzzeling Evidence?

- The normative answer:
  - The subgame reached in the HTG is identical to the TG.
  - Nothing can be learned from observing no hostage being posted.
  - Observed behavior is inconsistent under complete and perfect information.

- The behavioral answer:
  - A hostage obviously is a signaling device – whether posted or not!
    ★ A trustworthy trustee would post a hostage since he will not loose it.
    ★ A untrustworthy trustee will not post a hostage since he will not run the risk of loosing it.
  - Assuming to play the subset of “bad” trustees seems to decrease the probability to trust.

- Can this experimentally observed behavior form an equilibrium?
Situation where one actor is unsure about the incentives of the other actor, we thus use a Trust Game under incomplete information.

Assume the following:

- The trustee is of the good type with probability $0 < \pi^{TG} < 1$.
- The trustee of the bad type with probability $1 - \pi^{TG}$.
- The type of trustees can be distinguished via a minor refinement of their reward-payoff $R_2$.
- The good trustee’s “fairness payoff” is $\Delta_G \geq T_2 - R_2$ which compensates him for cooperation ($\Delta_G + R_2 > T_2$) or makes him indifferent ($\Delta_G + R_2 = T_2$).
- The bad trustee’s “fairness payoff” is $\Delta_B < T_2 - R_2$ which does not completely compensate him for cooperation.
The Trust Game with Incomplete Information II

- The trustor trusts the trustee iff $\pi^{TG} R_1 + (1 - \pi^{TG}) S_1 > P_1$.
- Trust is placed if the prior exceeds $\pi^{TG} > \frac{P_1 - S_1}{R_1 - S_1} =: \pi^*$. 

\[ \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}, \begin{pmatrix} S_1 \\ T_2 \end{pmatrix}, \begin{pmatrix} R_1 \\ R_2 + \Delta_G \end{pmatrix}, \begin{pmatrix} P_1 \\ T_2 \end{pmatrix}, \begin{pmatrix} S_1 \\ R_2 + \Delta_B \end{pmatrix} \]
Distinguish three different types of trustees, the good type with probability $0 < \pi_G < 1$, the mediocre type with probability $0 < \pi_M < 1$, and the bad type with probability $\pi_B = 1 - \pi_G - \pi_M$.

Again assume that the trustee receives an additional payoff $\Delta$ due to fairness considerations if he does not abuse trust if placed.

The trustor can observe whether or not a hostage has been posted by a trustee before he moves.

The trustee loses his hostage if posted, trust is then placed but subsequently abused ($T_2^- = T_2 - H$).

Let the trustor’s payoff then be

$$S_1^+ = \begin{cases} S_1 & \text{if the hostage goes to a third party} \\ S_1 + H & \text{if the hostage goes to the trustor} \end{cases}$$
The Hostage Trust Game with Incomplete Information II

Payoffs: \( R_2^\Delta \equiv R_2 + \Delta \); \( T_2^- \equiv T_2 - H \); \( H^- = \) ‘not posting a hostage’, \( H^+ = \) ‘posting a hostage; \( nt = \) no trust, \( pt = \) place trust, \( at = \) abuse trust, \( ht = \) honor trust.
The Hostage Trust Game with Incomplete Information III

- Explaining the empirical findings asks to show that there exists an equilibrium in the Hostage Trust Game with incomplete information in which the trustor’s assessment about facing a good trustee falls short of $\pi^*$. 

- It need be shown that $\pi_G < \pi^* < \pi^{TG}$ can exist. 

- Due to the asymmetric information, perfect Bayesian equilibria are used. 

- In case of out-of-equilibrium behavior we propose behavior that does not contradict Bayes’s Rule. 

