Guilt in Games

Pierpaolo **Battigalli** (Bocconi University)

Martin Dufwenberg (University of Arizona)

January, 2007

Abstract

Presented at the session on "Beliefs in the Utility Function" of the ASSA meeting 2007.

We apply our *dynamic psychological games* theoretical framework to model two types other-regarding preferences related to guilt. In a game with *simple guilt* a player dislikes letting other players down. In a game with *guilt from blame* a player dislikes that other players think that he intended to let them down.

Beliefs in the utility function help explain "non-standard" behavior in

(1) decision problems: e.g. avoidance of anxiety (Caplin & Leahy, QJE01) or of disappointment $\longrightarrow$ dynamic consistency may be an issue

(2) interactive decision problems with other-regarding preferences: e.g. reciprocity (Rabin AER-93, Dufwenberg & Kirchsteiger GEB-04), conformity/social respect (Bernheim JPE-94, Dufwenberg & Lundholm EJ-01), concern of experts for the emotions of others (Caplin&Leahy EJ-04) $\longrightarrow$ endogenous higher-order beliefs are crucial (Geanakoplos *et al.* GEB-89).

Our paper "Dynamic Psychological Games" (DPG) provides a theoretical framework covering (1) and (2), but the main focus is (2). DPG argues that conditional higher-order beliefs, beliefs of others and plans of action should be arguments of the utility function (on top of material consequences). Here we apply DPG to analyze guilt in games.

Psychologists: "if people feel guilt for hurting their partners ... and for failing to live up to their expectations, they will alter their behavior (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship" (Baumeister *et al*, *Psychological Bulletin*, 1994).

A quite substantial body of experimental evidence on trust games supports this view (Dufwenberg & Gneezy, GEB-00, Bacharach *et al.* mimeo-02, Guerra & Zizzo JEBO-04, Charness & Dufwenberg, Econometrica-06, Attanasi & Nagel mimeo-06).

Building on previous work on trust games (Dufwenberg JEBO-02), we model guilt in two ways:

Say that player i *lets* player j *down* (disappoints j) if as a result of i's choice of strategy j gets a lower monetary payoff than j initally expected.

Simple guilt

Player i's guilt may depend on how much i believes he lets j down

Guilt from blame

Player i's guilt may also depend on how much i believes j believes i intended to let j down.

1 Formalism

Finite extensive game forms:

- *material* consequences of terminal nodes: $m_i = \mathbf{m}_i(z) \in \mathbb{R}$ (*money*)

- player i's *information* $h = H_i(t) \in H_i$ specified *at every node* t (even if i is *not* active), including root t^0 and terminal nodes z, $h^0 := \{t^0\} \in H_i$, assume *perfect recall and observation of* $\mathbf{m}_i(z)$ (default assumption: coarsest terminal info.)

- *chance*=fictitious player with exogenous behavior strategy σ_c

- derive *outcome function* $z = \mathbf{z}(s)$ (s=strategy profile)

- *behavior strategies* $\sigma_i = (\sigma_i(\cdot|h))_{h \in H_i}$, derive $\mathsf{Pr}_{\sigma_i}(s_i|h)$

Hierarchical conditional beliefs:

1st-order cond. belief system $\alpha_i = (\alpha_i(\cdot|h))_{h \in H_i}$, $\alpha_i(\cdot|h) \in \Delta(S_{-i})$

2nd order: $\beta_i = (\beta_i(h))_{h \in H_i}$, $\beta_i(h)$=point belief of i given h about $\alpha_{-i} = (\alpha_j)_{j \neq i}$

3rd order: $\gamma_i = (\gamma_i(h))_{h \in H_i}$,, $\gamma_i(h)$=point belief of i given h about $\beta_{-i} = (\beta_j)_{j \neq i}$

...

