Incomplete Information Models of Guilt Aversion in the Trust Game

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

\textsuperscript{1}Paris School of Economics
\textsuperscript{2}Bocconi University, Milan
\textsuperscript{3}Bicocca University, Milan

Sintelnet workshop, IRIT UT3
Psychological Game Theory

- Traditional Game Theory: \( u_i = u_i(\text{actions}) \)
- Problem 1: the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend directly on beliefs (about choices, beliefs, or information).
- Extension 1: Geanakoplos, Pearce and Stacchetti (1989) introduce belief-dependent motivations: \( u_i = u_i(\text{actions, beliefs}_i) \)
  \( \text{beliefs}_i: \) initial (pre-play) beliefs of player \( i \) about strategies and about beliefs of others.
- Problem 2: conditional (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.
- Extension 2: Battigalli and Dufwenberg (2009) introduce conditional beliefs: \( u_i(\text{actions, cond.beliefs}_i, \text{cond.beliefs}_{-i}) \)
- Summary: In games with belief-dependent motivations there are two channels through which beliefs and information affect behavior.
Psychological Game Theory

- **Traditional Game Theory:** \( u_i = u_i(\text{actions}) \)

- **Problem 1:** the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend *directly* on beliefs (about choices, beliefs, or information).

- **Extension 1:** Geanakoplos, Pearce and Stacchettii (1989) introduce *belief-dependent motivations*: \( u_i = u_i(\text{actions, beliefs}_i) \)
  - \( \text{beliefs}_i \): initial (pre-play) beliefs of player \( i \) about strategies and about beliefs of others.

- **Problem 2:** *conditional* (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.

- **Extension 2:** Battigalli and Dufwenberg (2009) introduce *conditional beliefs*: \( u_i(\text{actions, cond.beliefs}_i, \text{cond.beliefs}_{-i}) \)

- **Summary:** In games with belief-dependent motivations there are two channels through which beliefs and information affect behavior.
Psychological Game Theory

- Traditional Game Theory: \( u_i = u_i(\text{actions}) \)

- **Problem 1**: the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend directly on beliefs (about choices, beliefs, or information).

- **Extension 1**: Geanakoplos, Pearce and Stacchetti (1989) introduce belief-dependent motivations: \( u_i = u_i(\text{actions}, \text{beliefs}_i) \)
  
  \text{beliefs}_i: initial (pre-play) beliefs of player \( i \) about strategies and about beliefs of others.

- **Problem 2**: conditional (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.

- **Extension 2**: Battigalli and Dufwenberg (2009) introduce conditional beliefs: \( u_i(\text{actions}, \text{cond.beliefs}_i, \text{cond.beliefs}_{-i}) \)

- **Summary**: In games with belief-dependent motivations there are two channels through which beliefs and information affect behavior.
Psychological Game Theory

- **Traditional Game Theory:** $u_i = u_i(\text{actions})$

- **Problem 1:** the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend directly on beliefs (about choices, beliefs, or information).

- **Extension 1:** Geanakoplos, Pearce and Stacchetti (1989) introduce belief-dependent motivations: $u_i = u_i(\text{actions}, \text{beliefs}_i)$
  
  - $\text{beliefs}_i$: initial (pre-play) beliefs of player $i$ about strategies and about beliefs of others.

- **Problem 2:** conditional (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.

- **Extension 2:** Battigalli and Dufwenberg (2009) introduce conditional beliefs: $u_i(\text{actions}, \text{cond.\,beliefs}_i, \text{cond.\,beliefs}_{-i})$

- **Summary:** In games with belief-dependent motivations there are two channels through which beliefs and information affect behavior.
Psychological Game Theory

- **Traditional Game Theory:** $u_i = u_i(\text{actions})$

- **Problem 1:** the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend *directly* on beliefs (about choices, beliefs, or information).

- **Extension 1:** Geanakoplos, Pearce and Stacchetti (1989) introduce *belief-dependent motivations:* $u_i = u_i(\text{actions}, \text{beliefs}_i)$ 
  
  - **beliefs**$_i$: initial (pre-play) beliefs of player $i$ about strategies and about beliefs of others.

- **Problem 2:** *conditional* (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.

