Incomplete Information Models of Guilt Aversion in the Trust Game

G. ATTANASI\textsuperscript{1}  P. BATTIGALLI\textsuperscript{2}  E. MANZONI\textsuperscript{3}

\textsuperscript{1}University of Strasbourg, BETA
\textsuperscript{2}Bocconi University, IGIER
\textsuperscript{3}University of Milan-Bicocca

Milan, 24/05/2013
Introduction

- We focus on the analysis of the Trust Minigame, a binary version of the Trust Game.

- Experimental work has shown that players in the Trust Game have non-selfish motivations that are not representable through distributional preferences (e.g., Cooper and Kagel 2013). We propose an approach in which players may display guilt aversion (for experimental evidence see, e.g., Charness and Dufwenberg 2006).

- We aim to develop results that may organize experimental evidence. Since a subject cannot know the preferences of the stranger he is paired with, an incomplete information analysis is necessary.
We model players’ subjective beliefs following Harsanyi’s approach: types parametrize beliefs. Hence we analyze *subjective* Bayesian equilibria: beliefs are not derived from an objective distribution of types.

This generates *heterogeneous first- and second-order beliefs* over co-players’ strategies, which is consistent with the experimental evidence (Charness and Dufwenberg 2006, Attanasi, Battigalli, Nagel 2013).

Such heterogeneity cannot be obtained by assuming that players know the objective distribution of types in the population, given the random matching of subjects in the lab.
We focus on *non-trivial* equilibria, in which Ann goes In with positive probability (i.e. at least a positive fraction of Ann’s types go In).
We apply the model of simple guilt by Battigalli & Dufwenberg (2007).

Player i’s guilt depends on

- his guilt sensitivity $\theta_i \geq 0$;
- the expected co-player’s disappointment (given his subjective beliefs).

Player i’s psychological utility function is

$$u_i = m_i - \theta_i \max\{0, \mathbb{E}_{-i}[m_{-i}] - m_i\},$$

We start from a model with role-dependent guilt sensitivity: $\theta_A = 0$ and $\theta_B \geq 0$ and this is common knowledge.
We introduce the following notation for players’ endogenous beliefs:

<table>
<thead>
<tr>
<th>Belief</th>
<th>Notation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann’s initial first-order belief</td>
<td>$\alpha_A$</td>
<td>$P_A[S]$</td>
</tr>
<tr>
<td>Bob’s initial first-order belief</td>
<td>$\alpha_B$</td>
<td>$P_B[I]$</td>
</tr>
<tr>
<td>A feature of Bob’s initial second-order belief</td>
<td>$\beta_B^\emptyset$</td>
<td>$E_B[\alpha_A]$</td>
</tr>
<tr>
<td>A feature of Bob’s conditional second-order belief</td>
<td>$\beta_B^I$</td>
<td>$E_B[\alpha_A</td>
</tr>
</tbody>
</table>

We let $m_i(z)$ denote the material payoff of $i$ at terminal history $z$. 
Trust Minigame with psychological utilities
Role-dependent guilt aversion

Ann can only be disappointed after \((I, K)\); her disappointment is

\[
\max\{0, \mathbb{E}_A[m_A] - m_A(I, K)\} = 2 \cdot \alpha_A + 0 \cdot (1 - \alpha_A) = 2\alpha_A.
\]
Bob’s psychological utility of $z = (I, K)$ is

$$u_B(I, K, \alpha_A) = m_B(I, K) - \theta_B \max\{0, \mathbb{E}_A[m_A] - m_A(I, K)\} = 4 - 2\theta_B \alpha_A.$$
Example:

Bob Shares if his conditional second-order belief satisfies $\beta_B^I > \frac{1}{\theta_B} = \frac{1}{3}$ and Keeps otherwise.
Role-dependent guilt aversion with complete information

$\theta_B = 3$

There is the trivial equilibrium $(O, K)$:

- Bob Keeps $\implies$ Ann stays Out;
- Ann expects Bob to Keep $\implies$ Bob Keeps ($\alpha_A = \beta^\cap_B = 0, \beta^I_B \leq \frac{1}{3}$).
Role-dependent guilt aversion with complete information

\( \theta_B = 3 \)

There is also the non-trivial equilibrium \((I, S)\):

- Bob Shares with probability one \(\Rightarrow\) Ann goes In.
- Ann expects Bob to Share with probability one \(\Rightarrow\) Bob Shares

\( (\alpha_A = \beta^\emptyset_B = \beta^I_B = 1) \).
We model incomplete information about $\theta_B$ using Harsanyi’s methodology: implicit, self-referential representation of interactive beliefs;

- $\theta_i$ is the guilt type (or utility type);
- $e_i \in [0, 1]$ is the epistemic type that parametrizes player $i$’s beliefs;
- $t_i = (\theta_i, e_i)$ is the overall type.

We have a continuum of types on both sides.
Role-dependent guilt aversion with incomplete information

Type structure, Ann’s types

- $\theta_A = 0$;
- $T_A = T_A^e = [0, 1]$;
- $t_A = P_{t_A}[\vartheta_B = \theta^H]$;
- Ann believes that $\vartheta_B$ and $e_B$ are independent;
- common marginal beliefs on $e_B$ given by cdf $F : \mathbb{R} \to [0, 1]$ (continuous cdf with support $[0, 1]$).

Ann’s belief is

$$\tau_A(t_A)[\vartheta_B = \theta^H \cap e_B \leq y] = t_A F(y),$$  \hspace{1cm} (1)

for all $t_A, y \in [0, 1]$.  

Attanasi, Battigalli, Manzoni ()   
Incomplete Info Models of Guilt Aversion   
Milan, 24/05/2013  \hspace{1cm} 13 / 29
\( \theta_B \in \{ \theta^L, \theta^H \} \), with \( \theta^L = 0 \) and \( \theta^H = 3 \) (in this presentation, for simplicity);

\( T_B = \{0, 3\} \times [0, 1] \)

Heterogeneous beliefs of \( e_B \) about \( t_A \) are given by cdf \( F_{e_B} : \mathbb{R} \rightarrow [0, 1] \) (continuous with support \([0, 1]\));

Conditional expectations \( E_{e_B}[t_A|t_A > x] \) are strictly increasing in \( e_B \).

Therefore Bob’s belief is

\[
\tau_B(\theta_B, e_B)[t_A \leq x] = F_{e_B}(x),
\]

for all \( e_B, x \in [0, 1], \theta_B \in \{0, 3\} \).
Bayesian equilibrium

Deﬁnitions

A Bayesian equilibrium of the Trust Minigame with guilt aversion and incomplete information is given by

\[(\sigma_A : T_A \rightarrow \{I, O\}, \sigma_B : T_B \rightarrow \{S, K\})\]

such that for every \(i\) and \(t_i\), \(\sigma_i(t_i)\) maximizes \(i\)'s expected utility given \(t_i\)'s endogenous beliefs about the co-player’s choice and beliefs.
Endogenous beliefs

- Ann’s first-order endogenous belief:
  \[
  \alpha_A(t_A) = \tau_A(t_A)[\sigma_B = S]
  \]

- Bob’s second-order initial belief:
  \[
  \beta^\sigma_B(t_B) = \mathbb{E}_{t_B}[\alpha_A] = \int \alpha_A(t_A)\tau_B(t_B)[dt_A].
  \]

- Bob’s second-order conditional belief:
  \[
  \beta^I_B(t_B) = \mathbb{E}_{t_B}[\alpha_A|\sigma_A = I] = \frac{\int_{(\sigma_A)^{-1}(I)} \alpha_A(t_A)\tau_B(t_B)[dt_A]}{\tau_B(t_B)[\sigma_A = I]}
  \]
Role-dependent guilt aversion with incomplete information

Equilibrium analysis

Ann \xrightarrow{\text{In}} Bob \xrightarrow{\text{Share}} (2, 2)

Out \quad \text{Keep}

(1, 1) \quad (0, 4 - 2\theta_B a_A)

Bob’s types \((\theta^L, e_B)\) choose K.
Role-dependent guilt aversion
Equilibrium analysis

Only \( t_A \) such that \( \alpha_A(t_A) > \frac{1}{2} \) choose \( \text{In} \) in equilibrium.