(it is "common knowledge" that Bayes rule holds)

$D_j(z, s_j, \alpha_j) := \max\{0, \mathsf{E}_{s_j,\alpha_j}[m_j|h^0] - \mathbf{m}_j(z)\}$
=how much j is let down (*disappointed*) at z [s_j consistent w/ z]

$G_{ij}(z, s_{-i}, \alpha_j) := D_j(z, s_j, \alpha_j) - \min_{s_i} D_j(\mathbf{z}(s_i, s_{-i}), s_j, \alpha_j)$
=how much i lets j down at z [s_{-j} consistent w/ z]

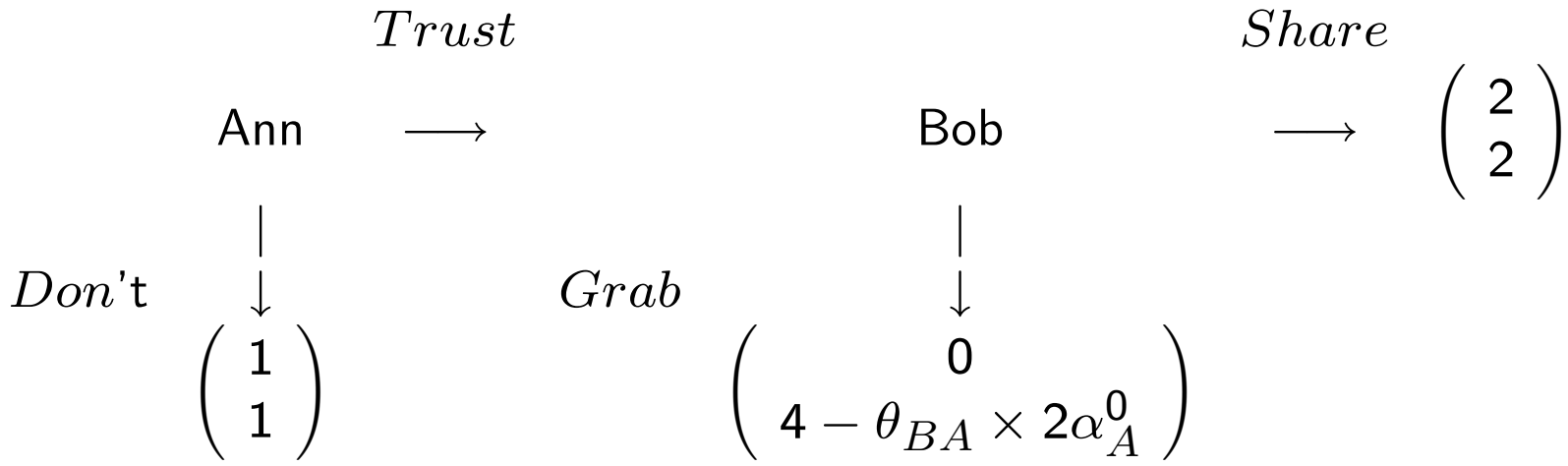
Psychological payoff function with *simple guilt*

$$u_i^{SG}(z, s_{-i}, \alpha_{-i}) = \mathbf{m}_i(z) - \sum_{j \neq i} \theta_{ij} G_{ij}(z, s_{-i}, \alpha_j),\ s_{-i} \in S_{-i}(z),$$

$\theta_{ij} \geq 0$ is i's guilt sensitivity toward j.

Of course, u_i^{SG} is *not "experienced" utility*, it just represents i's *motivations*: $\max_{s_i} \mathsf{E}_{s_i,\alpha_i,\beta_i}[u_i^{SG}|h]$

Example 0



Trust Game with simple guilt aversion, $\theta_{AB} = 0$, $\alpha_A^0 := \alpha_A(Share|h^0)$

$2\alpha_A^0$=Ann's disappointment at $z = (Trust, Grab)$
$4 - \theta_{AB} \times 2\alpha_B^0$=Bob's "state-dependent" utility at $z = (Trust, Grab)$

Bob *Share*s if $2 > \mathsf{E}_{\beta_B}[4 - 2\theta_{AB}\alpha_B^0|Trust]$

$G^0_{ij}(s_i, \alpha_i, \beta_i) := \mathsf{E}_{s_i,\alpha_i,\beta_i}[G_{ij}|h^0]$
=how much i (initially) intends to let j down