- **Extension 2:** Battigalli and Dufwenberg (2009) introduce *conditional beliefs:* $u_i(\text{actions}, \text{cond. beliefs}_i, \text{cond. beliefs}_{-i})$

- **Summary:** In games with belief-dependent motivations there are *two channels* through which beliefs and information affect behavior.
Psychological Game Theory

- Traditional Game Theory: \( u_i = u_i(\text{actions}) \)
- **Problem 1:** the motivation of decision makers who are affected by and care for emotions, reciprocity, or social conformity may depend *directly* on beliefs (about choices, beliefs, or information).
- **Extension 1:** Geanakoplos, Pearce and Stacchetti (1989) introduce *belief-dependent motivations*: \( u_i = u_i(\text{actions}, \text{beliefs}_i) \)
  \( \text{beliefs}_i: \) initial (pre-play) beliefs of player \( i \) about strategies and about beliefs of others.
- **Problem 2:** *conditional* (second-order) beliefs are not considered. Yet they are crucial for applications and theoretical interpretations of experimental findings.
- **Extension 2:** Battigalli and Dufwenberg (2009) introduce *conditional beliefs*: \( u_i(\text{actions}, \text{cond. beliefs}_i, \text{cond. beliefs}_{-i}) \)
- **Summary:** In games with belief-dependent motivations there are *two channels* through which beliefs and information affect behavior.
Main Features of Psychological Games
(focusing on Guilt and Reciprocity in a Trust Game)

1. Attanasi, Battigalli and Nagel (2012): experimental data

2. **Role-dependent** Guilt Aversion:
   complete and incomplete information analysis

3. **Role-independent** Guilt Aversion:
   complete and incomplete information analysis

4. Discussion

5. Attanasi, Battigalli, Manzoni
Main Features of Psychological Games (focusing on Guilt and Reciprocity in a Trust Game)

Attanasi, Battigalli and Nagel (2012): experimental data

Role-dependent Guilt Aversion: complete and incomplete information analysis

Role-independent Guilt Aversion: complete and incomplete information analysis

Discussion
Road map

1. Main Features of Psychological Games (focusing on Guilt and Reciprocity in a Trust Game)


3. **Role-dependent** Guilt Aversion: complete and *incomplete information* analysis

4. **Role-independent** Guilt Aversion: complete and *incomplete information* analysis

5. Discussion
Main Features of Psychological Games (focusing on Guilt and Reciprocity in a Trust Game)

Attanasi, Battigalli and Nagel (2012): experimental data

1. Role-dependent Guilt Aversion: complete and incomplete information analysis
2. Role-independent Guilt Aversion: complete and incomplete information analysis

Discussion
Main Features of Psychological Games
(focusing on Guilt and Reciprocity in a Trust Game)

Attanasi, Battigalli and Nagel (2012): experimental data

Role-dependent Guilt Aversion:
complete and incomplete information analysis

Role-independent Guilt Aversion:
complete and incomplete information analysis

Discussion
Main Features of Psychological Games

EXAMPLE: Trust Game with material payoffs

\[ \begin{align*}
\alpha &= \Pr_A[\text{Share if Continue}] : \text{initial 1st-order belief of } A \\
\beta &= \mathbb{E}_B[\alpha|\text{Continue}] : \text{conditional 2nd-order belief of } B
\end{align*} \]
**Main Features of Psychological Games**

**EXAMPLE: Trust Game with material payoffs**

\[
\begin{bmatrix}
A \\
| Dissolve \\
| \\
(1) \\
(1)
\end{bmatrix} \quad \begin{array}{r}
\text{Continue} \\
\longrightarrow
\end{array} \quad \begin{bmatrix}
B \\
| Take \\
| \\
(0) \\
(4)
\end{bmatrix}
\]

\[
\alpha = \Pr_A[\text{Share if Continue}] : \text{initial 1st-order belief of } A
\]

\[
\beta = \mathbb{E}_B[\alpha | \text{Continue}] : \text{conditional 2nd-order belief of } B
\]
Main Features of Psychological Games

EXAMPLE: Trust Game with material payoffs

\[
\begin{array}{c|c|c|c|c}
 & \text{A} & \text{Continue} & \text{B} & \text{Share} \\
\hline
\text{Dissolve} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \rightarrow & \begin{pmatrix} 0 \\ 4 \end{pmatrix} & \begin{pmatrix} 2 \\ 2 \end{pmatrix} \\
\text{Take} & \end{array}
\]

- \( \alpha = \Pr_A[\text{Share if Continue}] : \text{initial 1st-order belief of A} \)

- \( \beta = \mathbb{E}_B[\alpha|\text{Continue}] : \text{conditional 2nd-order belief of B} \)
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSION [Dufwenberg (2002)]

If $B$ dislikes letting $A$ down (Guilt Aversion), his utility from $(\text{Continue, Take})$ depends negatively on his expectation of $\alpha$.