A positive fraction of Ann’s types chooses \( \text{I} \) in equilibrium
\( \implies \beta_B^I(e_B) > \frac{1}{2} \) by Bayes’ rule.
Role-dependent guilt aversion
Equilibrium analysis

\[
\begin{array}{ccc}
\text{Ann} & \xrightarrow{\text{In}} & \text{Bob} & \xrightarrow{\text{Share}} & (2, 2) \\
\text{Out} & \downarrow & \text{Keep} & \downarrow & \\
(1, 1) & & (0, 4 - 2\theta_B\alpha_A) & & \\
\end{array}
\]

- \((\theta^H, e_B)\) Shares if \(\beta^I_B(e_B) > \frac{1}{\theta^H}\).
- \(\beta^I_B(e_B) > \frac{1}{2} > \frac{1}{3} = \frac{1}{\theta^H} \implies \sigma_B(\theta^H, e_B) = S\) for every \(e_B\).
- Given \(\sigma_B(\theta^k, \cdot)\) as above, \(\alpha_A(t_A) = P(\theta_B = \theta^H) = t_A\).
- \(\sigma_A(t_A) = 1\) iff \(\alpha_A(t_A) > \frac{1}{2} \implies \sigma_A(t_A) = 1\) iff \(t_A > \frac{1}{2}\).
Proposition

In the model with role-dependent guilt aversion, the unique non-trivial Bayesian equilibrium is \((\sigma_A, \sigma_B)\), such that for every \(t_A\) and \(e_B\)

(a) \(\sigma_A (t_A) = 0\) iff \(t_A \leq \frac{1}{2}\);

(b) \(\sigma_B (\theta^L, e_B) = K\) and \(\sigma_B (\theta^H, e_B) = S\).

Notice that there is heterogeneity of the endogenous beliefs. In particular:

- \(\alpha_A (t_A) = t_A\);
- \(\beta^l_B (e_B) = \left(1 - F_{e_B} \left(\frac{1}{2}\right)\right)^{-1} \int_{\frac{1}{2}}^{1} t_A dF_{e_B} (t_A)\).
Role-independent guilt aversion

Also Ann cares about the co-player’s disappointment. Bob can be disappointed if Ann goes Out, in this case Bob’s disappointment is:

\[
\max \{0, \mathbb{E}_B[m_B] - m_B(O)\} = \mathbb{E}_B[m_B] - 1 = \left\{ \begin{array}{ll}
\alpha_B, & \text{if } s_B = S \\
3\alpha_B, & \text{if } s_B = K.
\end{array} \right.
\]
Role-independent guilt aversion

We let

\[ A = E_A [E_B [m_B] - m_B(O)] \]

be Ann’s expectation of Bob’s disappointment.
Role-independent guilt aversion

Type structure

- We model incomplete information about $\theta$ using Harsanyi’s methodology.
- We have a continuum of types on both sides.
- $\theta_i \in \{\theta^L, \theta^H\}$ is player $i$’s, with $\theta^L = 0$, and $\theta^H = 3$ (for simplicity, in these slides).
- $e_i = \mathbb{P}_{(\theta_i, e_i)}[\vartheta_i = \theta^H]$ is player $i$’s epistemic type.
- Each player $i$ believes that $\vartheta_i$ and $e_i$ are independent.
- Player $i$’s marginal beliefs on $e_i$ are given by cdf $F : \mathbb{R} \rightarrow [0, 1]$ (continuous with support $[0, 1]$).

