ANALYSIS ON TRUST GAME BY RECIPROCAL AGENTS

Hidetoshi OKAYASU

Department of Value and Decision Science
Graduate School of Decision Science and Technology
Tokyo Institute of Technology
2-12-1 O-okayama Meguro-ku, Tokyo 152-8552, JAPAN

Telephone number: +81-3-5734-3768
Facsimile number: +81-3-5734-3768
E-mail address: okayasu@valdes.titech.ac.jp

ABSTRACT

In this paper, the author proposes a game-theoretical model of trust among reciprocal agents. Our model, a trust game, is a non-cooperative game in extensive form. By considering about this game, we can define clearly the concept of trust behavior in general games in extensive form. But just using ordinary equilibrium concept (e.g. subgame perfect equilibrium), we cannot explain the trust behavior in some situations. This result contradicts with some observations in real world. So, we have to adopt another solution concept, sequential reciprocity equilibrium (SRE), which is suggested by Dufwenwerg. Adopting this SRE concept, we analyze repeated trust game (RTG). As a result of analysis, I find the condition of reciprocity to trust others, and reciprocal agents can get higher payoff than non-reciprocal agents when the length of game is enough long.

Keywords: Game Theory; Trust; Reciprocity; Social Network

1. INTRODUCTION

The purpose of this research is to develop a game theoretical model of trust among reciprocal agents. Through analysis on this model, we find conditions of establishing trust relation. This study can be employed for further application: for example, mechanism designs for creating highly-trust social network, evaluation for strength of trust relation between service provider and consumer, etc.

Trust has been studied in many areas, sociology, economics, politics, anthropology, psychology and so on. These studies can divides into three approaches. (Sato, 1999) First one is functionalistic approach, which views trust as a function in social systems, as represented by Luhmann. He insisted that trust is a mechanism for reduction of complexity. Secondly, psychological approach supposes trust as a personal characteristic. On this approach, researchers compare the tendency of trust among countries, cultures etc. by experiments and surveys. The last one is rationalistic approach, which develops trust model and analyzes it by economic method, e.g. game theory, expected-utility theory. The representative study from this approach is Coleman's theory based on rational choice theory. He established a trust model, and explained "place trust in others" action by utility-maximize action. (Coleman, 1990) We adopt this rationalistic approach.

The structure of this paper is as follows: in Section2, a basic model of trust, we call "trust game", will be suggested. For analyzing this model, we introduce a special equilibrium concept, sequential reciprocal equilibrium, in Section 3, and we adopt this concept for trust game and repeated trust game in Section 4. Finally, Section 5 is devoted to mention concluding remarks and further study.

2. MODEL: TRUST GAME

We start with Coleman's trust model, which expressed trust as risky decision making. It is extensive form game illustrated in Figure 2.1. This game is played by two players, Truster and Trustee. Just considered Truster's utility, Trustee make a probabilistic choice between 'Trustworthy' and 'Not Trustworthy'. Supposing that probability of choosing 'Trustworthy' is p, potential gain and loss of Truster when Trustee chooses 'Trustworthy' and 'Not Trustworthy' is G and L respectively. In this trust model, we can obtain a condition below about Truster's rational decision making immediately.

- Truster place trust if and only if $p > \dfrac{L}{G+L}$
- Truster don't place trust if and only if $p < \dfrac{L}{G+L}$

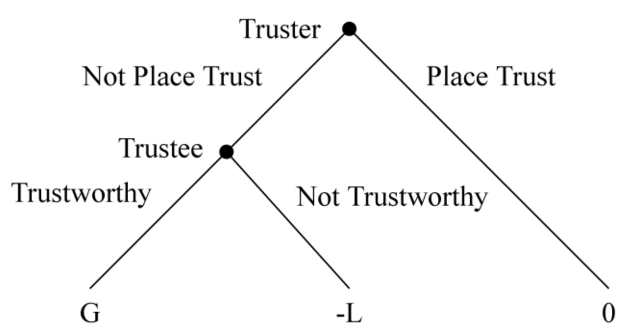

Figure 2.1 Coleman's Trust Model

Generally speaking, probability p is not known to Truster, so a mechanism of shaping expectation on Trustee's action is interesting. But Coleman's model cannot explain how Truster's expectation shapes, because it treats Trustee's action probabilistically.

