Deception and Image Concerns: Theory
Lecture 19, *Experimental Econ. & Psychology*

Pierpaolo Battigalli
Bocconi University

18 November 2020

Abstract
Introspection and evidence tell us that people lie much less than justified by the maximization of their expected material payoff. Here we consider models of image concerns, whereby an agent dislikes to be perceived as someone who has lied (see Dufwenberg & Dufwenberg 2018), or who does not have an intrinsic motivation to tell the truth (e.g., Gneezy et al. 2018, Khalmetski & Sliwka 2019). We mostly focus on the first kind of motivation, i.e., a concern for the opinion of others about one's own good or bad behavior.

Introduction

- For decades, economists have assumed that individuals are willing to misreport private information, if this maximizes their expected material payoff. Yet, evidence shows that this is not a correct description of human nature (e.g., Fischbacher & Föllmi-Heusi 2013, Abeler et al. 2019).
- We already analyzed deception in the context of Sender-Receiver Cheap-Talk Game forms and showed that (partial) truthtelling may be explained by belief-dependent other-regarding preferences, such as guilt aversion, on top of a mere dislike for lying. Yet such preferences cannot play a role in mere reporting game forms where there is no other party whose payoff can be affected.
- Thus, here we analyze a different reason for truthtelling, **image concerns**:
  - the dislike for being perceived as having lied (main focus of the lecture),
  - or the dislike for being perceived as someone who has a low intrinsic motivation to tell the truth.

Truthtelling and image concerns

- For each path (terminal history) z, let $L_i(z) \geq 0$ denote the extent of i's lies in path z; for example $z = (x, y)$ with x=privately observed realization, y=report about x. A special case is the indicator function: $L_i(z) = 1$ if i lied ($x \neq y$), $L_i(z) = 0$ if i told the truth ($x = y$). Also recall that $\mathcal{P}_j(z)$ is the *ex post* information set of j if z occurs, i.e., the set of paths that j cannot distinguish from z given his *ex post* information feedback.
- **Image concerns:**
  1. **Concern for others'** (*ex post*) **opinion about good/bad behavior**:

$$u_i(z, \alpha) = \pi_i(z) - \theta_i \mathbb{E}_{\alpha_j}(L_i | \mathcal{P}_j(z)).$$

  2. **Concern for others'** (*ex post*) **opinion about good/bad traits**: $0 \leq \theta_i^{\mathbf{I}}$=**intrinsic**-motivation trait, $0 \leq \theta_i^{\mathbf{R}}$=**reputational**-motivation trait,

$$u_i(z, \alpha) = \pi_i(z) - \theta_i^{\mathbf{I}} L_i(z) + \theta_i^{\mathbf{R}} \mathbb{E}_{\alpha_j}\left(\widetilde{\theta_i^{\mathbf{I}}} | \mathcal{P}_j(z)\right)$$

A theoretical analysis of cheating: model

- In the model of Dufwenberg & Dufwenberg (2018, D&D), the game form comes from the seminal experiment of Fischbacher & Föllmi-Heusi (2013):
  - Chance (pl. 0) move with realization $x \in \{0, 1, ..., 5\}$, $p_0(x) = \frac{1}{6}$;
  - Player 1 (the only active person) privately observes x and reports $y \in \{0, 1, ..., 5\}$;
  - Player 2 observes y; thus, $\mathcal{P}_2(x, y) = \{(x', y') : y' = y\} = \{0, 1, ..., 5\} \times \{y\}$ (simply written below as y)
  - $\pi_1(x, y) = y$, $\pi_2(x, y) = const.$
- Here, $z = (x, y)$, $\alpha_2 = \left(\alpha_2(\cdot|\varnothing), (\alpha_2(\cdot|y))_{y=0}^{5}\right)$ [with $\alpha_2(\cdot|\varnothing) \in \Delta(Z)$] satisfies Bayes rule whenever possible.
  - D&D assume: $u_1((x, y), \alpha) = y - \theta_1 \cdot \sum_{x'} \alpha_2(x'|y)[y - x']^+$.
  - **Note:** $u_1((x, y), \alpha)$ is *independent* of x (and α_1), as player 1 only cares about his material payoff y and 2's perception given y and 2's system of (endogenous) 1^{st}-order beliefs α_2.

