Higher Order Beliefs and Emotions in Games: Theoretical Framework

Pierpaolo Battigalli

Bocconi University and IGIER

U. East Anglia, July 3-4 2017
Psychological Game Theory Summer School

Introduction

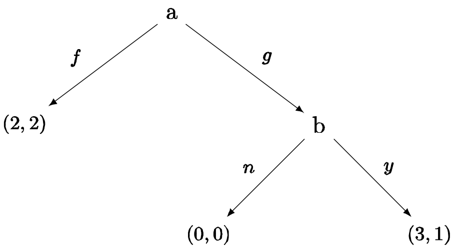
- Credible promises/threats and reliable communication are essential for cooperation.
- According to standard theory, credibility (incentive compatibility) is related to the value of future interaction.
- But often people keep their word and communicate truthfully even when this is not incentivized by future interactions.
- Emotions like guilt, anger, shame and pride can make people act against their selfish material interests in ways that are often (not always) beneficial to cooperation.
- Many emotions are triggered by beliefs, including beliefs about the beliefs of others (higher-order beliefs).
- Emotions affect behavior in two ways:
  - *direct*: induced action tendencies (e.g., frustration-aggression$\Rightarrow$carry out threats);
  - *indirect:* anticipated feelings (valence) modify material incentives (e.g., keep costly promises to avoid guilt).

- By letting psychological utility in games depend on beliefs we can model such phenomena.

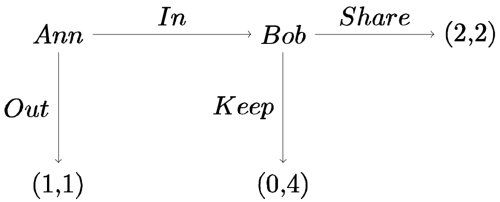
$$
\begin{array}{lcl}
\text{Beliefs} & \longrightarrow & \text{Emotions} \\
\uparrow & \searrow & \downarrow \\
\text{Experience} & \longleftarrow & \text{Actions}
\end{array}
$$

- We develop a methodology and illustrate it with some examples/applications.
- We adopt a *subjective* notion of *rationality:* (sequential) best reply to subjective beliefs, with psychological motivations. We do not consider bounded computational abilities, nor do we model how emotions can interfere with cognition.

Stylized dilemmas with implicit threat or promises



Ultimatum Minigame



Trust Minigame

Motivations & Examples

The following is *in*consistent with standard social preferences (e.g., inequity or lying aversion), but consistent with our framework and model(s):

- **Psychology**:
  - desire to live up to others' expectations to avoid guilt feelings (Baumeister *et al.*, 1994; Tangney, 1995);
  - frustration-aggression hypothesis (Dollard *et al.*, 1939; Frijda, 1993);
  - moral behavior to avoid the feeling of shame (Tangney, 1995).

Motivations & Examples (continue)

- **Facts (casual evidence, empirics):**
  - Non-returning customers give tips.
  - Low offers are often rejected leaving money on the table.
  - Unexpected losses by home football/soccer teams are associated with increased domestic violence (Card & Dahl, 2011) or violent crime (Munyo & Rossi 2013).
- **Facts (experimental):**
  - *Trust Minigame:* correlation between sharing and with 2^{nd}-order beliefs of sharing; game-form invariant treatments affect beliefs and behavior (Charness & Dufwenberg, 2006; Tadelis, 2011; Attanasi *et al.* 2013).
  - *Ultimatum Game:* Rejections correlate with (manipulated) initially expected offers (Sanfey, 2009; Xiang *et al.*, 2013, with fMRI).
  - *Lying/truth-telling* is not categorical (Fischbacher & Föllmi-Heusi, 2008), it depends on the payoffs of receivers (Gneezy, 2005; Battigalli *et al.* 2013) and on exposure to passive observers (Gneezy *et al.*, 2016).

