Enrico Mattia Salonia
Game Theory - Spring 2026

The Trust Game

Ann has 10 euros and can give any part of it to Bob. Bob receives three times what Ann gives him. He can then decide how much of what he has to give back to Ann.

Denote by a the amount Ann gives to Bob and by b the amount Bob gives back to Ann. Their monetary rewards at the end of the game are as follows:

$$m_A(a,b) = 10 - a + b \qquad m_B(a,b) = 3a - b.$$

Exercise 1. Explain why this interaction constitutes a finite extensive-form game with perfect information.

From now on, assume Ann is selfish and greedy: she only wants to maximize her monetary reward. Bob, instead, might have one of two different preference types. He could be a *Beckerian altruist* (Becker, 1974), i.e., he cares about Ann's monetary reward; in that case, his preferences can be represented by the following utility function:

$$U_B(a,b) = m_B(a,b) + \alpha\, m_A(a,b).$$

Here, α is Bob's *altruism* parameter, which captures how much he cares about Ann's monetary reward. Alternatively, Bob might be *inequity averse* à la Fehr and Schmidt (1999), i.e., he prefers monetary rewards not to be too unequal. In that case, his utility function is:

$$U_B(a,b) = m_B(a,b) - \alpha\, \max\Big(m_A(a,b) - m_B(a,b), 0\Big) - \beta\, \max\Big(m_B(a,b) - m_A(a,b), 0\Big).$$

The parameter α represents Bob's *envy*, how much he dislikes Ann having more, and the parameter β represents Bob's *guilt*, how much he dislikes having more than Ann.

Exercise 2. Find the backward-induction solution under the assumption that Bob is a *Beckerian altruist*. How does the outcome depend on the altruism parameter α?

Solution to Exercise 2. We solve the game using backward induction, starting with Bob's decision. Bob chooses the amount b to maximize his utility function:

$$U_B(a,b) = m_B(a,b) + \alpha\, m_A(a,b)$$

Substituting the monetary payoff functions, with Bob receiving three times what Ann gives, we get:

$$U_B(a,b) = (3a - b) + \alpha(10 - a + b) = 10\alpha + a(3 - \alpha) + b(\alpha - 1)$$

Since Bob chooses b, his optimal action depends entirely on the coefficient of b, which is $(\alpha - 1)$:

- If $\alpha < 1$, the coefficient is negative. Bob maximizes his utility by returning the minimum possible amount: $b^* = 0$.

- If $\alpha > 1$, the coefficient is positive. Bob maximizes his utility by returning the maximum possible amount: $b^* = 3a$.

Moving backward to Ann's decision, she anticipates Bob's response to maximize her own selfish monetary payoff, $m_A(a, b) = 10 - a + b$:

- **Case 1 ($\alpha < 1$):** Ann knows $b^* = 0$. Her payoff is $10 - a$. To maximize this, she chooses $a^* = 0$.
- **Case 2 ($\alpha > 1$):** Ann knows $b^* = 3a$. Her payoff is $10 - a + 3a = 10 + 2a$. To maximize this, she chooses the maximum possible amount: $a^* = 10$.

The outcome depends on whether α is greater or less than 1. If $\alpha > 1$, Bob is sufficiently altruistic, and full trust is achieved ($a^* = 10$, $b^* = 30$).

Exercise 3. Find the backward-induction solution under the assumption that Bob is *inequity averse* à la Fehr and Schmidt (1999). Proceed in steps:

1. First, assume β is equal to 0.
2. Second, assume α is equal to 0. In which of the two cases does Ann give something to Bob?

Solution to Exercise 3. Let us first determine the difference in their monetary payoffs, which is central to Bob's utility:

$$m_B(a, b) - m_A(a, b) = (3a - b) - (10 - a + b) = 4a - 2b - 10$$

1. Assume $\beta = 0$, Bob only feels envy, $\alpha > 0$:

$$U_B(a, b) = m_B(a, b) - \alpha \max\Big(m_A(a, b) - m_B(a, b), 0\Big)$$

Any increase in b directly reduces Bob's monetary payoff and increases Ann's. If Ann has more money, giving her money increases Bob's envy penalty. If Bob has more money, giving her money reduces his wealth without any mitigating guilt penalty. Therefore, increasing b strictly decreases Bob's utility. Bob's best response is $b^* = 0$. Anticipating this, Ann maximizes $10 - a$ by choosing $a^* = 0$.

2. Assume $\alpha = 0$, Bob only feels guilt, $\beta > 0$:

$$U_B(a, b) = m_B(a, b) - \beta \max\Big(m_B(a, b) - m_A(a, b), 0\Big)$$

If Bob is ahead, i.e. $m_B > m_A$, his utility is:

$$U_B(a, b) = (3a - b) - \beta(4a - 2b - 10)$$

Taking the derivative of U_B with respect to b, we get $-1 + 2\beta$. This yields two sub-cases:

- If $\beta < 0.5$, the derivative is negative. Bob's greed outweighs his guilt, so he chooses $b^* = 0$. Anticipating this, Ann chooses $a^* = 0$.
- If $\beta > 0.5$, the derivative is positive. Bob's guilt outweighs his greed. He will increase b until the payoffs are equal ($m_A = m_B$), which occurs when $4a - 2b - 10 = 0$, or $b^* = 2a - 5$.

If $\beta > 0.5$, Ann knows Bob will equalize the payoffs. Equal payoffs mean they evenly split the total wealth generated in the game, which is $(10 - a) + 3a = 10 + 2a$. Ann's guaranteed payoff is half of this:

$$m_A = \frac{10 + 2a}{2} = 5 + a$$

To maximize $5 + a$, Ann will give the maximum amount: $a^* = 10$.

Ann only gives something to Bob in the second case, when $\alpha = 0$ and $\beta > 0.5$. Because the multiplier is 3, Ann's final payoff when giving $a = 10$ is 15 euros, giving her a strict incentive to trust Bob when he is sufficiently inequity-averse.

This analysis illustrates how trust can emerge when individuals care about fairness or others' welfare, helping explain why trust and reciprocity are frequently observed in experiments.

References

Becker, G. S. (1974). A theory of social interations. *Journal of Political Economy*, *82*(6), 1063–1093.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.