$\Rightarrow \mathsf{E}_{\alpha_j,\beta_j,\gamma_j}[G^0_{ij}|H_j(z)]$=how much j would "blame" i if z were reached

Psychological payoff function with *guilt from blame*

$$u_i^{GB}(z, \alpha_{-i}, \beta_{-i}, \gamma_{-i}) = \mathbf{m}_i(z) - \sum_{j \neq i} \theta_{ij} \mathsf{E}_{\alpha_j,\beta_j,\gamma_j}[G^0_{ij}|H_j(z)]$$

Again, u_i^{GB} is not an "experienced" utility...

Sequential Equilibrium (SE)

Fix psychological utility functions u_i ($u_i = u_i^{SG}, u_i^{GB}$, or other)

Assessment $(\sigma, \alpha, \beta, ...) = (\sigma_i, \alpha_i, \beta_i, ...)_{i \in N}$ is *consistent if* $\exists \sigma^k \to \sigma$ s.t. $\forall i$
(a) each $\alpha_i(\cdot|h)$ is derived from $\lim_{k \to \infty} \sigma^k$
(b) each higher-order belief $\beta_i(h), \gamma_i(h), ...$ is correct

$(\sigma, \alpha, \beta, ...)$ is a *SE* if it is consistent and

$$\forall i, \forall h \in H_i, \forall s_i^*, \Pr_{\sigma_i}(s_i^*|h) > 0 \Rightarrow s_i^* \in \arg \max_{s_i \in S_i(h)} \mathsf{E}_{s_i, \alpha_i, \beta_i, ...}[u_i|h]$$

Comment on SE: in equilibrium, players never change their mind about co-players' beliefs; they are only forced to update beliefs about co-players' strategies.

This is a disturbing feature of SE. We think that other solution concepts should be explored and applied in the context of psychological games, e.g. solution concepts allowing for forward induction reasoning (see DPG).

2 Some results and examples

Observation 1. *In any* 2*-person,* simultaneous-*move game form* without chance *moves, the pure stategy SE assessments of the psychological games with simple guilt and guilt from blame coincide* (same θ_{ij}s)*.*

Intuition: by consistency and perfect recall, i predicts that j will find out how much i lets j down.

The assumptions are *tight*. In other games, a SE with simple guilt need not be a SE with guilt from blame, and vice versa.

Observation 1 does not extend to n-person game forms:

$Ann \backslash^{Bob}$	*abstain*	*steal*
abstain	0,0,2	0,2,**0**
steal	2,0,**0**	1,1,**0**

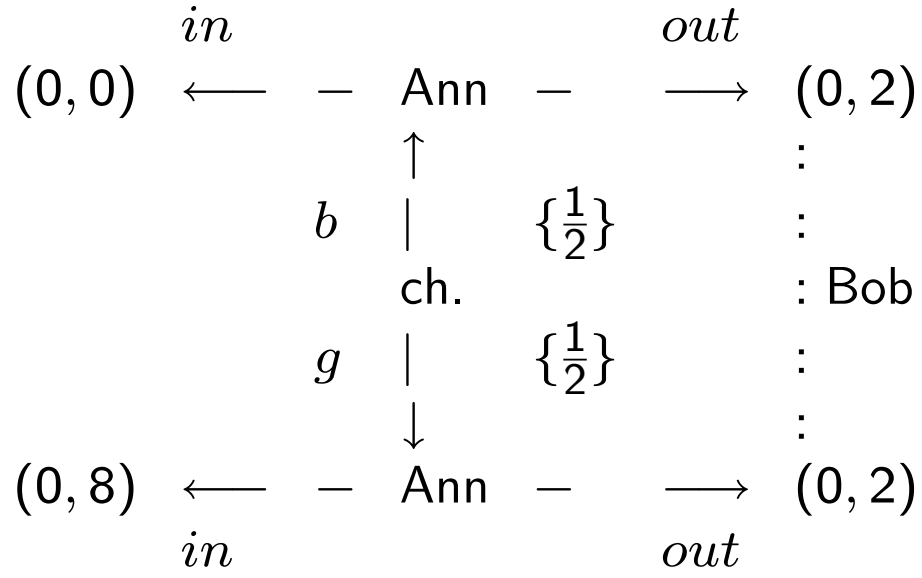
A 3-person game form

Example 1. Cleo is *passive*. $1 < \theta_{AC} = \theta_{BC} < 2 \Rightarrow$ (*abstain*, *abstain*) is (part of) a SE with simple guilt but not a SE with guilt from blame.