Setting: game tree

We consider *finite, multistage game forms with observable actions and incomplete information* (easy cases: leader-follower and dictator games).
Game tree $\left(I, \left(A_i, \mathcal{A}_i\left(\cdot\right)\right)_{i\in I}\right)$ where:

- **Players**: $i \in I$.
- **Actions, action profiles**: $a_i \in A_i$ finite, $wait \in A_i$ (trick), $a = \left(a_i\right)_{i\in I} \in \times_{i\in I} A_i := A$.
- **Histories**: $\varnothing$ =empty history, and $h = \left(a^k\right)_{k=1}^{t} \in A^t$, $a^t = \left(a_i^t\right)_{i\in I}$, $t = 1, 2, ..., T$ ($h \preceq h'$, "prefix" relation).

Setting: game tree

- **Feasible actions and profiles:** $h \mapsto \mathcal{A}_i(h) \subseteq A_i$, $\mathcal{A}(h) := \times_{i \in I} \mathcal{A}_i(h) \subseteq A$; $\mathcal{A}_i(h) = \{w\}$ if i **inactive** at h; $\mathcal{A}(h) = \varnothing$ (empty set) if **game over.**
- **Feasible histories:** $\varnothing$ (empty hist.=root of tree) is feasible, $h = \left(a^k\right)_{k=1}^{t}$ is feasible if $a^1 \in \mathcal{A}(\varnothing)$ and $a^{k+1} \in \mathcal{A}\left(a^1, ..., a^k\right)$, $k = 1, ..., t-1$.
- **Nonterminal and terminal:** $H := \{h : h \text{ feasible}\}$; terminal (play paths): $Z := \{h \in H : \mathcal{A}(h) = \varnothing\}$; nonterminal: $H \backslash Z$.
- **Personal histories of i:** $H_i := H \cup \{(h, a_i) : h \in H \backslash Z, a_i \in \mathcal{A}_i(h)\}$ (as soon as i chooses a_i at h he knows that $h_i = (h, a_i)$ has occurred; important later). Prefix relation $\preceq$ easily generalized for H_i, for all $i \in I$.
- **Terminal continuations of h_i:** $Z(h_i) = \{z \in Z : h_i \preceq z\}$

Setting: game form

Game form $\left(I, \left(A_i, \mathcal{A}_i\left(\cdot\right), \Theta_i, \pi_i\left(\cdot,\cdot\right)\right)_{i\in I}\right)$: add to the game tree information types and the material payoffs/outcome functions:

- **Type** of i: $\theta_i \in \Theta_i$ exogenous trait (finite only for simplicity), private information of i (ability, degree of altruism, aversion to lying, aversion to guilt, ...); profiles of types $\theta \in \Theta = \times_{i\in I}\Theta_i$, $\theta_{-i} \in \Theta_{-i} = \times_{j\neq i}\Theta_j$.
- "**Monetary" payoffs/outcomes** (material consequences) $(z,\theta) \mapsto \pi(z,\theta) = (\pi_i(z,\theta))_{i\in I} \in \mathbb{R}^I$ (π_i is *not* the *utility* of i).

(Conditional) Beliefs

Beliefs of the first and second order are *conditional probability systems* (CPS's) about paths (including own behavior) and types of others that satisfy obvious *independence* restrictions, and possibly other restrictions deemed plausible in applications (symmetry, positivity, known prob. of chance moves,...). First-order conditional beliefs concern behavior (paths) and information types, and satisfy natural properties relating beliefs conditional on different (personal) histories:

- **First-order beliefs of** i: Consider set of CPSs $B_i^1 \subseteq \left[\Delta\left(Z \times \Theta_{-i}\right)\right]^{H_i}$, where $\beta_i^1 = \left(\beta_i^1\left(\cdot|h_i\right)\right)_{h_i \in H_i} \in B_i^1$ only if (with obvious abbreviations for marg. and cond. probabilities): for *all* $h_i, h_i' \in H_i \backslash Z$, $z \in Z$ with $h_i \preceq h_i' \prec z$, $h \in H \backslash Z$, $a \in \mathcal{A}\left(h\right)$, $a_i' \in \mathcal{A}_i\left(h\right)$, $\theta_{-i} \in \Theta_{-i}$,
  - *chain rule* (CR^1)
    $\beta_i^1\left(z, \theta_{-i}|h_i\right) = \beta_i^1\left(z, \theta_{-i}|h_i'\right)\beta_i^1\left(h_i'|h_i\right) = \beta_i^1\left(\theta_{-i}|z\right)\beta_i^1\left(z|h_i\right)$
    (note: $\beta_i^1\left(h_i'|h_i\right) = \beta_i^1\left(Z\left(h_i'\right)|h_i\right)$);
  - *own-action independence* (OAI^1):
    $\beta_i^1\left(a_{-i}, \theta_{-i}|h\right) = \beta_i^1\left(a_{-i}, \theta_{-i}|h, a_i'\right)$ (note: $\left(h, a_i'\right) \in H_i$; beliefs about types and simultaneous actions of others are independent of own action).

(Conditional) Beliefs

Second-order conditional beliefs concern both behavior-types *and* the first-order CPS's of others:

- **Second-order beliefs**: Consider set of CPS's

$$B_i^2 \subseteq \left[\Delta\left(Z \times \Theta_{-i} \times B_{-i}^1\right)\right]^{H_i}$$

  where $\beta_i^2 = (\beta_i^2(\cdot|h_i))_{h_i \in H_i} \in B_i^2$ only if it satisfies CR and OAI restrictions similar to those for first-order CPS's: for *all* $h_i, h_i' \in H_i \backslash Z$ with $h_i \preceq h_i'$, $z \in Z$, $h \in H \backslash Z$, $a \in \mathcal{A}(h)$, $a_i' \in \mathcal{A}_i(h)$, $\theta_{-i} \in \Theta_{-i}$, (Borel) $E_{-i} \subseteq B_{-i}^1$,

  - *chain rule* (CR^2) $\beta_i^2(\{(z,\theta_{-i})\} \times E_{-i}|h_i) = \beta_i^2(\{(z,\theta_{-i})\} \times E_{-i}|h_i')\,\beta_i^2(h_i'|h_i) = \beta_i^2(\{\theta_{-i}\} \times E_{-i}|z)\,\beta_i^2(z|h_i)$,
  - *own-action independence* (OAI^2) $\beta_i^2(\{(a_{-i},\theta_{-i})\} \times E_{-i}|h) = \beta_i^2(\{(a_{-i},\theta_{-i})\} \times E_{-i}|h, a_i')$ (beliefs about **simultaneous** actions, types and 1st-ord. beliefs of others are independent of own action).

- **Result (technical):** *For any finite game form and any player $i \in I$, B_i^1 and B_i^2 are compact metrizable (hence Polish) topological spaces.*

Comments and notation about beliefs

- Interpretation of (1^{st}-order) beliefs about one's own behavior: plan of the player, that is,
  - by OAI, $\beta_i^1\left(\left(a_i, a_{-i}\right)|h\right) = \beta_i^1\left(a_i|h\right)\beta_i^1\left(a_{-i}|h\right)$,
  - $\beta_{i,i} = \left(\beta_i^1(a_i|h)\right)_{h\in H\setminus Z, a_i\in\mathcal{A}_i(h)}$ is the **plan** of i.
- 1^{st}-order beliefs can be derived from 2^{nd}-order beliefs by marginalization (conditional on each h): e.g., $\beta_i^1(\theta_{-i}) = \beta_i^2\left(\{\theta_{-i}\}\times B_{-i}^1\right)$; we then write

$$\beta_i^1 = \mathrm{marg}_{Z\times\Theta_{-i}}\beta_i^2$$

$$\beta_i \in B_i = \left\{\left(\beta_i^1, \beta_i^2\right)\in B_i^1\times B_i^2 : \beta_i^1 = \mathrm{marg}_{Z\times\Theta_{-i}}\beta_i^2\right\}$$

$$B_i\left(\bar{\beta}_i^1\right) = \left\{\left(\beta_i^1, \beta_i^2\right)\in B_i^2 : \beta_i^1 = \bar{\beta}_i^1\right\}$$

(B_i is isomorphic to B_i^2; $B_i\left(\bar{\beta}_i^1\right)$ is isomorphic to the section at $\bar{\beta}_i^1$ of B_i: it is the set of β_i consistent with $\bar{\beta}_i^1$).