Player $i$’s belief is

$$\tau_i(t_i)[\vartheta_i = \theta^H \cap e_i \leq x] = e_i F(x),$$  \hspace{1cm} (3)

with $\mathbb{E}[e_i] > \frac{1}{3\theta^H} = \frac{1}{9}$ \hspace{1cm} (4)
Role-independent guilt aversion
Equilibrium analysis I

\[ \text{Ann} \overset{In}{\rightarrow} \text{Bob} \overset{Share}{\rightarrow} (2, 2) \]

\[ \text{Out} \downarrow \quad \text{Keep} \downarrow \]

\[ (1 - \theta_A \bar{\beta}_A, 1) \quad (0, 4 - 2\theta_B \beta^I_B) \]

- \( \sigma_B \left( \theta^L, e_B \right) = K \) for every \( e_B \).
- If also \( \sigma_B \left( \theta^H, e_B \right) = K \) for every \( e_B \) \( \implies \) \( \sigma_A \left( \theta^L, e_A \right) = O \).
If $\sigma_B(\theta^k, e_B) = K$ for every $(\theta^k, e_B)$ and $\bar{\beta}_A$ is sufficiently high, $(\theta^H, e_A)$ may go In not to disappoint Bob;

$E[e_B] > 1/9$ ensures that when $\sigma_A(\theta^H, e_A) = 1$, in equilibrium $\bar{\beta}_A$ is high enough to make it optimal for Ann to go In.
If in equilibrium $\alpha_A = 0$, then $\sigma_B (\theta^k, e_B) = K$ for every $(\theta^k, e_B)$. 
Proposition

The following is a non-trivial equilibrium of the role-independent model: for every $e_A$ and $e_B$

(a) $\sigma_A (\theta^L, e_A) = O$;
(b) $\sigma_A (\theta^H, e_A) = I$;
(c) $\sigma_B (\theta^L, e_B) = \sigma_B (\theta^H, e_B) = K$.

In such equilibrium

- $\sigma_A (\theta^H, e_A) = I$ because a guilt averse Ann does not want to disappoint Bob;
- $\sigma_B (\theta^H, e_B) = K$ because $\alpha_A = 0$.

Notice that also in this case there is heterogeneity of endogenous beliefs. In particular $\alpha_B (e_B) = e_B$. 

Proposition

All the non-trivial equilibria \((\sigma_A, \sigma_B)\) of the role-independent model have the following structure: for every \(e_A\) and \(e_B\)

\[
\sigma_A (\theta^L, e_A) = \begin{cases} 
O, & \text{if } e_A \leq \hat{e}_A^L \\
I, & \text{otherwise}
\end{cases},
\]

\[
\sigma_A (\theta^H, e_A) = \begin{cases} 
O, & \text{if } e_A \leq \hat{e}_A^H \\
I, & \text{otherwise}
\end{cases},
\]

\[
\sigma_B (\theta^L, e_B) = K,
\]

\[
\sigma_B (\theta^H, e_B) = \begin{cases} 
K, & \text{if } e_B \geq \hat{e}_B^H \\
S, & \text{otherwise}
\end{cases}.
\]

with \(0 \leq \hat{e}_A^H < \hat{e}_A^L \leq 1\) and \(0 \leq \hat{e}_B^H < 1\).
Empirical predictions

- Subjective beliefs do not coincide with the objective distribution of types in the population.
- Knowledge of the objective distribution is necessary to make empirical predictions.
- If the objective distribution of types has a rich support, and $\theta^H$ is sufficiently high we expect:
  - heterogeneous behavior;
  - heterogeneous hierarchical beliefs about behavior.
- If $e_i$ and $\vartheta_i$ are statistically independent we expect positive correlation between pro-social actions and endogenous second order beliefs (cf. Charness and Dufwenberg 2006).