To solve this problem, we develop Coleman's model into trust game in Figure 2.2 by introducing Trustee's utility. The following is subgame perfect equilibrium of this game.

- If $0 < b < 1$, subgame perfect equilibrium is (T, r)
- If $1 < b$, subgame perfect equilibrium is (N, e)

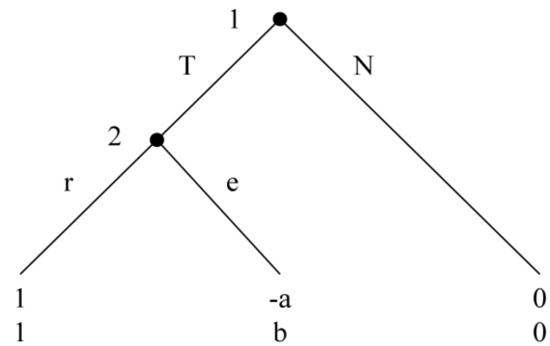

Figure 2.2 Trust Game

Even though $1 < b$, we can observe Truster(Player 1) places trust on Trustee(Player 2) and Trustee rewards Truster's trust in real life. To explain this phenomenon, we introduce other-regarding preference. Specifically, we adopt other equilibrium concept, which is considered with agents' reciprocity in next section.

3. EQUILIBRIUM CONCEPT

This section gives the definition of sequential reciprocal equilibrium (SRE). (Cox, 2004) define that (positive) reciprocity is a motivation to repay generous or helpful actions of another by adopting actions that are generous or helpful to the other person, which distinct from the unconditional kindness motivated by altruism. But how generous or helpful Player 1's action is for Player 2 depends on the 2's belief on 1's action as well as 1's action itself. So utility function is not only determined by pair of strategies, but also pair of players' beliefs. It is the essence of psychological game (Gianakoplos et al., 1989), on which SRE concept is based. (Dufwenberg et al., 2004) defined SRE as below.

Let $N = \{1,2,...n\}$ be the set of players. Let H be the set of choice profiles, or histories. Let A_i be the set of behavior strategies of $i \in N$. Define $A = \prod_{i \in N} A_i$. Let $\pi_i : A \to \mathbf{R}$ be player i's material payoff function. Let $B_{ij} = A_j$ be the set of possible beliefs of player i's about the strategy of player j. Let $C_{ijk} = B_{jk} = A_k$ be the set of possible beliefs of player i's about the belief of player j about the strategy of player k .With $a_i \in A, h \in H$, let $a_i(h)$ be the update strategy that prescribes the same choices as a_i, except for the choices that define history h which are made with probability 1. For $b_{ij} \in B_{ij}, c_{ijk} \in C_{ijk}$, define update beliefs $b_{ij}(h)$, $c_{ijk}(h)$ in same fashion to update strategies.

Here we move on to kindness and utility function, which appeared in definition of SRE.

Definition 3.1 (kindness)

The kindness of player i to another player $j \neq i$ at history $h \in H$ is given by the function $\kappa_{ij} : A_i \times \prod_{j \neq i} B_{ij} \to \mathbf{R}$ defined by

$$\kappa_{ij}(a_i(h),(b_{ij}(h))_{j\neq i}) = \pi_j(a_i(h),(b_{ij}(h))_{j\neq i}) - \pi_j^{e_i}((b_{ij}(h))_{j\neq i})$$

$\pi_j^{e_i}((b_{ij}(h))_{j\neq i}$ is the equitable payoff for player j, defined by

$$\pi_j^{e_i}((b_{ij}(h))_{j\neq i} = \frac{1}{2}[\max_{a_i \in A_i} \pi_j(a_i(h),(b_{ij}(h))_{j\neq i}) + \min_{a_i \in A_i} \pi_j(a_i(h),(b_{ij}(h))_{j\neq i})]$$

Definition 3.2 (belief about kindness)

Player i's beliefs about how kind player $j \neq i$ is to i at history $h \in H$ is given the function $\lambda_{iji} : B_{ij} \times \prod_{k \neq j} C_{ijk} \to \mathbf{R}$ defined by

$$\lambda_{iji}(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) = \pi_i(b_{ij}(h),(c_{ijk}(h))_{k\neq j}) - \pi_i^{e_j}((c_{ijk}(h))_{k\neq j})$$