Expectations

- For all $h \in H$, θ_i, β_i, and (measurable) function $\widetilde{x} : Z \times \Theta \times B^1_{-i} \to \mathbb{R}$ (random variable) we can compute the expectation of $\widetilde{x}$ *conditional* on h (or $h_i = (h, a_i)$), *given* (θ_i, β_i)

$$\mathbb{E}\left[\widetilde{x}|h;\theta_i,\beta_i\right] = \int \widetilde{x}\left(z,\theta_i,\theta_{-i},\beta^1_{-i}\right)\beta^2_i\left(\mathrm{d}z,\mathrm{d}\theta_{-i},\mathrm{d}\beta^1_{-i}|h\right).$$

- For a belief-independent r.v. (e.g., $\widetilde{x} = \pi_j$)

$$\mathbb{E}\left[\widetilde{x}|h;\theta_i,\beta_i\right] = \sum_{z,\theta_{-i}} \widetilde{x}\left(z,\theta_i,\theta_{-i}\right)\beta^1_i\left(z,\theta_{-i}|h\right).$$

Psychological preferences

"Experience utility"

We assume that the "value" or "**experience utility**" of a path z for i depends on (some aspects of) $\theta = \left(\theta_j\right)_{j\in I}$ and $\beta^1 = \left(\beta^1_j\right)_{j\in I}$:

$$v_i : Z \times \Theta \times B^1 \rightarrow \mathbb{R}$$

Examples ($[x]^+ = \max\{x, 0\}$):

- *selfish risk neutral:* $v_i = \pi_i$;
- *guilt/pity aversion*: $v_i\left(z, \theta, \beta^1\right) = \pi_i\left(z\right) - \theta_i \cdot \left[\mathbb{E}\left[\pi_{-i}; \beta^1_{-i}\right] - \pi_{-i}(z)\right]^+$ (no own-plan dep.);

Psychological preferences

- *disappointment aversion*: $v_i\left(z,\theta,\beta^1\right)=\pi_i\left(z\right)-\theta_i\cdot\left[\mathbb{E}\left[\pi_i;\beta_i^1\right]-\pi_i(z)\right]^+$ (own-plan dep., see also loss aversion with ref. point=lagged expect. as in Koszegi & Rabin);
- *pride/shame, ...* : $v_i\left(z,\theta,\beta^1\right)=\pi_i\left(z\right)+\theta_i^r\cdot\rho\left(\mathbb{E}\left[\widetilde{\theta}_i^g|z;\beta_{-i}^1\right]\right)$, $\rho'>0$, $\theta_i=\left(\theta_i^g,\theta_i^r\right)$, θ_i^g =goodness, $\mathbb{E}\left[\widetilde{\theta}_i^g|z;\beta_{-i}^1\right]$=reputation of i according to $-i$, θ_i^r=reputational concern (non-instrumental).

Psychological preferences

"Decision utility"

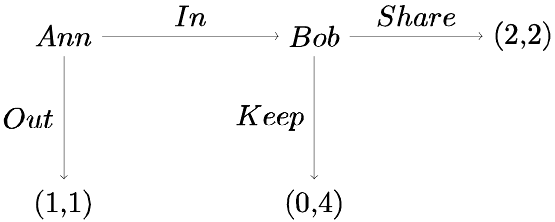
The "utility" of an action a_i given non-terminal history h is what drives the decision of the player i active at h. It may just be the expected value of v_i conditional on h given (θ_i, β_i), or a modification of such expectation that captures the action tendencies of an emotion, e.g., desire to harm given anger. Assuming additive separability,

$$u_i\left(h, a_i; \theta_i, \beta_i\right) = \mathbb{E}\left[v_i | h, a_i; \theta_i, \beta_i\right] + \mathbb{E}\left[\delta_i\left(h, \theta_i, \beta_i^1, \widetilde{\pi}_{-i}, \widetilde{\theta}_{-i}, \widetilde{\beta}_{-i}^1\right) | h, a_i; \beta_i\right]$$

