Trust and Guilt Aversion: Theory
Lecture 13, *Experimental Econ. & Psychology*

Pierpaolo Battigalli
Bocconi University

28 October 2020

Abstract

Psychologists argue that "the prototypical cause of *guilt* would be the infliction of harm, loss, or distress" and that 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." In PGT, we model such guilt avoidance following Battigalli & Dufwenberg (2007): people feel guilty for making others get less than they expected. Since guilt has "negative valence", people are willing to trade off some personal material gains to decrease the probability with which and the extent to which they let others down. This has important economic implications. We focus on trust and deception.

Introduction: emotions

- For a long time, neither psychologists nor economists paid much attention to emotions and how they shape behavior, although founding figures in psychology like C. Darwin and W. James did (Keltner & Lerner's 2010 handbook chapter).
- Economist J. Elster (1996, 1998) argues that economists have neglected to study the emotions, although "all human satisfaction comes in the form of emotional experiences" (1996, p. 1368). Not recognizing this, economists may fail to get a correct grip on how decisions are formed.
- That view is corroborated by more recent developments in psychology. Keltner & Lerner (2010) argue that "a robust science of emotion ... emerged" (p. 317), indicating that a variety of emotions impacts well-being and behavior.
- The appraisal-tendency approach is often stressed (Lerner & Keltner 2000, 2001): **appraisal tendencies** are goal-directed processes through which emotions exert effects on judgments and decisions.

Introduction: guilt

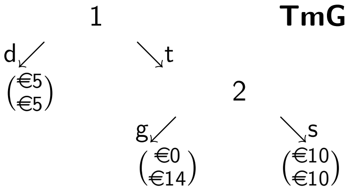
- Psychologists Baumeister et al. (1994) argue that "the prototypical cause of *guilt* would be the infliction of harm, loss, or distress" and that if "people feel guilt for hurting their partners ... and for failing to live up to their expectations, they will alter their behavior (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship" (see p. 245; cf. Tangney 1995).
- Economists assume that people feel guilty for making others get less than they expected. Since guilt has "negative valence", people are willing to trade off some personal material gains to decrease the probability with which and the extent to which they let others down (Battigalli & Dufwenberg 2007).
- Although some prominent psychologists stress guilt avoidance (see above), psychology mostly focuses on the *action tendency* following the experience of guilt, that is, "repair behavior." (See, e.g., Silfver 2007 for a discussion, and Attanasi et al. 2013 for a model of—*inter alia*—such tendency.)

Modeling guilt avoidance

- To model guilt avoidance in 2-person situations, we can posit the following psychological utility function (the meaning of math. symbols was explained in Lectures 9-12, in particular, $[x]^+ = \max\{0, x\}$): let $v_i(\pi_i)$ be the utility of money, with $v_i' > 0$, $v_i'' \leq 0$, then
  - $u_i(z, \alpha_{-i}) = v_i(\pi_i(z)) - G_i\left(\left[\mathbb{E}_{\alpha_{-i}}(\pi_{-i}) - \pi_{-i}(z)\right]^+\right)$, $G_i' \geq 0$,
  - e.g., $u_i(z, \alpha_{-i}) = \pi_i(z) - \theta_i\left[\mathbb{E}_{\alpha_{-i}}(\pi_{-i}) - \pi_{-i}(z)\right]^+$.
- In *n*-person situations:
  - $u_i(z, \alpha_{-i}) = v_i(\pi_i(z)) - G_i\left(\left(\left[\mathbb{E}_{\alpha_j}(\pi_j) - \pi_j(z)\right]^+\right)_{j \neq i}\right)$, $G_{i,j}' \geq 0$,
  - e.g., $u_i(z, \alpha_{-i}) = \pi_i(z) - \sum_{j \neq i} \theta_{ij}\left[\mathbb{E}_{\alpha_j}(\pi_j) - \pi_j(z)\right]^+$.
- It may be argued that "excessive expectations" (e.g., above equal sharing of max surplus) do not matter (cf. Balafoutas et al. 2017):
  - let $\bar{\pi}_j$ be the "legitimate limit" to $\mathbb{E}_{\alpha_j}(\pi_j)$, then
  - $u_i(z, \alpha_{-i}) = \pi_i(z) - \sum_{j \neq i} \theta_{ij}\left[\min\left\{\mathbb{E}_{\alpha_j}(\pi_j), \bar{\pi}_j\right\} - \pi_j(z)\right]^+$.