Definition 3.3 (utility function)

Player i's utility at history $h \in H$ is a function $U_i : A_i \times \prod_{j\neq i}(B_{ij} \times \prod_{k\neq j} C_{ijk}) \to \mathbf{R}$ defined by

$$\begin{aligned}&U_i(a_i(h),(b_{ij}(h),(c_{ijk}(h))_{k\neq j})_{j\neq i})\\ &= \pi_i(a_i(h),(b_{ij}(h))_{j\neq i}) + \sum_{j\in N\backslash i}(Y_{ij}\kappa_{ij}(a_i(h),(b_{ij}(h))_{j\neq i}) \cdot \lambda_{iji}(b_{ij}(h),(c_{ijk}(h))_{k\neq j}))\end{aligned}$$

Especially in case of $N = \{1,2\}$, utility functions are following,

$U_1(a_1(h),b_{12}(h),c_{121}(h)) = \pi_1(a_1(h),b_{12}(h)) + Y_{12} \cdot \kappa_{12}(a_1(h),b_{12}(h)) \cdot \lambda_{121}(b_{12}(h),c_{121}(h))$

$U_2(a_2(h),b_{21}(h),c_{212}(h)) = \pi_2(a_2(h),b_{21}(h)) + Y_{21} \cdot \kappa_{21}(a_2(h),b_{21}(h)) \cdot \lambda_{212}(b_{21}(h),c_{212}(h))$

Definition 3.4 (SRE)

The profile $a^* = (a_i^*)_{i\in N}$ is sequential reciprocity equilibrium (SRE) if for all $i \in N$ and for each history $h \in H$ it holds that

- $a^* \in \arg\max_{a_i \in A_i(h,a^+)} U_i(a_i(h),(b_{ij}(h),(c_{ijk}(h))_{k\neq j})_{j\neq i})$

  $A_i(h,a) \subseteq A_i$ is the set of strategies, that prescribe the same choices as the strategy $a_i(h)$ for all histories other than h.

- $b_{ij} = a_j^*$ for all $j \neq i$

- $b_{ijk} = a_k^*$ for all $k \neq j$

As above, this definition consists of three conditions, 'utility maximization', 'consistency with first-order belief' and 'consistency with second-order belief'. If $Y_{ij} = 0$ for any i, j, each player is motivated by material payoff. In this case, SRE is same as subgame perfect equilibrium.

4. ANALYSIS

Trust Game

Then, we complete the preparation of equilibrium concept. Using SRE, we begin to analysis on trust game in this section. We simplify notation Y_1 and Y_2 instead of Y_{12} and Y_{21}, since we analyze only two-player games.

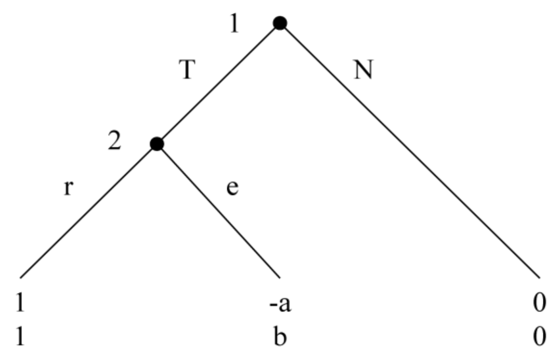

Figure 4.1 Trust Game

Proposition 4.1

For any SRE a^*, $a_1^* = T$ and $Y_2 > \dfrac{2(b-1)}{a+1} \Rightarrow a_2^* = r$

(Proof)

Focused on 2's decision making. 1's equitable payoff is $\pi_1^{e_2} = \dfrac{1-a}{2}$. Then 2's kindness to 1 is $\kappa_{21}(r) = 1 - \pi_1^{e_2} = \dfrac{1+a}{2}$, $\kappa_{21}(e) = -a - \pi_1^{e_2} = -\dfrac{1+a}{2}$. Supposing that 2's belief on 1's belief about 2's strategy is $c_{212} = p''r + (1-p'')e$. Then 2's equitable payoff is $\pi_2^{e_1} = \dfrac{1}{2}((p''\cdot 1 + (1-p'')b) + 0)$. Therefore 1's belief on 2's kindness to 1 is $\lambda_{121}(T, p''r + (1-p'')e) = \dfrac{1-b}{2}p'' + \dfrac{b}{2}$. Hence 2's utility as follows