Examples: anger of Bob (when frustrated) from blaming Ann's behavior or intentions (Battigalli *et al.*, 2015); it increases the decision utility of rejecting the greedy offer in the ultimatum game when Bob expected a fair offer, because of the harm inflicted on Ann.

Note: If there is own-plan dependence of experience utility, or if decision utility is different from the conditional expectation of experience utility, then maximization of decision utility may differ from what i would like to *covertly commit to* ex ante (dynamic inconsistency of preferences).

Trust Minigame with Guilt Aversion



Trust Minigame

- Ann is commonly known to be selfish: $u_1(In; \beta_1) = 2\beta_1^1(Share|In)$, *In* only if $\beta_1^1(Share|In) \geq 1/2$.
- Bob is guilt averse:
  $u_2\left(In, Keep; \theta_2, \beta_2\right) = 4 - 2\theta_2 \mathbb{E}\left[2\widetilde{\beta}_1^1(Share|In)|In; \beta_2^2\right]$.

“Psychological” equilibrium?

- Geanakoplos *et al.* 1989 (GPS), with indirect methods, and Battigalli & Dufwenberg 2009 (BD), with direct methods, define adapted notions of “psychological” Nash and sequential equilibrium.
- One can show that it is enough to apply Harsanyi's method of Bayesian games, which complements information types θ_i with “epistemic types” e_i to obtain “Harsanyi types” $t_i = (\theta_i, e_i)$ which implicitly determine exogenous hierarchies of beliefs, and then look at Bayesian equilibrium decision functions $t_i \mapsto \sigma_i(t_i)$. This generates endogenous hierarchies of beliefs in equilibrium. When $T_i \cong \Theta_i$ we get back the equilibria defined “ad hoc” by GPS and BD (see Attanasi, Battigalli, & Manzoni, 2016).
- Problem of this “rational-expectations” equilibrium approach: NO FOUNDATIONS after several decades since its introduction in GT and Theoretical Economics!

Rationalizability

- **Rationality (subjective!):** i is rational if he plans rationally given his subjective beliefs (one-shot dev. property) and his action on path is one he planned to choose with positive probability.
- **Strong belief** (informal)**:** i strongly believes an event E if he is certain of E conditional on each $h \in H$ consistent with E.
- k-**rationalizability** ($k \in \mathbb{N}$): set of tuples $\left(z, \theta, \beta^1\right)$ consistent with *rationality and* $(k-1)$*-mutual strong belief in rationality* (see Battigalli, Corrao & Sanna, 2017); *note:* we look at possible values of the variables that affect v_i and δ_i, because the relevant expectations are taken with respect to beliefs about such variables [with non-belief-dependent preferences we look at (z, θ)].

Rationalizability (continues)

Rationality: Recall, plan of i at h: $\beta_{i,i}(\cdot|h) = \mathrm{marg}_{\mathcal{A}_i(h)}\beta_i(\cdot|h)$ $(h \in H \backslash Z)$. Belief β_i satisfies **rational planning** if, for each h where i is active

$$\beta_{i,i}(a_i|h) > 0 \Rightarrow \arg\max_{a_i \in \mathcal{A}_i(h)} u_i(h, a_i; \theta_i, \beta_i).$$

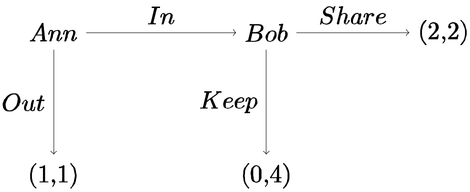
Given "prediction set" $P \subseteq Z \times \Theta \times B^1$ and type-belief $\left(\bar{\theta}_i, \bar{\beta}_i^1\right)$, $P_{\bar{\theta}_i, \bar{\beta}_i^1}$ is the **section** of P at $\left(\bar{\theta}_i, \bar{\beta}_i^1\right)$:

- $P_{\bar{\theta}_i, \bar{\beta}_i^1} = \left\{\left(z, \theta, \beta^1\right) \in P : \theta_i = \bar{\theta}_i, \beta_i^1 = \bar{\beta}_i^1\right\}$;
- similarly, $P_h = \left\{\left(z, \theta, \beta^1\right) \in P : h \preceq z\right\}$ for each $h \in H$.

k-**rationalizable set**: trivial prediction $P^0 = Z \times \Theta \times B^1$.
For $k > 0$, require rational planning, *strong belief* in (the section of) P^{k-1}, and on-path choice of planned actions:

$$P^k = \left\{ \left(\bar{z}, \bar{\theta}, \bar{\beta}^1\right) \in P^{k-1} : \begin{array}{l} \forall i, \exists \beta_i \in B_i\left(\bar{\beta}_i^1\right) \text{ s.t. rational planning} \\ \forall h \in H, P_h^{k-1} \neq \emptyset \Rightarrow \beta_i^2\left(P_{\bar{\theta}_i, \bar{\beta}_i^1}^{k-1} | h\right) = 1 \\ \forall \bar{h} \prec \bar{z}, \beta_{i,i}(\bar{a}_i | \bar{h}) > 0 \end{array} \right\}.$$

Rationalizability in Trust Game with Guilt Aversion



Trust Minigame

- Ann (pl. 1) commonly known to be selfish:
  $u_1(In; \beta_1) = 2\beta_1^1(Share|In)$, *In* only if $\beta_1^1(Share|In) \geq 1/2$.
- Bob (pl. 2) guilt averse:
  $u_2\left(In, Keep; \theta_2, \beta_2\right) = 4 - \theta_2 \mathbb{E}\left[2\widetilde{\beta}_1^1(Share|In)|In; \beta_2^2\right]$.
- Step 2: $\mathbb{E}\left[\widetilde{\beta}_1^1(Share|In)|In; \beta_2^2\right] \geq 1/2$ (strong belief in rationality)
  - $(In, Share)$ if $\beta_1^1(Share|In) > 1/2$ and $\theta_2 > 2$,
  - $(In, Keep)$ if $\beta_1^1(Share|In) > 1/2$ and $\theta_2 < 1$, *etc.*
- Step 3: *In* if $\beta_1^1\left(\widetilde{\theta}_2 > 2\right) > 1/2$, *Out* if $\beta_1^1\left(\widetilde{\theta}_2 < 1\right) > 1/2$.

Application: Guilt and Reciprocity in Trust Game

Attanasi, Battigalli & Nagel (2013, rev. 2017):

- Clever way to elicit θ_2 (sensitivity to both guilt and reciprocity) and make it "common knowledge" *via* disclosure of filled-in questionnaire.
- Correlation in strategies and beliefs induced *via* disclosure predicted (partially) by rationalizability, steps 1-3.
- With incomplete information (no disclosure), Steps 1-2 are still valid, Step 3 is silent: no further implication on top of step 2.
- Meaningful qualitative predictions across treatments, data move in the predicted direction.

Sequential Equilibrium

Assume for simplicity that $\theta = (\theta_i)_{i \in I}$ is common knowledge ($\forall i \in I$, $\Theta_i = \{\theta_i\}$) $\Rightarrow$ suppress θ.
Let

$$\sigma_i = (\sigma_i(\cdot|h))_{h \in H \setminus Z} \in \times_{h \in H \setminus Z} \Delta(\mathcal{A}_i(h))$$

denote a **behavioral strategy** of i.