- $-2\alpha =$ how much $A$ would feel ‘let down’ after $(\text{Continue, Take})$
- $-2\beta =$ $B$’s expectation of $-2\alpha$, given $\text{Continue}$
- $G := B$’s sensitivity to Guilt
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSION [Dufwenberg (2002)]

If $B$ dislikes letting $A$ down (Guilt Aversion), his utility from $(\text{Continue}, \text{Take})$ depends negatively on his expectation of $\alpha$.

- $-2\alpha = \text{how much } A \text{ would feel ‘let down’ after } (\text{Continue}, \text{Take})$
- $-2\beta = B's \text{ expectation of } -2\alpha, \text{ given } \text{Continue}$
- $G := B's \text{ sensitivity to Guilt}$
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERTION [Dufwenberg (2002)]

<table>
<thead>
<tr>
<th>A</th>
<th>Continue</th>
<th>B</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dissolve</td>
<td></td>
<td>Take</td>
<td>(2 2)</td>
</tr>
<tr>
<td>(1)</td>
<td></td>
<td>(0)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

If B dislikes letting A down (Guilt Aversion), his utility from (Continue, Take) depends negatively on his expectation of $\alpha$.

- $-2\alpha =$ how much A would feel ‘let down’ after (Continue, Take)
- $-2\beta =$ B’s expectation of $-2\alpha$, given Continue
- $G :=$ B’s sensitivity to Guilt
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSION [Dufwenberg (2002)]

If $B$ dislikes letting $A$ down (*Guilt Aversion*), his utility from $(\text{Continue, Take})$ depends negatively on his expectation of $\alpha$.

- $-2\alpha = $ how much $A$ would feel ‘let down’ after $(\text{Continue, Take})$
- $-2\beta = $ $B$’s expectation of $-2\alpha$, given $\text{Continue}$
- $G := $ $B$’s sensitivity to Guilt
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSION [Dufwenberg (2002)]

If $B$ dislikes letting $A$ down (Guilt Aversion), his utility from $(\text{Continue}, \text{Take})$ depends negatively on his expectation of $\alpha$.

- $-2\alpha =$ how much $A$ would feel ‘let down’ after $(\text{Continue}, \text{Take})$
- $-2\beta =$ $B$'s expectation of $-2\alpha$, given $\text{Continue}$
- $G := B$'s sensitivity to Guilt
Main Features of Psychological Games

EXAMPLE: Trust Game with GUilt AVERSION [Dufwenberg, 2002]

\[
\begin{pmatrix}
1 \\
1 \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
2 \\
2 \\
\end{pmatrix}
\]

In both cases, \( \alpha \) is unknown to B.

\( G \) commonly known \( \implies \) Psy game with Complete Info

\( G \) not commonly known \( \implies \) Psy game with Incomplete Info
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSIÓN [Dufwenberg, 2002]

![Game Diagram]

- **A**
  - Dissolve
  - \[
  \begin{pmatrix}
  1 \\
  1
  \end{pmatrix}
  \]

- **B**
  - Take
  - \[
  \begin{pmatrix}
  4 - G \cdot 2\beta \\
  0
  \end{pmatrix}
  \]

- **Share**
  - \[
  \begin{pmatrix}
  2 \\
  2
  \end{pmatrix}
  \]

\( G \) *commonly known* \( \implies \) Psy game with *Complete Info*

\( G \) *not commonly known* \( \implies \) Psy game with *Incomplete Info*

In both cases, \( \alpha \) is *unknown* to \( B \).
Main Features of Psychological Games

EXAMPLE: Trust Game with GUILT AVERSİON [Dufwenberg, 2002]

\[
\begin{array}{ccc}
A & \text{Continue} & B \\
\mid \downarrow \text{Dissolve} & \mid \downarrow \text{Take} \\
\begin{pmatrix} 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 4 - G \cdot 2\beta \end{pmatrix} \\
\end{array}
\]

\[\begin{align*}
G \text{ commonly known} & \implies \text{Psy game with Complete Info} \\
G \text{ not commonly known} & \implies \text{Psy game with Incomplete Info}
\end{align*}\]

In both cases, \(\alpha\) is unknown to \(B\).
EXAMPLE: Trust Game with RECIPROCITY [Rabin, 1993; Dufwen. and Kirchst., 2004]

If $B$ is motivated by (intention-based) *Reciprocity*, his utility from (*Continue*,*Share*) depends negatively on his expectation of $\alpha$.