- 2 chooses strategy 'r', $U_2' = 1 + Y_2(\dfrac{1+a}{2})(\dfrac{1-b}{2}p'' + \dfrac{b}{2})$
- 2 chooses strategy 'e', $U_2'' = b + Y_2(-\dfrac{1+a}{2})(\dfrac{1-b}{2}p'' + \dfrac{b}{2})$

$$U_2' > U_2'' \Leftrightarrow 1-b+2Y_2(\frac{1+a}{2})(\frac{1-b}{2}p''+\frac{b}{2}) > 0$$

By consistency condition $p''=1$, we obtain $Y_2 > \frac{2(b-1)}{a+1}$.

Next proposition can be proved similarly fashion to Proposition 4.1.

Proposition 4.2

For any SRE a^*, $Y_2 > \frac{2(b-1)}{a+1} \Rightarrow a_1^* = T$

From proposition 4.1 and 4.2, following theorem derived immediately.

Theorem 4.3

If $Y_2 > \frac{2(b-1)}{a+1}$, unique SRE is $a^* = (a_1^*, a_2^*) = (T, r)$

Repeated Trust Game(RTG)
Next, we try to analyze repeated trust game(RTG), which is illustrated Figure 4.2. Nodes with Odd number are 1's move, the others are 2's move. Supposing that total number of move M is even. On each move, player has two alternatives T_i and N_i, corresponding T and N in trust game respectively. If 1 chooses T_k and 2 chooses T_{k+1}, both players' payoff increase 1. On the other hand, though 1 chooses T_k, 2 chooses N_{k+1}, 1's payoff decrease $a(0<a<1)$ and 2's payoff increase $b(1<a<2)$ and the game is terminated. If 1 chooses N_k, the game is over without payoff change. Think along the similar analysis on trust game, we try to find conditions for trust actions included in SRE.

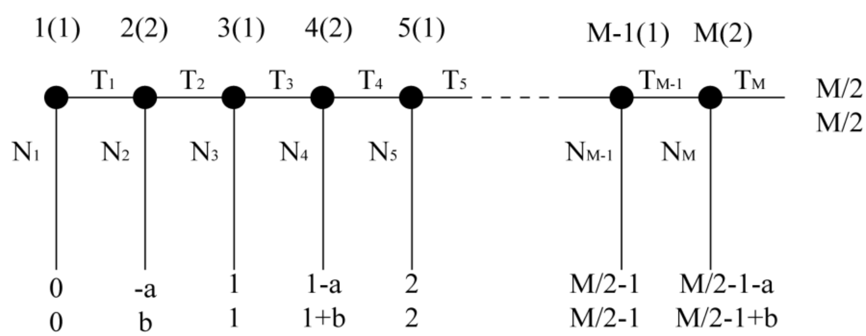

Figure 4.2 Repeated Trust Game

Proposition 4.4

For any SRE a^*, which reaches node M, $Y_2 > \dfrac{4(b-1)}{M(a+1)} \Rightarrow a_2^* = (*,\cdots,*,T_M)$

(Proof)

Focused on 2's decision making on node M. 1's equitable payoff is $\pi_1^{e_2} = \dfrac{M}{4} - \dfrac{a}{2}$. Then 2's kindness to 1 is $\kappa_{21}(T_{2,} \cdots, T_{M-2}, T_M) = -\dfrac{M}{4} + \dfrac{a}{2}$, $\kappa_{21}(T_{2,} \cdots, T_{M-2}, N_M) = \dfrac{M}{4} - \dfrac{a}{2} - 1$.