Guilt and Trust

- The Trust mini-Game is a very simple game form representing a stylized social dilemma:



- Would 2 **s**hare? Should 1 **t**rust 2? Assume *common knowledge that* $\theta_1 = 0$*, but* $\theta_2 \geq 0$ *is unknown to* 1 (cf. Attanasi et al. 2016). $\mathbb{E}_{\beta_2}\left(\mathbb{E}_{\alpha_1}(\pi_1)\,|\mathrm{t}\right)$=2's expectation of $\mathbb{E}_{\alpha_1}(\pi_1)$ cond. on t.
  - Pl. 2 prefers to **s**hare given **t**rust if $10 > 14 - \theta_2 \mathbb{E}_{\beta_2}\left(\mathbb{E}_{\alpha_1}(\pi_1)\,|\mathrm{t}\right)$.
  - Let pl. 1 **t**rust only if $\mathbb{E}_{\alpha_1}(\pi_1) \geq 5$ and let pl. 2 be certain of this (also *after* he observes t), then $\mathbb{E}_{\beta_2}\left(\mathbb{E}_{\alpha_1}(\pi_1)\,|\mathrm{t}\right) \geq 5$, and the sharing condition is $10 > 14 - 5\theta_2$, i.e., $\theta_2 > 0.8$.
  - Assuming pl. 1 is certain of this, she **t**rusts if $\alpha_1\left(\widetilde{\theta}_2 > 0.8\right) > 0.5$.

Guilt and deception: game forms, question [optional]

- **Cheap-Talk Sender-Receiver (CTSR) game form:** Pl. 1 sends a message $m \in \{m^A, m^B\}$, pl. 2 takes an action $a \in \{A, B\}$, payoffs depend only on a, but *only pl. 1 knows how.* Message m^A (m^B) says *"action A (B) gives you more money"*. Consider the following 3 cases, where only *pl. 1 knows the payoffs, pl. 2 knows nothing:*

$\Delta_i = \pi_i(B) - \pi_i(A)$	action	π_1	π_2
low stakes:	A	\$5	\$6
$\Delta_1 = 1 = -\Delta_2$	B	\$ (5 + 1)	\$ (6 − 1)
asymmetric stakes:	A	\$5	\$15
$10\Delta_1 = 10 = -\Delta_2$	B	\$ (5 + 1)	\$ (15 − 10)
high stakes:	A	\$5	\$15
$\Delta_1 = 10 = -\Delta_2$	B	\$ (5 + 10)	\$ (15 − 10)

- *How do the three cases of CTRS compare in terms of propensity to deceive (lie)?*

Guilt & deception: hypotheses [optional]

- *CK that $\theta_2 = 0$, but $\theta_1 \geq 0$ is unknown.* Let: Y=[do A iff m^A] (**trusting strategy**), N=[do A iff m^B] (**contrarian strategy**), $\Pi_2^X = \mathbb{E}_{\alpha_{2,1},X}(\pi_2)$=2's expected payoff from strat. X given $\alpha_{2,1}$. *Assume* the following about $\alpha_{2,1}$ and $\beta_{1,2}$ *for each case* (Why does it make sense? Because pl. 2 cannot distinguish!):