A theoretical analysis of cheating: equilibrium beliefs

- Like most papers on this topic D&D obtain (α_1, α_2) from *equilibrium* analysis: pl. 1 believes that pl. 2 (observer) knows his **plan** $\left(\alpha_1(\cdot|x)_{x=0}^{5}\right)$ (because pl. 1 is "his own audience").
- Thus, pl. 1 believes that pl. 2 *initially* assigns to every path (x, y) probability
$$\alpha_2(x, y|\varnothing) = \frac{1}{6}\alpha_1(y|x),$$
where $\alpha_1(y|x)$ is the *planned and actual* probability of pl. 1 reporting y given realization x.
- By **Bayes rule**, for all x, x', and y, **if** $\sum_{x=0}^{5}\alpha_1(y|x) > 0$ **then**
$$\alpha_2(x'|y) = \frac{\frac{1}{6}\alpha_1(y|x')}{\sum_{x=0}^{5}\frac{1}{6}\alpha_1(y|x)} = \frac{\alpha_1(y|x')}{\sum_{x=0}^{5}\alpha_1(y|x)}. \qquad \text{(BR)}$$
- Thus, the 2^{nd}-order belief of 1 about the audience (pl. 2) satisfies: $\beta_{12}(\tilde{\alpha}_2 = \alpha_2|x) = 1$ and $\bar{u}_{1,x}(y, \beta_{12}) = u_1(x, y, \alpha_2)$ for all x, y, where α_2 is given by (BR).

A theoretical analysis of cheating: equilibrium incentives

- Plan α_1 is **rational** given β_{12} if, for each realization x, it assigns positive probability only reports y that maximize the expected value of u_1.
- Since such expected value is $\bar{u}_{1,x}(y, \beta_{12}) = u_1(x, y, \alpha_2)$, the **incentive condition of rational planning** is
    - for all $\bar{x}, \bar{y} \in \{0, 1, ..., 5\}$,

$$\alpha_1(\bar{y}|\bar{x}) > 0 \Longrightarrow \bar{y} \in \arg\max_{y \in \{0,1,\dots,5\}} \left(y - \theta_1 \sum_{x'=0}^{5} [y - x']^+ \alpha_2(x'|y) \right),$$

    - where $\alpha_2(x'|y) = \frac{\alpha_1(y|x')}{\sum_{x=0}^{5} \alpha_1(y|x)}$ by (BR).

A theoretical analysis of cheating: results, no truthtelling

- **Observation:** *Truthtelling ($\alpha_1(x|x) = 1$ for each x) is not an equilibrium.*
- **Proof:** We show that truthtelling and (BR) imply an incentive to deviate. Intuitively, player 1 would want to report $y = 5$ being certain of being believed. Formally:
  - By (BR), if $(\forall x, \alpha_1(x|x) = 1)$ then $(\forall y, \alpha_2(y|y) = 1)$.
  - Then $\sum_{x'=0}^{5} [5 - x']^+ \alpha_2(x'|5) = [5 - 5]^+ = 0$.
  - Then, $\forall x < 5$, $u_1(x, 5, \alpha_2) = 5 > x = u_1(x, x, \alpha_2)$. ■

- Define **pure cheating** as player 1 planning to report always $y = 5$: $\alpha_1(5|x) = 1$ for each x.
- **Observation: (1)** *If $\theta_1 < 2$ (low image concerns), pure cheating is an equilibrium.* **(2)** *If $\theta_1 > 2$ (high image concerns), pure cheating is not an equilibrium; thus, if $\theta_2 > 2$ there are partial lies in equilibrium.*
- **Proof of (2):** We show that pure cheating and (BR) imply an incentive to under-report. Intuitively, reporting $y = 0$ yields 0 payoff and 0 perceived cheating, because *under-reporting is not a relevant form of cheating* in this model. Reporting $y = 5$ makes the audience believe that pl. 1 is likely to have cheated. Formally:
  - By (BR), if $(\forall x, \alpha_1(5|x) = 1)$ then $\left(\forall x', \alpha_2(x'|5) = \frac{1}{6}\right)$ (the posterior after $y = 5$ is equal to the prior).
  - Then, $u_1(x, 5, \alpha_2) = 5 - \theta_1(5 + 4 + 3 + 2 + 1)\frac{1}{6} = 5 - \theta_1\frac{5}{2} < 0 = u_1(x, 0, \alpha_2)$ if $\theta_1 > 2$. ■

A theoretical analysis of cheating: a plausible equilibrium

- D&D prove the following:

Proposition

If $\theta_1 > 2$ (high image concerns), there is a unique equilibrium such that:

- *each report has positive probability: for all y,* $\sum_{x=0}^{5} \alpha_1 (y|x) > 0$;
- *there are no downward lies: for all x and all* $y < x$, $\alpha_1 (y|x) = 0$;
- *there is uniform upward cheating: for all* $x < 4$ *and all* $y', y'' > x$, $\alpha_1 (y'|x) = \alpha_1 (y''|x)$.

- This equilibrium explains well the data of Fischbacher & Föllmi-Heusi (2013) (see Fig. 2 at p 255 of D&D).