Supposing that 2's belief on 1's belief about 2's strategy is $c_{212} = (T_{2,} \cdots, T_{M-2}, p''T_M + (1-p'')N_M)$. Then 2's equitable payoff is $\pi_2^{e_1} = \dfrac{M}{4} + \dfrac{1-p''}{2}(-1+b)$. Therefore 1's belief on 2's kindness to 1 is $\lambda_{121}(T_{2,} \cdots, T_{M-2}, p''T_M + (1-p'')N_M)$. Hence 2's utility as follows

- 2 chooses strategy 'T_M', $U_2' = \dfrac{M}{2} + Y_2(\dfrac{M}{4} + \dfrac{1-p''}{2}(-1+b))(\dfrac{M}{4} + \dfrac{a}{2})$

- 2 chooses strategy 'N_M', $U_2'' = \dfrac{M}{2} + Y_2(\dfrac{M}{4} + \dfrac{1-p''}{2}(-1+b))(\dfrac{M}{4} - \dfrac{a}{2} - 1)$

$$U_2' > U_2'' \Leftrightarrow 1 - b + Y_2(\frac{M}{4} + \frac{1-p''}{2}(-1+b)(a+1)) > 0$$

By consistency condition $p''=1$, we obtain $Y_2 > \dfrac{4(b-1)}{M(a+1)}$.

Proposition 4.5

For any SRE a^*, if there exists k such that players choose $T_{k'}$ at any node k' ($k < k' \le M$), players choose T_k at node k.

(Proof)

Supposing that k is odd.

- 1 chooses strategy 'T_k', $U_2' = \dfrac{M}{2} + Y_1 \cdot \kappa_{12}(*,\cdots,*,T_k,T_{k+2},\cdots,T_{M-1}) \cdot \lambda_{121}(\cdot)$

- 1 chooses strategy 'N_k', $U_2'' = \dfrac{k-1}{2} + Y_1 \cdot \kappa_{12}(*,\cdots,*,N_k,T_{k+2},\cdots,T_{M-1}) \cdot \lambda_{121}(\cdot)$

Because of $b_{12} = (*,\cdots,*,T_{k+1},\cdots,T_M)$,

$\kappa_{12}(*,\cdots,*,T_k,T_{k+2},\cdots,T_{M-1}) - \kappa_{12}(*,\cdots,*,N_k,T_{k+2},\cdots,T_{M-1}) = \dfrac{M}{2} - \dfrac{k-1}{2} > 0$. Hence 2's utility as follows. $U_2' - U_2'' = (\frac{M}{2} - \frac{k-1}{2}) + Y_1(\frac{M}{2} - \frac{k-1}{2}) \cdot \lambda_{121}(\cdot) > 0$. Therefore player choose T_k at node k. We can prove similarly when k is even.

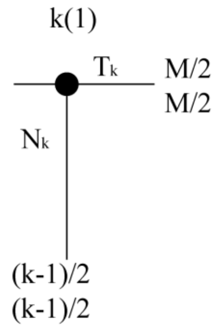

Figure 4.3 Decision Making on Node k (k: odd)

Combined these propositions, following theorem can be proven inductively.

Theorem 4.6

If $Y_2 > \frac{4(b-1)}{M(a+1)}$, unique SRE is $a^* = (a_1^*, a_2^*) = ((T_{1,} \cdots, T_{M-1}), (T_{2,} \cdots, T_M))$

5. CONCLUSIONS

This paper gave a game theoretical model of trust. Introducing SRE concept, 'place trust' and 'reward' action could be expressed as a solution of extensive form game. Through analysis on repeated trust game, we found the condition for building trust among reciprocal agents. The main outcome is Theorem 4.6; it shows that the longer RTG, the easier to build trust relation between agents. The next step of this research is extending the number of players. Though SRE concept could be adopted for game played by 3+ players, the analysis may be too complicated to solve algebraically. So, agent-based simulation will be powerful tool for analysis.

REFERENCES

Coleman, James. (1990). *Foundation of Social Theory.*, Harvard University Press.

Cox, J. C. (2004). How to identify trust and reciprocity, *Games and Economic Behavior.*, 46: 260-281

Dufwenberg, M., and Kirchsteiger, G. (2004). A theory of sequential reciprocity, *Games and Economic Behavior.*, 47: 268-298

Geanakoplos, J., Pearce, D. (1989). Psychological games and sequential rationality, *Games and Economic Behavior.*, 1:60-79

Luhmann, Niklas. (1973). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexitat.*, Ferdinand Enke Verlag.

Sato, Yoshimichi. (1999). Trust and Communication, *Sociological Theory and Methods.*, 13: 169-182