1. (**H.1. α-Symmetry**) $\alpha_{2,1}$ is s. t. $\mathbb{E}_{\alpha_{2,1},X}\left(\pi_2|m^A\right) = \Pi_2^X = \mathbb{E}_{\alpha_{2,1},X}\left(\pi_2|m^B\right)$ for each $X \in \{Y, N\}$ (m^A and m^B are perceived as equally truthful/deceiving). Thus Y (resp. N) is the unique BR iff $\Pi_2^Y > \Pi_2^N$ (resp. $\Pi_2^Y < \Pi_2^N$).
2. (**H.2: Belief in rationality, H.1, and trust**) 1's belief $\beta_{1,2}$ is s.t.
   1. H.1 certainly holds;
   2. pl. 2 (receiver) is certainly rational;
   3. pl. 2 (receiver) is likely to trust 1: $\mathbb{P}_{\beta_{1,2}}\left(\Pi_2^Y > \Pi_2^N\right) > 0.5$.
3. **(H.3: β-Symmetry)** Exp. disapp. depends only on realized payoff: $\forall x$, $\mathbb{E}_{\beta_{1,2}}\left(\left[\Pi_2^Y - x\right]^+ |\Pi_2^Y > \Pi_2^N\right) = \mathbb{E}_{\beta_{1,2}}\left(\left[\Pi_2^N - x\right]^+ |\Pi_2^Y < \Pi_2^N\right)$

Guilt & deception: comparative predictions [optional]

Let $D(x)$ denote the expected disappointment of pl. 2, according to 1's belief $\beta_{1,2}$, if 2 gets \$$x$. H.2-3 (+technical assumption) imply:

Lemma

Function D is strictly decreasing and convex ($D' < 0$, $D'' > 0$ where differentiable).

Corollary

The ratio $(D(x) - D(x+h))/h$ is strictly decreasing in $h > 0$.

Given $\beta_{1,2}$, let $\hat{\theta}_1^t = (\pi_1^t(B) - \pi_1^t(A)) / (D(\pi_2^t(B)) - D(\pi_2^t(A)))$, with $t \in \{\mathrm{ls}, \mathrm{as}, \mathrm{hs}\}$: *player 1 (sender) lies in t iff $\theta_1 < \hat{\theta}_1^t$.*

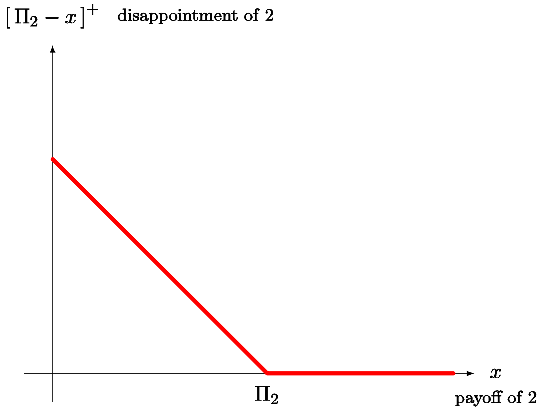
Proposition

Under H.2-3, the thresholds satisfy $0 < \hat{\theta}^{\mathrm{as}} < \hat{\theta}^{\mathrm{ls}} < \hat{\theta}^{\mathrm{hs}}$.

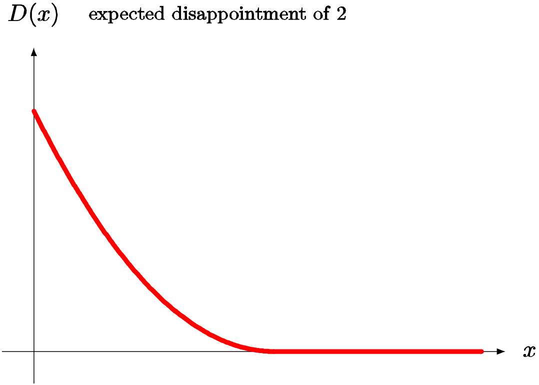
Guilt & deception: intuition for predictions [optional]

(1) Disappointment is convex

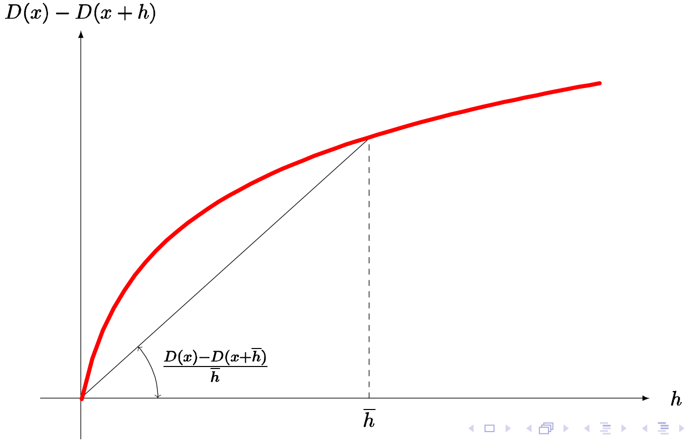
The higher the payoff of 2 the less he is disappointed. Furthermore, disappointment $\left[\Pi_2^X - x\right]^+$ is a convex function of realized payoff x.