- Is *Cont.* a fair action? It depends on *intentions*:
  - *Cont.* is kind if $\alpha$ is low. *Cont.* is perceived as kind by $B$ if $\beta$ is low.
  - $R := B$’s sensitivity to *Reciprocity*
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Rabin, 1993; Dufwen. and Kirchst., 2004]

If $B$ is motivated by (intention-based) *Reciprocity*, his utility from \((\text{Continue}, \text{Share})\) depends negatively on his expectation of $\alpha$.

- Is $\text{Cont.}$ a fair action? It depends on intentions:
  - $\text{Cont.}$ is kind if $\alpha$ is low. $\text{Cont.}$ is perceived as kind by $B$ if $\beta$ is low
  - $R := B$’s sensitivity to Reciprocity
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Rabin, 1993; Dufwen. and Kirchst., 2004]

If $B$ is motivated by (intention-based) *Reciprocity*, his utility from $(\text{Continue, Share})$ depends negatively on his expectation of $\alpha$.

- Is $\text{Cont.}$ a fair action? It depends on intentions:
  - $\text{Cont.}$ is kind if $\alpha$ is low. $\text{Cont.}$ is perceived as kind by $B$ if $\beta$ is low.
  - $R := B$’s sensitivity to Reciprocity.
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Rabin, 1993; Dufwen. and Kirchst., 2004]

If \( B \) is motivated by (intention-based) Reciprocity, his utility from \((\text{Continue,Share})\) depends negatively on his expectation of \( \alpha \).

- Is Cont. a fair action? It depends on intentions:
- Cont. is kind if \( \alpha \) is low. Cont. is perceived as kind by \( B \) if \( \beta \) is low
- \( R := B \)'s sensitivity to Reciprocity
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Rabin, 1993; Dufwen. and Kirchst., 2004]

If $B$ is motivated by (intention-based) Reciprocity, his utility from $(\text{Continue, Share})$ depends negatively on his expectation of $\alpha$.

- Is $\text{Cont.}$ a fair action? It depends on intentions:
  - $\text{Cont.}$ is kind if $\alpha$ is low. $\text{Cont.}$ is perceived as kind by $B$ if $\beta$ is low
  - $R := B$’s sensitivity to Reciprocity
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Dufwenberg and Kirchsteiger, 2004]

\[ \begin{align*}
A & \quad \text{Continue} \quad \text{Share} \\
\text{Dissolve} & \quad \text{Take} \\
1 & \quad 0 \\
1 + R \left( \beta - \frac{3}{2} \right) & \quad 4 \\
\end{align*} \]

\[ 2 + R \left( 3 - 2\beta \right) \]

\( R \) commonly known \( \implies \) Psy game with Complete Info
\( R \) not commonly known \( \implies \) Psy game with Incomplete Info

In both cases, \( \alpha \) is unknown to \( B \).
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Dufwenberg and Kirchsteiger, 2004]

\[
\text{Continue} \quad \begin{align*}
A & \quad \xrightarrow{\text{Dissolve}} \quad 1 + R\left(\beta - \frac{3}{2}\right) \\
B & \quad \xrightarrow{\text{Take}} \quad \begin{pmatrix} 0 \\ 4 \end{pmatrix}
\end{align*}
\]

\[
\text{Share} \quad \begin{align*}
2 + R(3 - 2\beta)
\end{align*}
\]

- \( R \) commonly known \( \iff \) Psy game with Complete Info
- \( R \) not commonly known \( \iff \) Psy game with Incomplete Info

In both cases, \( \alpha \) is unknown to \( B \).
Main Features of Psychological Games