- More precisely, we assume that the trustor’s priors do not change after observing out-of-equilibrium behavior by a trustee.
Define the three types of trustees:

- The good trustee: He always honors trust, therefore
  \[ \pi_G = \Pr(\Delta > T_2 - R_2). \]

- The mediocre trustee: He honors trust after posting a hostage but abuses trust after refraining from posting the hostage, therefore
  \[ \pi_M = \Pr(T_2 - R_2 - H < \Delta < T_2 - R_2). \]

- The bad trustee: He always abuses trust, therefore
  \[ \pi_B = \Pr(\Delta < T_2 - R_2 - H). \]

Define the following probabilities and sets:

- \( q^+ = \Pr(\text{place trust}|H^+) \), the probability that the trustor places trust after observing a trustee posting a hostage.

- \( q^- = \Pr(\text{place trust}|H^-) \), the probability that the trustor places trust after observing a trustee not posting a hostage.

- \( J_G \), the set of good trustees.
Equilibrium Analyses II

- $J_M$, the set of mediocre trustees.
- $J_B$, the set of bad trustees.
- $\mathcal{H}^+$, the set of hostage posting trustees.
- $\mathcal{H}^-$, the set of none hostage posting trustees.
- $\mathcal{H}^{(0,1)}$, the set of trustees mixing with a probability in the open interval $(0, 1)$ over whether or not to post a hostage.
- $p_G = \Pr(H^+|\text{good trustee})$, the probability that a good trustee posts a hostage.
- $p_M = \Pr(H^+|\text{mediocre trustee})$, the probability that a mediocre trustee posts a hostage.
- $p_B = \Pr(H^+|\text{bad trustee})$, the probability that a bad trustee posts a hostage.
Equilibrium Analyses III

Equilibrium I (Pooling Equilibrium)

Suppose that \((J_G \cup J_M \cup J_B) \subset H^+, H^- = \emptyset,\) and \(H^{(0,1)} = \emptyset;\) then and only then an equilibrium of the following type exists:

All types of trustees post a hostage with probability one while the trustor subsequently places trust if and only if \(\pi_G < \frac{P_1 - S_1}{R_1 - S_1}\) and \(\pi_G + \pi_M > \frac{P_1 - S_1^+}{R_1 - S_1^+}\) are simultaneously fulfilled. The trustees consequently play their equilibrium strategies.

- Trust is placed if the assessment about the pool of good trustees falls short of \(RISK,\) and the pool of good and mediocre type trustees exceeds \(RISK^+;\)
- Equilibrium behavior of the trustor asks for hostage posting with probability one by all types of trustees.
- This does not provide any signal about the type of a trustee.
- It nevertheless makes sure that a good and a mediocre type trustee will honor placed trust.
- To trust in the first place pays off for the trustor.
Equilibrium Analyses IV

**Equilibrium II**

Suppose that \((J_M \cup J_B) \subset \mathcal{H}^-, J_G \subset \mathcal{H}^{(0,1)}\), and \(\mathcal{H}^+ = \emptyset\); then and only then an equilibrium of the following type exists:

A mediocre and a bad type trustee refrain from posting a hostage with probability one while a good type trustee posts a hostage with probability \(0 < p_G < 1 - \frac{1 - \pi_G}{\pi_G\left(\frac{1}{R_1-P_1}\right)}\). The trustor subsequently places trust if and only if \(\pi_G > \frac{P_1-S_1}{R_1-S_1}\). The trustees consequently play their equilibrium strategies.

- Good trustees are mixing over whether to post the hostage, but \(p_G\) can be zero.
- A mediocre and a bad type trustee refrain from posting a hostage.
- Whenever the trustor sees a hostage being posted, she can be sure to face a good type trustee. No hostage reveals no clear information.
- By not observing a hostage, the trustor faces the same decision problem as in the Trust Game with incomplete information.
- The condition for cooperation in the HTG with incomplete information is identical to the one in the TG with incomplete information.
Equilibrium Analyses V

Equilibrium IIIa

Suppose that \((J_G \cup J_M \cup J_B) \subset \mathcal{H}^{(0,1)}\), \(\mathcal{H}^+ = \emptyset\), and \(\mathcal{H}^- = \emptyset\); then and only then an equilibrium of the following type exists:

All types of trustees provide a hostage with probability \(0 < p_G = p_M = p_B < 1\) and the trustor withholds trust if and only if \(\pi_G + \pi_M < \frac{P_1 - S_1^+}{R_1 - S_1^+}\). The trustees consequently play their equilibrium strategies.