References: general

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

References: deception/truthtelling (optional)

- ABELER, J., D. NOSENZO, AND C. RAYMOND (2019): "Preferences for Truth-Telling," *Econometrica*, **87**, 1115-1153.
- DUFWENBERG, M., AND M.A. DUFWENBERG (2018): "Lies in Disguise – A Theoretical Analysis of Cheating". *Journal of Economic Theory*, **175**, 248-264.
- FISCHBACHER, U., AND F. FÖLLMI-HEUSI (2013): "Lies in Disguise – An Experimental Study on Cheating," *Journal of the European Economic Association*, **11**, 525-547.
- GNEEZY, U., A. KAJACKAITE, AND J. SOBEL (2018): "Lying Aversion and the Size of the Lie," *American Economic Review*, **108**, 419-453.
- KHALMETSKI, K. AND D. SLIWKA (2019): "Disguising Lies - Image Concerns and Partial Lying in Cheating Games," *American Economic Journal: Microeconomics*, **11**, 79-110.

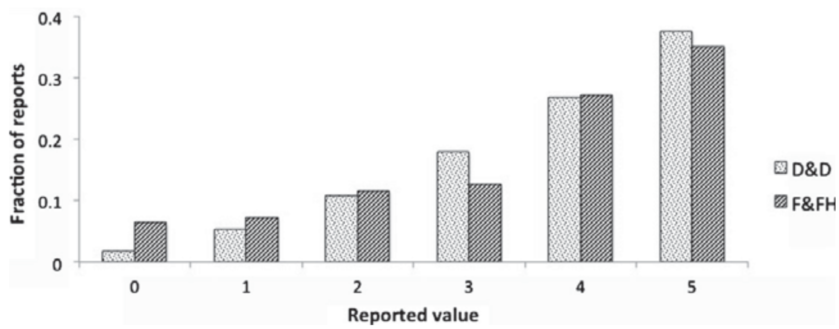

Fig. 2. D&D theory vs F&FH data.

Since $\varepsilon_y > 0$, (9) has a solution iff $y - \theta \cdot \sum_{x'<y}(\frac{\pi_{x'}}{(\sum_{x<y}\pi_x)+\pi_y} \cdot [y - x']) < 0$, or $\theta > y/\sum_{x'<y}(\frac{\pi_{x'}}{\sum_{x\leq y}\pi_x} \cdot [y - x']) > 1$. The solution is unique.

Inspecting the conditions on θ needed for solutions of (6) & (9) to exist (mentioned after (6) & (9)), one infers that an SE exists iff:

$$\theta > \widehat{\theta}((\pi_x)_{x\leq n}) = \max_{y\in\{1,\ldots,n\}} y/\sum_{x'<y}(\frac{\pi_{x'}}{\sum_{x\leq y}\pi_x} \cdot [y - x']). \tag{10}$$

If (10) holds, since (6) and (9) uniquely define $\varepsilon_y \in (0, 1)$ for $y \in \{1, \ldots, n\}$, the SE s is uniquely defined while satisfying all the desired properties.

If x is drawn from the uniform distribution we get $\theta > \widehat{\theta}((\pi_x)_{x\leq n}) = 2$. To see this, plug $\pi_x = \pi_{x'} = \frac{1}{n+1}$ for all x, x' into the rhs of (10), which then equals $y/(y \cdot \frac{\frac{1}{n+1}}{(y+1)\cdot\frac{1}{n+1}} \cdot \frac{y+1}{2}) = 2$. □

It is useful to have a name for the SE highlighted in the Proposition. It involves that if DM observes x then he reports each $y > x$ with positive probability; we suggest *sailing-to-the-ceiling* as an apt monicker.

The critical value $\widehat{\theta}((\pi_x)_{x\leq n})$, defined in (10), depends on the distribution $(\pi_x)_{x\leq n}$, and may be affected by n in interesting ways (although n is irrelevant if the distribution is uniform). We return to this topic in section 4.2.

3.4. The proposition vs. F&FH's data

How does the sailing-to-the-ceiling SE of the Proposition stand up to data? Most studies assume that x is drawn from a uniform distribution. This simplifies computation. Namely, plug $\pi_x = \frac{1}{n+1}$ for all x into (6) & (9). For $0 < y \leq n$ we get

$$y - \theta \cdot \frac{1-\varepsilon_y}{y\cdot(1-\varepsilon_y)+1} \cdot \frac{y\cdot(y+1)}{2} = 0$$

$$\Longleftrightarrow \qquad (1-\varepsilon_y) = \frac{2}{y\cdot(\theta-2)+\theta}. \tag{11}$$

Focus on F&FH's die-roll setting: $n = 5$. Using (11), we can compute the SE, which can be eerily similar to F&FH's data (and other studies; see AN&R). This is illustrated in Fig. 2 where the prediction is generated with $\theta = 3$.[8]

See section 5 for more discussion of experimental tests.

[8] Using (11), we get $(1-\varepsilon_5) = \frac{2}{5(3-2)+3} = \frac{1}{4}$, so $\sum_{x\leq 5}\pi_x \cdot s(x)(5) = 5(\frac{1}{6}\cdot\frac{1}{4}) + \frac{1}{6}\cdot 1 = \frac{3}{8}$, etc.