EXAMPLE: Trust Game with RECIPROCITY [Dufwenberg and Kirchsteiger, 2004]

\[
\begin{align*}
\text{Continue} & \quad \text{A} & \text{Share} & \quad \text{B} \\
\text{Dissolve} & \quad 1 & \quad \text{Take} & \quad \begin{pmatrix} 2 \\ 2 + R(3 - 2\beta) \end{pmatrix}
\end{align*}
\]

\[
\begin{pmatrix}
1 + R(\beta - \frac{3}{2}) \\
0 \\
4
\end{pmatrix}
\]

\begin{align*}
R \text{ commonly known} & \quad \implies \quad \text{Psy game with Complete Info} \\
R \text{ not commonly known} & \quad \implies \quad \text{Psy game with Incomplete Info}
\end{align*}

In both cases, \(\alpha\) is unknown to B.
### Attanasi, Battigalli and Nagel (2012): experimental results

**Disclosure of Belief-Dependent Preferences in the Trust Game**

<table>
<thead>
<tr>
<th>Categories of Types</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NoQ-QnoT</td>
</tr>
<tr>
<td>(More) Guilt</td>
<td>40</td>
</tr>
<tr>
<td>(More) Reciprocity</td>
<td>10</td>
</tr>
<tr>
<td>Mix Guilt-Reciprocity</td>
<td>5</td>
</tr>
<tr>
<td>Fairness</td>
<td>2</td>
</tr>
<tr>
<td>Self Interest</td>
<td>13</td>
</tr>
<tr>
<td>Unclassified</td>
<td>10</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>80</strong></td>
</tr>
</tbody>
</table>

*B’s psychological types according to the questionnaire.*
### Categories of Types

<table>
<thead>
<tr>
<th>Categories of Types</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NoQ-QnoT</td>
</tr>
<tr>
<td>(More) Guilt</td>
<td>40</td>
</tr>
<tr>
<td>(More) Reciprocity</td>
<td>10</td>
</tr>
<tr>
<td>Mix Guilt-Reciprocity</td>
<td>5</td>
</tr>
<tr>
<td>Fairness</td>
<td>2</td>
</tr>
<tr>
<td>Self Interest</td>
<td>13</td>
</tr>
<tr>
<td>Unclassified</td>
<td>10</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>80</strong></td>
</tr>
</tbody>
</table>

*B’s psychological types according to the questionnaire.*
Disclosure of Belief-Dependent Preferences in the Trust Game

Bs’ frequency of Share.

Bs’ 2nd-order belief of Share.
As’ frequency of *Continue*.  

As’ 1st-order belief of *Share*.  

Attanasi, Battigalli and Nagel (2012): experimental results  
Disclosure of Belief-Dependent Preferences in the Trust Game
We have built a design to test a guilt+reciprocity model of psychological game:
- elicitation of belief-dependent preferences through a structured questionnaire;
- transmission procedure allowing to make elicited preferences public information between truster and trustee.

Guilt averse trustees cooperate more than reciprocity concerned and than self-interested trustees under incomplete information.

Public information of trustee’s feelings sensitivity polarizes distribution of outcomes.

Public information of trustee’s Guilt Aversion leads to trust and cooperation within the pair. Public information of trustee’s Reciprocity does not.
We have built a design to test a guilt+reciprocity model of psychological game:
- elicitation of belief-dependent preferences through a structured questionnaire;
- transmission procedure allowing to make elicited preferences public information between truster and trustee.

Guilt averse trustees cooperate more than reciprocity concerned and than self-interested trustees under incomplete information.

Public information of trustee’s feelings sensitivity polarizes distribution of outcomes.

Public information of trustee’s Guilt Aversion leads to trust and cooperation within the pair. Public information of trustee’s Reciprocity does not.
We have built a design to test a guilt+reciprocity model of psychological game:
- elicitation of belief-dependent preferences through a structured questionnaire;
- transmission procedure allowing to make elicited preferences public information between truster and trustee.

Guilt averse trustees cooperate more than reciprocity concerned and than self-interested trustees under incomplete information.

Public information of trustee’s feelings sensitivity polarizes distribution of outcomes.

Public information of trustee’s Guilt Aversion leads to trust and cooperation within the pair. Public information of trustee’s Reciprocity does not.
We have built a design to test a guilt + reciprocity model of psychological game:
- elicitation of belief-dependent preferences through a structured questionnaire;
- transmission procedure allowing to make elicited preferences public information between truster and trustee.

Guilt averse trustees cooperate more than reciprocity concerned and than self-interested trustees under incomplete information.

Public information of trustee’s feelings sensitivity polarizes distribution of outcomes.