Intuition. *Simple guilt:* only the initial beliefs of Cleo matter: i has no incentive to deviate as $2 - 2\theta_{iC} < 0$.
Guilt from blame: if i deviates from (*abstain*, *abstain*) and steals, since *Cleo observes only her material payoff of \$0* she cannot be sure whom to blame. This partially shelters i from some pangs of blamed guilt $\Rightarrow$ higher incentive to deviate (the threshold 2 comes from consistency of beliefs).

Observation 1 does not extend to game forms with chance moves:

Example 2. (in, in) is (part of) a SE with guilt from blame, not with simple guilt (intuition: not trivial)



Game form with asymmetric info. about chance move

Intuition. (in, in) cannot be a SE strategy w/ simple guilt: Bob's expected payoff is 4, if Ann observes b, she minimizes Bob's disappointment by choosing out ($D_B = 4 - 2$) instead of in ($D_B = 4 - 0$).

But (in, in) is part of a SE w/ guilt from blame where Bob, upon observing Out, would believe that Ann plays [in if b,out if g] thus blaming Ann in the amount $[(\frac{1}{2}(4 - 0) + \frac{1}{2}(4 - 2) - 1] = 2$ (where 1 is Bob's unavoidable disappointment if he initially expects (in, in)).

Observation 2. *In any* simultaneous *game form* without chance *moves all the pure strategy SE assessments of the material payoff game are also SE of the psychological games with simple guilt and guilt from blame* (whatever θ_{ij}s)

Intuition. Without deviations every j gets exactly what he expected. If i deviates from material equil. he can only decrease (weakly) his material payoff and increase (weakly) the disappointment of every $j \neq i$.

The assumptions are *tight*.

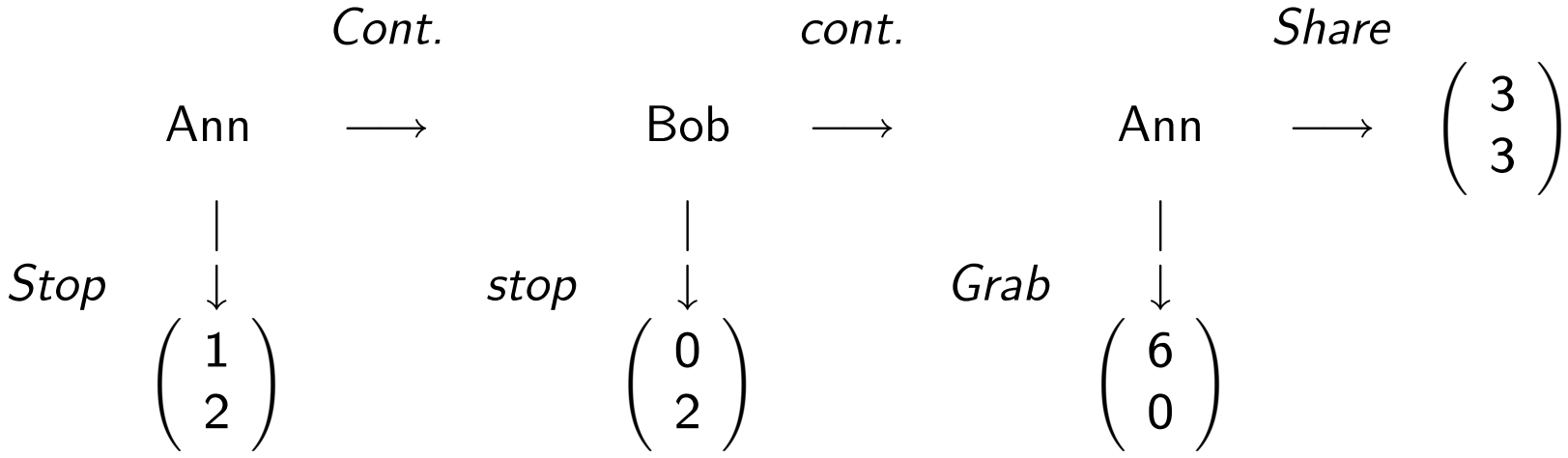
Observation 2 does not extend to game forms with chance moves:

Example 3. Bob is *passive*, the game form is:

$Ann \backslash^{chance}$	*g* $[\frac{1}{2}]$	*b* $[\frac{1}{2}]$
up	ε, 2	ε, 0
down	0,1	0,1

up is part of a material SE, but not part of a SE with guilt for moderate values of θ_{AB}: given up with 50% chance Bob is disappointed.

Observation 2 does not extend to sequential game forms:



A game form with perfect information

Example 4. [(*Stop*,*Grab*),*stop*] is the unique material SE, but it is not SE with simple guilt if θ_{AB} is high enough: if Ann *Share*d at (*Cont.*,*cont.*) she would spare Bob a disappointment of 2.

On the other hand, [(*Stop*,*Grab*),*stop*] is a SE with guilt from blame. Key: Bob sees that if Ann *Cont*inues she does not expect to disappoint him.

$\Rightarrow$ *Observation 1 does not extend to sequential game forms.*

Observation 3. *In any game form* without chance *moves every strictly efficient path (terminal node) of the material payoff game can be supported as a SE of the game with simple guilt for sufficiently high* θ_{ij}s.

Intuition. Fix strictly (materially) efficient z^*. Define threshold θ^* s.t. no incentive to deviate from the z^*-path for $\theta_{ij} > \theta^*$ all i,$j \neq i$. Fix such θ_{ij}s. Use a "trembling hand+fixed point" argument to show that there is a consistent assessment that yields z^* with prob. 1 and s.t. there is no incentive to deviate off the z^*-path.

The no-chance move assumption is necessary. The observation can be extended to guilt from blame if there are "observable deviators" at the end of the game.

3 Public good games with linear technology

n players

Possible contributions: $A_i = S_i = \{0, 1, ..., K\}$

B=increase in material payoff of every player for each contributed dollar:

$$\mathbf{m}_i(a_1, ..., a_n) = B\left(\sum_{j=1}^{n} a_j\right) - a_i, \quad \left(\frac{1}{n} < B < 1\right)$$

$\theta_{ij} = \theta$ common guilt sensitivity

Simple guilt: *If* $\theta B(n-1) \geq (1-B)$ *then every strategy profile is part of a SE, if* $\theta B(n-1) < (1-B)$ *then the only SE strategy profile is* $(0, ..., 0)$.

Intuition:

$\theta B(n-1)$=guilt from withholding one dollar

$(1-B)$=net material benefit from withholding one dollar

Guilt from blame: equilibria with positive contributions are more difficult to support; symmetric equilibria are more difficult to support than asymmtric equilibria.

k=no. of "donors" (i is a donor if $a_i > 0$)
Best chance to support k donors as equilibrium if a shortfall is equally blamed on all the supposed donors (the "least blamed" has highest incentive to deviate)
$(k-1)B\frac{1}{k-1} = B$=guilt blamed from other donors (if $k > 1$), for unilaterally withholding one dollar
$B(n-k)/k$=guilt blamed from non-donors

There is a SE with $k \geq 1$ *donors if and only if*
$\theta B[I_{k>1}(k) + (n-k)/k] \geq (1-B)$

4 Concluding remarks

Possible economic applications: contribution to public goods (as we have seen), trust, contracts...

Importance of *conditional* higher order beliefs

Importance, in guilt from blame, of "*terminal information* structure" and conditional beliefs of order higher than 2nd.