Therefore, $D(x)$, the expected disappointment of 2 according to 1's (2nd-ord.) belief $\beta_{1,2}$ is decreasing and convex in the payoff x of 2.



Thus, $(D(x) - D(x+h))/h$ is strictly decreasing in $h > 0$, because $f(h) = D(x) - D(x+h)$ is strictly increasing and concave, with $f(0) = 0$.



Guilt & deception: intuition for predictions [optional]

With this, **Proposition 1** follows from the payoff differences in the 3 cases:

- $\hat{\theta}^{\text{as}} < \hat{\theta}^{\text{ls}}$ because

$$\frac{1}{D(5) - D(5+10)} < \frac{1}{D(5) - D(5+1)}$$

(by (2), D is decreasing);

- $\hat{\theta}^{\text{ls}} < \hat{\theta}^{\text{hs}}$ because

$$\frac{1}{D(5) - D(5+1)} < \frac{10}{D(5) - D(5+10)}$$

(by (3), $h/(D(x) - D(x+h))$ increasing in h).

Guilt & deception: the role of payoff consequences

- Gneezy (2005) designed a clever experiment with 3 (main) treatments:

$\Delta_i = \pi_i(B) - \pi_1(A)$	action	π_1	π_2
1: low stakes: $\Delta_1 = 1 = -\Delta_2$	A	\$5	\$6
	B	\$(5+1)	\$(6−1)
2: asymm. stakes: $10\Delta_1 = 10 = -\Delta_2$	A	\$5	\$15
	B	\$(5+1)	\$(15−10)
3: high stakes: $\Delta_1 = 10 = -\Delta_2$	A	\$5	\$15
	B	\$(5+10)	\$(15−10)

- According to **Proposition 1**, under the stated assumptions about guilt aversion and 2^{nd}-ord. beliefs, *senders tend to lie the least in treatment 2 (as) and the most in treatment 3 (hs).* The frequencies of lies across treatments are *consistent with this prediction*.

References

Battigalli P., G. Charness, and M. Dufwenberg (2013): "Deception: The Role of Guilt," *Journal of Economic Behavior & Organization*, **93**, 227-232. [Downloadable from webpage, optional.]

Battigalli, P., C. Corrao, and M. M. Dufwenberg (2019): "Incorporating Belief-Dependent Motivation in Games," *Journal of Economic Behavior & Organization*, **167**, 185-218. [Downloadable from webpage, optional.]

Battigalli P., and M. Dufwenberg (2007): "Guilt in Games," *American Economic Review (P&P)*, **97**, 170-176. [Downloadable from webpage, optional.]

Battigalli, P., and M. Dufwenberg (2020): "Belief-Dependent Motivations and Psychological Game Theory," *Journal of Economic Literature*, forthcoming.

Additional references: economics (all optional)

- 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 W.P. 506, Bocconi University.
- Balafoutas, L., and H. Fornwagner (2017): "The Limits of Guilt". *Journal of the Economic Science Association*, **3**, 137-148.
- Gneezy, U. (2005): "Deception: the Role of Consequences," *American Economic Review*, **95**, 384-394.

Additional references: psychology (all optional)

- BAUMEISTER, R., A. STILLWELL, AND T. HEATHERTON (1994): "Guilt: An Interpersonal Approach," *Psychological Bulletin*, **115**, 243-267.
- SILFVER, M. (2007): "Coping with Guilt and Shame: A Narrative Approach," *Journal of Moral Education*, **36**, 169-183.
- TANGNEY, J.P. (1995): "Recent Advances in the Empirical Study of Shame and Guilt," *American Behavioral Scientist*, **38**, 1132-1145.