Public information of trustee’s Guilt Aversion leads to trust and cooperation within the pair. Public information of trustee’s Reciprocity does not.
We provide an analysis of trust games with guilt aversion and incomplete information.

We follow Battigalli and Dufwenberg (2009).

We consider **four** models, depending on the following characteristics:

- Guilt aversion: role-dependent (only B is guilt averse) or role-independent;
- Subjective beliefs: homogeneous or heterogeneous.
Incomplete information models of guilt aversion in the trust game

Why is it relevant?

- We show how to model incomplete information with non-standard preferences.
- Useful for experiments which focus on non-standard preferences where it is unlikely that preferences are known.

Let’s start from the analysis of:
- role-dependent guilt aversion case with complete information.
The disappointment of A after (ln, D), given her belief \( \alpha \) is

\[
D_A(ln, D, \alpha) = \max \{0, m_A(ln, s_B, \alpha) - m_A(ln, D)\} \\
= [\alpha \cdot m_A(ln, C) + (1 - \alpha) \cdot m_A(ln, D)] - m_A(ln, D) = 2\alpha
\]
The psychological utility of $B$, given his guilt parameter $\theta$, is

$$u_B(In, D, \alpha) = m_B(In, D) - \theta \cdot D_A(In, D, \alpha) = 4 - \theta \cdot 2\alpha$$
Role-dependent guilt aversion
Complete information

\[ \begin{array}{c}
\text{A} & \rightarrow & \text{B} \\
\text{Dissolve} & & \text{Take} \\
\begin{pmatrix} 1 \\ 1 \end{pmatrix} & & \begin{pmatrix} 0 \\ 4 - \theta \cdot 2\alpha \end{pmatrix}
\end{array} \]

\( (\text{Out}, D) \) is an equilibrium for every value of \( \theta \):

- if \( A \) believes that \( B \) will choose \( D \) it is optimal for her to choose \( \text{Out} \);
- with \( \alpha = 0 \) it is optimal for \( B \) to choose \( D \).
Role-dependent guilt aversion
Complete information

\[
\begin{array}{ccc}
\text{A} & \text{Continue} & \text{B} \\
\text{Dissolve} & \rightarrow & \text{Take} \\
1 & \rightarrow & 0 \\
1 & \rightarrow & 4 - \theta \cdot 2\alpha \\
\end{array}
\]

\((ln, C)\) is an equilibrium for \(\theta > 1\):

- if \(A\) believes that \(B\) will choose \(C\) it is optimal for her to choose \(ln\);
- if \(\alpha = 1\), then, given that \(\theta > 1\), it is optimal for \(B\) to choose \(C\).
In order to model the incomplete information in this game, we need to introduce a description of types and beliefs.

**Types:** We consider structures of *Harsanyi types* where the set of types $T_i$ can be factorized in:

- a set of *payoff types* $\Theta_i$;
- a set of *epistemic types* identified with the interval $[0, 1]$, that is $t_i = (\theta_i, e_i) \in \Theta_i \times [0, 1] = T_i$.

When we model role-dependent guilt aversion we assume $\Theta_A \neq \Theta_B$, where $\Theta_A$ is a singleton and $\Theta_B = \{\theta^L, \theta^H\}$, with $\theta^H > \theta^L = 0$. Therefore $T_A = [0, 1]$ and $T_B = \{\theta^L, \theta^H\} \times [0, 1]$. 
Beliefs:
We call *exogenous* a belief about an exogenous variable or a parameter. We call *endogenous* a belief about a variable we try to explain with the strategic analysis of the game.

In our case the relevant beliefs are:

- **exogenous beliefs**: we consider the beliefs of A on $T_B$, $\tau_A$ and the beliefs of B on $T_A$, $\tau_B$;

- **endogenous beliefs**: the endogenous beliefs that are relevant for our analysis are A’s first order endogenous belief $\alpha_A$ and B’s second order conditional belief $\beta_B$ where

$$\alpha_A = \Pr_A[C]$$

$$\beta_B = E_B[\alpha_A \mid In]$$
We assume that, in A’s view, B’s epistemic component $e_B$ is independent of $\theta_B$ (coherent with Attanasi, Battigalli & Nagel, 2012).