Definition

A profile $(\sigma, \beta) = (\sigma_i, \beta_i)_{i \in I}$ is a **sequential equilibrium** if for all $i, j \in I$, for all $h \in H \setminus Z$ and $a = (a_i)_{i \in I} \in \mathcal{A}(h)$,

- **(agreement, independence & correct beliefs)**
  - $\beta_i^1(a|h) = \prod_{j \in I} \sigma_j(a_j|h)$,
  - $\mathrm{marg}_{B_{-i}^1} \beta_i^2(\cdot|h) = \delta_{\beta_{-i}^1}$ ($\delta_{\beta_{-i}^1}$ is the degenerate measure that assigns probability 1 to β_{-i}^1);
- **(rational planning)**

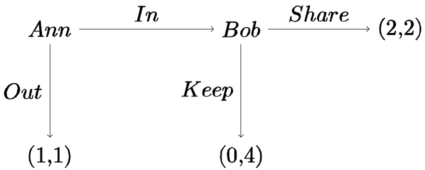
$$\sigma_i(a_i|h) > 0 \Rightarrow a_i \in \arg\max_{a_i' \in \mathcal{A}_i(h)} u_i(h, a_i; \beta_i).$$

Sequential Equilibrium: Comments

The psychological games framework requires higher-order (conditional) beliefs. Introducing higher-order beliefs allows to uncover (undesirable) conceptual features of Sequential Equilibrium (SE) in both psychological and standard games:

- SE is a notion of equilibrium in beliefs. 2^{nd}-order beliefs are always correct, hence they cannot change$\Rightarrow$*beliefs about plans of others never change!*
- Trembling-hand interpretation: Deviations from equilibrium plans/strategies are always interpreted as *unintentional mistakes*, no future mistakes ar ever expected.
- Consistency of behavior with plans σ yields possible paths $Z(\sigma) \subseteq Z$.
- If σ is interpreted as a profile of truly randomized strategies (at each h players spin roulette wheels to decide what to do), then it makes sense to look at the distribution $\zeta(\sigma) \in \Delta(Z)$ induced by σ.

Seq. Equil. in the Trust Game with Guilt Aversion



Trust Minigame

Suppose $\theta_2 > 2$ (commonly known), is $(Out, Keep)$ part of a SE? No!

- Bob is always certain that Ann's plan is *Out* $\Rightarrow$ after *In* Bob would still believe that Ann expected €1.
- $u_2(In, Keep; \beta_2) = 4 - \theta_2 \cdot (1 - 0) < 2 = u_2(In, Share; \beta_2)$.
- **Exercise**: Prove that
  - If $\theta_2 < 1$, unique SE outcome and unique rationalizable outcome is *Out*.
  - If $1 < \theta_2 < 2$, both *Out* and $(In, Share)$ are SE as well as rationalizable outcomes (nonexhaustive list).
  - If $\theta_2 > 2$, then the unique SE as well as the unique rationalizable outcome is $(In, Share)$.
- Compare with the analysis in BD (2009), why is it different?

Self-confirming equilibrium

- Characterization of stable distributions in population games played recurrently with feedback about outcomes (see Battigalli et al. 2015).
- Players are subjectively rational, their beliefs may be incorrect, but each player's beliefs are confirmed by what he observes (feedback), e.g., the frequencies of monetary payoffs given the chosen actions.
- Concept to be used to analyze stable pattern of behavior in empirical data, or stabilized behavior in experiments with repeated play and random matching in each period.
- Trust Minigame (feedback=own payoff):
  - a fraction of agents in pop. 1 stay *Out*, and their beliefs may be incorrect,
  - the complementary fraction of agents go *In*, and their beliefs of the conditional frequency of Share must be *correct*,
  - the fraction of agents in pop. 2 who Share (given In) is determined by the distribution of types and (with belief-dependent preferences) the distribution of 2^{nd}-order beliefs, which may be incorrect.