- Assume \(p_G = p_M = p_B\) tends to, but do not reach zero.
- As long as \(\pi_G + \pi_M < \text{RISK}^+\), the trustor is not willing to place trust.
- Since \(\pi_G + \pi_M\) together need to be smaller than \(\text{RISK}^+\) but \(\text{RISK}^+ < \text{RISK}\), we have that \(\pi_G < \pi^* < \pi^{TG}\) can be fulfilled.
Equilibirum Analyses VI

Equilibirum IIIb

Suppose that \((J_G \cup J_M \cup J_B) \subset \mathcal{H}^{(0,1)}, \mathcal{H}^+ = \emptyset, \) and \(\mathcal{H}^- = \emptyset;\) then and only then an equilibrium of the following type exists:

All types of trustees provide a hostage with probability \(0 < p_G < 1, 0 < p_M < 1,\) and \(0 < p_B < 1\) for all \(p_G \neq p_M \neq p_B,\) the trustor then withholds trust if and only if

\[
\frac{\pi_G p_G + \pi_M p_M}{\pi_G p_G + \pi_M p_M + \pi_B p_B} < \frac{P_1 - S_1}{R_1 - S_1} \quad \text{and} \quad \frac{\pi_G (1-p_G)}{\pi_G (1-p_G) + \pi_M (1-p_M) + \pi_B (1-p_B)} < \frac{P_1 - S_1}{R_1 - S_1}
\]

are simultaneously fulfilled. The trustees consequently play their equilibrium strategies.

- Equilibrium IIIb is a somewhat more complex version of Equilibirum IIIa.
- Assume \(p_G \neq p_M \neq p_B\) tends to, but do not reach zero.
- Parameter values can be found such that \(\pi_G < \pi^* < \pi^{TG}\) is fulfilled. For example:
  
  - \(T_2 = 75, R_1 = R_2 = 50, P_1 = P_2 = 40, S_1 = 15\) and \(H = 5.\)
  - \(p_G = .01, p_M = .02,\) and \(p_B = .03.\)
  - Therefore, \(\text{RISK} = 0.7143\) and \(\text{RISK}^+ = 0.6667.\)
  - Prior beliefs are \(\pi_G = 0.5, \pi_M = 0.2,\) and \(\pi_B = 0.3.\)
Discussion I

- Hostages are posted in equilibrium. However, none of these equilibria are separating.
- Normatively, hostages can thus not be used as a signaling device in this simple game.
- It always pays off for the bad trustee to mimic a good or mediocre trustee (note, the hostage was not binding!).
- No equilibrium exists which fully supports the empirical findings.
- If no hostage has been posted, $\pi_G < \pi^* < \pi^{TG}$ is never fulfilled in equilibrium.
- However, for $p_G = p_M = p_B \to 0$ and $p_G \neq p_M \neq p_B \to 0$ we find support for the empirical findings.
Discussion II

- What should we conclude? Is no news bad news? Yes!
- Not posting the hostage seems to bear some information (about the trustee).
- The analyses suggest the following:
  - Not using a potential signaling device serves as an (unintended) signal.
  - Send clear signals about your own cooperative intentions and do not rely upon diffuse signals.
  - Diffuse signals or no signals may be misunderstood by other players.
- The analysis also showed that an equilibrium exists in which all types of trustees place a hostage. A simple, even though clear signal may thus not be sufficient.
Arthur Conan Doyle. 1892. *Silver Blaze*

Inspector Gregory: “Is there any other point to which you would wish to draw my attention?”

Holmes: “To the curious incident of the dog in the night-time.”

“The dog did nothing in the night time”

“That was the curious incident,” remarked Sherlock Holmes.