A-subjects have heterogeneous beliefs about $\theta_B$, parametrized by $t_A$, and common beliefs about $e_B$ given by a cdf $F(\cdot)$:

$$\tau_A(t_A)[\theta_B = \theta^H, e_B \leq y] = t_A F(y)$$

B-subjects have heterogeneous beliefs about A’s type parametrized by $e_B$; such beliefs are distributed according to a perturbation of the uniform distribution on $[0, e_B]$:

$$\tau_B(\theta_B, e_B)[t_A \leq x] = (1 - \varepsilon) \min \left(1, \frac{x}{e_B} \right) + \varepsilon x$$
A *Bayesian perfect equilibrium* is a pair of decision functions $(\sigma_A : T_A \rightarrow \{\text{In, Out}\}, \sigma_B : T_B \rightarrow \{C, D\})$ such that:

- for each player $i = A, B$ and type $t_i \in T_i$, choice $\sigma_i(t_i)$ maximizes $i$’s expected (psychological) utility, given the beliefs of type $t_i$ about the co-player’s choice and the endogenous beliefs induced by $\sigma_{-i}$;

- in the case of $i = B$, the maximization is conditional on $\text{In}$, with conditional beliefs computed by Bayes rule, if possible.
Role-dependent guilt aversion
Incomplete information: optimal strategies

- \( \sigma_B(\theta^L, e_B) = D \); given \( \theta^L = 0 \), if B's payoff type is \( \theta^L \) he has standard preferences and he maximizes his monetary payoff.

- \( \sigma_B(\theta^H, e_B) = C \) if and only if \( e_B > y \); the higher \( e_B \), the higher is B's distribution on A's types, the higher is \( E_B[\alpha \mid ln] \), and therefore his incentives to cooperate.

- \( \sigma_A(t_A) = ln \) if and only if \( t_A > x \); the higher is \( t_A \), the higher is the probability that B is guilt averse, the higher are B's incentives to cooperate and therefore A's incentives to choose \( ln \).
Role-dependent guilt aversion
Incomplete information: endogenous beliefs

A’s first order endogenous belief:

$$\alpha_A(t_A) = \tau_A(t_A)[\theta_B = \theta^H, e_B \geq y] = t_A(1 - F(y)).$$

B’s second order endogenous belief: Note that every type \((\theta_B, e_B)\) assigns positive probability to \(ln:\)

$$\tau_B(\theta_B, e_B)[\sigma_A = ln] = (1 - \varepsilon) \max \left( \frac{e_B - x}{e_B}, 0 \right) + \varepsilon (1 - x) > 0.$$  

Hence we obtain a conditional second-order belief by Bayes rule:

$$\beta_B(\theta_B, e_B) = E_{(\theta_B, e_B)}[\alpha_A|ln]$$

$$= \begin{cases} 
(1 - F(y)) \frac{(1+x)}{2}, & \text{if } e_B \leq x \leq 1, \\
(1 - F(y)) \frac{(1-\varepsilon)(e_B^2-x^2) + \varepsilon e_B(1-x^2)}{2[(1-\varepsilon)(e_B-x)+\varepsilon e_B(1-x)]}, & \text{if } 0 \leq x < e_B. 
\end{cases}$$
Since the expected payoff of $ln$ for type $t_A$ is
$$2 \cdot \alpha_A(t_A) = 2 \cdot t_A(1 - F(y)),$$
the incentive conditions for $x$ are:

$$0 < x < 1 \quad \Rightarrow \quad 2x(1 - F(y)) = 1$$

$$x = 1 \quad \Rightarrow \quad 2(1 - F(y)) \leq 1$$
Role-dependent guilt aversion
Incomplete information: B’s incentive conditions

Taking into account that all types \((\theta_B, e_B)\) with \(e_B \leq y\) or \(\theta_B = \theta^L\) defect, and that the conditional expected payoff of defection for type \((\theta_B, e_B)\) minus the payoff of cooperation is

\[
E(\theta_B, e_B)[u_B | ln, D] - 2 = 2 - \theta_B \cdot 2\beta_B(\theta_B, e_B),
\]

we obtain the incentive conditions for threshold \(y\):

\[
0 < y < 1 \quad \Rightarrow \quad \begin{cases} 
(1 - F(y)) \frac{(1+x)}{2} = \frac{1}{\theta^H}, & \text{if } x > y, \\
(1 - F(y)) \frac{(1-\epsilon)(y^2-x^2) + \epsilon y (1-x^2)}{2[(1-\epsilon)(y-x) + \epsilon y (1-x)]} = \frac{1}{\theta^H}, & \text{if } x \leq y,
\end{cases}
\]

\[
y = 1 \quad \Rightarrow \quad (1 - F(y)) \frac{(1+x)}{2} < \frac{1}{\theta^H}.
\]
Role-dependent guilt aversion
Incomplete information: Equilibria