Conclusions

- These notes draw on joint work with G. Attanasi, G. Charness, R. Corrao, M. Dufwenberg, E. Manzoni, R. Nagel, F. Sanna, and A. Smith.
- We introduce a framework to model belief-dependent emotions in games.
- Several experiments are driven by such framework and yield interesting evidence supporting the assumption that preferences are belief-dependent.
- Standard equilibrium (even of the "psychological" variety) is inadequate to organize experimental results.
- Rationalizability (2, 3 steps) and self-confirming equilibrium should be used more often, as appropriate to the design, to obtain predictions about behavior and elicited beliefs, and to organize data.

References

- ATTANASI, G., P. BATTIGALLI, AND E. MANZONI (2016): "Incomplete Information Models of Guilt Aversion in the Trust Game," *Management Science*, 62, 648-667.
- ATTANASI G., P. BATTIGALLI AND R. NAGEL (2013): "Disclosure of Belief-Dependent Preferences in the Trust Game," IGIER Working Paper 506, Bocconi University.
- BATTIGALLI P. AND M. DUFWENBERG (2007): "Guilt in Games," *American Economic Review, Papers and Proceedings*, 97, 170-176.
- BATTIGALLI P. AND M. DUFWENBERG (2009): "Dynamic Psychological Games," *Journal of Economic Theory*, 144, 1-35.
- BATTIGALLI, P., S. CERREIA, F. MACCHERONI, AND M. MARINACCI (2015): "Self-Confirming Equilibrium and Model Uncertainty," *American Economic Review*, 105, 646-677.
- BATTIGALLI P., G. CHARNESS AND M. DUFWENBERG (2013): "Deception: The Role of Guilt," *Journal of Economic Behavior and Organization*, 93, 227-232.

Battigalli P., M. Dufwenberg and A. Smith (2015): "Frustration and Anger in Games," IGIER Working Paper 539, Bocconi University.

Battigalli, P., R. Corrao, and F. Sanna (2017): "Epistemic Game Theory Without Type Structures. An Application to Psychological Games," typescript, Bocconi University.

Bosman, R. and F. van Winden, (2002): "Emotional Hazard in a Power-to-Take Experiment," *Economic Journal*, 112, 147–169.

Bosman, R., M. Sutter, and F. van Winden (2005): "The Impact of Real Effort and Emotions in the Power-to-Take Game," *Journal of Economic Psychology*, 26, 407–429.

Card, D. and G. B. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *The Quarterly Journal of Economics*, 126, 103–143.

Charness G. and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74, 1579-1601.

Dollard, J., L. Doob, N. Miller, O. Mowrer, and R. Sears (1939): *Frustration and Aggression*. Yale University Press, New Haven, CT.

Fischbacher, U., and F. Follmi-Heusi (2008): "Lies in disguise: An experimental study on cheating," *Journal of the European Economic Association*, 11, 525–547.

Frijda, N. H. (1993): "The Place of Appraisal in Emotion," *Cognition and Emotion*, 7, 357–387.

Geanakoplos J., D. Pearce and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60-79.

Gneezy, U. (2005): "Deception: The role of consequences," *American Economic Review*, 95, 384–394.

Gneezy, U., A. Kajackaite, and J. Sobel (2016): "Lie Aversion and the Size of a Lie," typescript.

Koszegi, B., and M. Rabin (2009): "Reference-dependent Consumption Plans," *American Economic Review*, 99, 909-936.

Munyo, I. and M. Rossi (2013): "Frustration, euphoria, and violent crime," *Journal of Economic Behavior & Organization*, 89, 136–142.

Reuben, E. and F. van Winden (2008): "Social Ties and Coordination on Negative Reciprocity: The Role of Affect," *Journal of Public Economics*, 92, 34–53.

Tadelis S. (2011): "The Power of Shame and the Rationality of Trust," typescript, UC Berkeley.

Sanfey, A. (2009): "Expectations and Social Decision-Making: Biasing Effects of Prior Knowledge on Ultimatum Responses," *Mind and Society* 8, 93–107.

Xiang, T., T. Lohrenz, and R. Montague (2013): "Computational Substrates of Norms and Their Violations during Social Exchange," *Journal of Neuroscience*, 33, 1099–1108.