If we assume $F(y) = y$, from the above incentive conditions we derive the following equilibria:

- for $\theta^H \in \left[\frac{4}{3}, 2\right]$:
  \[
  \begin{aligned}
  x &= \frac{\theta^H}{4-\theta^H} \\
  y &= \frac{3}{2} - \frac{2}{\theta^H}
  \end{aligned}
  \]

- for $\theta^H > 2$:
  \[
  \begin{aligned}
  x &= 1 \\
  y &= \frac{\theta^H - 1}{\theta^H}
  \end{aligned}
  \]

- for any value of $\theta^H$ we find the equilibrium ($Out, D$), that is:
  \[
  \begin{aligned}
  x &= 1 \\
  y &= 1
  \end{aligned}
  \]
Role-independent guilt aversion

Here we have:

- $\theta_i$ is the guilt parameter of player $i$;
- $\alpha_A = \Pr_A[C]$ is $A$’s first order exogenous belief;
- $\alpha_B = E_B[ln]$ is $B$’s first order exogenous belief.
Role-independent guilt aversion

\[ A \quad \rightarrow \quad B \]

\[ (1 - \theta_A \cdot D_B(Out, \alpha_B, e_B)) \]

\[ S \quad \rightarrow \quad \begin{pmatrix} 2 \\ 2 \end{pmatrix} \]

B’s disappointment is:

\[ D_B(Out, \alpha_B, s_B) = \max \{ 0, m_B(\alpha_B, s_B) - m_B(Out) \} \]

\[ = \begin{cases} \alpha_B & \text{if } s_B = C \\ 3\alpha_B & \text{if } s_B = D \end{cases} \]
There are three possible equilibria in pure strategies:

- \((\text{Out}, D)\) is an equilibrium for every pair \((\theta_A, \theta_B)\);
- \((\text{In}, D)\) is an equilibrium when \(\theta_A \geq \frac{1}{3}\);
- \((\text{In}, C)\) is an equilibrium when \(\theta_B \geq 1\).
Now $T_A = T_B = \{\theta^L, \theta^H\} \times [0, 1]$.

We assume that it is correctly and commonly believed that $\theta_i$ and $e_i$ are independent and that the marginal distribution of $e_i$ is given by cdf $F : [0, 1] \rightarrow [0, 1]$, a continuous and strictly increasing function.

$$\forall e_i \in [0, 1], \forall x, \tau_i(e_i)[\theta_{-i} = \theta^H, e_{-i} \leq x] = e_i F(x).$$
Role-independent guilt aversion
Incomplete information: optimal strategies

- $\sigma_B(\theta^L, e_B) = D$; given $\theta^L = 0$, if $B$’s payoff type is $\theta^L$ he has standard preferences and he maximizes his monetary payoff.

- $\sigma_A(\theta^L, e_A) = ln$ if and only if $e_A > x$; $A$ is not guilt averse, therefore the higher is $t_A$, the higher is the probability that $B$ is guilt averse, the higher are $B$’s incentives to cooperate and therefore $A$’s incentives to choose $ln$. 
Role-independent guilt aversion
Incomplete information: optimal strategies

\[ \sigma_B(\theta^H, e_B) = C \text{ if and only if } e_B < y; \text{ the more } B \text{ believes that } A \text{ is guilt averse (the higher } e_B), \text{ the lower } E_B[\alpha_A \mid ln], \text{ the lower } B\text{'s incentives to cooperate.} \]

\[ \sigma_A(\theta^H, e_A)\text{: mathematically it can be both increasing or decreasing; we conjecture that it has one of the following shapes: either } \sigma_A(\theta^H, e_A) = ln \text{ if and only if } e_A > z \text{ or } \sigma_A(\theta^H, e_A) = ln. \]
Discussion

Almost all trust games experimentally analyzed in the lab and in the field are games with incomplete information about players’ “psychological” types.

We show how to analyze trust games with incomplete information when there are belief-dependent preferences.

Comparing complete information and incomplete information analysis, we can observe that in the incomplete information case we have less polarized behavior and beliefs.

This is true both with role-dependent and with role-independent belief-dependent preferences.