Communication, Trust and Guilt: Experiments
Lecture 14, *Experimental Econ. & Psychology*

Pierpaolo Battigalli
Bocconi University

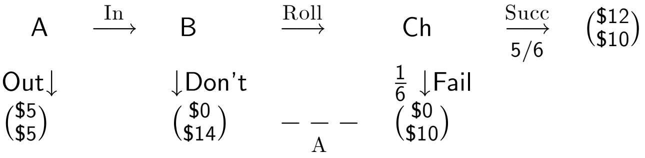
29 October 2020

Abstract
We present some experimental results about trust and communication. The experiments were either explicitly designed to test guilt-aversion (GA) models, or to test the effects of communication on trust and how it is affected by the material payoffs of the game form. The results are quite consistent with models of GA, but there are alternative explanations and possible confounds. We focus on the effects of communication on behavior and beliefs in a version of the Trust mini-Game (Charness and Dufwenberg 2006), and on how payoff consequences affect truthful communication in a simple Sender-Receiver Cheap-Talk game (Gneezy 2005).

Introduction

- Guilt is probably the most researched *emotion* in economics (see Azar 2019, we do not classify reciprocity as an emotion).
- The first experimental study testing guilt aversion (GA) (Dufwenberg & Gneezy 2000) measures 2^{nd}-order beliefs in an experimental "Lost Wallet Game" whereby the first mover chooses between an outside option and a trusting action letting the second mover play a Dictator game (the backdrop story is that returning a lost wallet promotes efficiency because it is more valuable for the owner than for the finder, it is also a trusting action because the owner may reward or not the finder, who could have kept the wallet).
- Here we focus on a seminal contribution by Charness & Dufwenberg (2006, henceforth C&D), who experimentally analyzed versions of the Trust mini-Game, prompting an interesting experimental literature (see Cartwright 2019).
- We also consider an experiment about deception (Gneezy, 2005).

Trust mini-Game with imperfect monitoring



	D	Roll
O	5, 5	5, 5
In	0,14	10,10

Strategic−form

- Charness & Dufwenberg (2006) analyze experimentally a variation of the TmG where the \$0 payoff of the first mover (pl. A) may be due to bad luck.
  - A-subjects observe *ex post* only their own payoff. This is not directly relevant for the simple guilt-aversion theory, but it may matter for other reasons, e.g., image concerns.
  - The game is played with the **strategy method**: B-subjects are asked to (covertly) commit in advance to the choice to be made in case A goes In. The resulting strategic form (with average payoffs given In-Roll) is the same as for the TmG shown in previous lectures.

Second-order beliefs

- In most PGT model, 2^{nd}-order beliefs $\beta_{i,-i}$ are key to understand incentives: best replies depend on $\beta_{i,-i}$ and personal traits θ_i. *Some* experiments about guilt try to elicit (=measure) θ_i (e.g., Bellemare et al. 2011, Attanasi et al. 2013), *most* experiments
  - try to elicit key features of $\beta_{i,-i}$—such as $\mathbb{E}_{\beta_{\mathrm{B},\mathrm{A}}}\left(\mathbb{E}_{\alpha_{\mathrm{A}}}\left(\pi_{\mathrm{A}}\right)|\mathrm{In}\right)$ in the TmG above—and analyze correlation with choices.
  - try to manipulate $\beta_{i,-i}$ by changes of the game form that are supposed to move $\beta_{i,-i}$ unambiguously in one direction, e.g., treatments that should move upward the empirical distribution, or average, of $\mathbb{E}_{\beta_{\mathrm{B},\mathrm{A}}}\left(\mathbb{E}_{\alpha_{\mathrm{A}}}\left(\pi_{\mathrm{A}}\right)|\mathrm{In}\right)$.
- C&D had many subjects in a room randomly assigned to roles A and B, and anonymously matched to play (the strategic form of) the TmG. After choices were made, they asked
  - A-subjects to guess the proportion of B's who chose Roll;
  - B-subjects to guess the average guess *of the A's who chose* In;
  - "almost correct" answers were rewarded with \$5.

Guilt, trust, and communication: design

- To test GA, C&D considered a control version of the TmG and several alternative treatments. The main treatment manipulations in the design were:
  - Most importantly, each B-subject was given the opportunity to send a free-form message to the paired A-subject [e.g., one message was: *"If you choose In, I will choose to roll. This way we both have an opportunity to make more that 5$!:)"*].
    - *Expected effect:* messages were expected to move $\hat{\pi}_{B,A}\left(\beta_{B,A}\right) = \mathbb{E}_{\beta_{B,A}}\left(\mathbb{E}_{\alpha_A}\left(\pi_A\right)|\text{In}\right)$ upward by a self-fulfilling expectation of trust: my message is likely to make you trust me, which increases my $\hat{\pi}_{B,A}\left(\beta_{B,A}\right)$, which makes you (who understand this) likely to trust me ... .
  - To check the robustness of the effects of communication, the **Out**side-option payoffs were changed from ($5,$5) to ($7,$7).
  - Also, A-subjects, instead of B-subjects, were given the opportunity to send a message.
- C&D found a significant (i) positive correlation of Roll with (elicited) $\hat{\pi}_{B,A}\left(\beta_{B,A}\right)$, and (ii) effect of communication.

Guilt, trust, and communication: results

- **Beliefs & Behavior:** C&D observe a strong correlation in the expected direction between beliefs and behavior for both A's and B's in all treatments: *B's who chose* $Roll$ *made significantly higher guesses about A's guesses compared to those who chose* $Don't$ (but this may also follow from false consensus, see Cartwright 2019):
  - (5,5) No Messages, average 2^{nd}-order guess of B's (cond. on In): 54% (given Roll) vs 40% (given $\mathrm{Don't}$);
  - (5,5) *Messages*, average 2^{nd}-order guess of B's (cond. on In): 73% (given Roll) vs 45% (given $\mathrm{Don't}$).
- **Communication:** More strikingly, C&D observe a *strong effect of communication:* approximately (see Fig 3, p 1587 of C&D),
  - (5,5), No Messages: 55% of A's go In, 45% of B's Roll, 22% of pairs choose In-Roll (no correlation);
  - (5,5) *Messages*: 75% of A's go In, 70% of B's Roll, 50% of pairs choose In-Roll (correlation of choices mediated by messages).

Deception, the role of payoff consequences: design

- Gneezy (2005) experimentally analyzes Cheap-Talk Sender-Receiver (CTSR) game forms where the Receiver (pl. 2) has no knowledge about payoffs, the Sender (pl. 1) knows them and can lie or tell the truth about the action that gives more to the Receiver. Treatments:

$\Delta_i = \pi_i(B) - \pi_i(A)$	action	π_{S}	π_{R}
1: low stakes: $\Delta_{\mathrm{S}} = 1 = -\Delta_{\mathrm{R}}$	A B	\$5 \$ (5 + 1)	\$6 \$ (6 − 1)
2: asymm. stakes: $10\Delta_{\mathrm{S}} = 10 = -\Delta_{\mathrm{R}}$	A B	\$5 \$ (5 + 1)	\$15 \$ (15 − 10)
3: high stakes: $\Delta_{\mathrm{S}} = 10 = -\Delta_{\mathrm{R}}$	A B	\$5 \$ (5 + 10)	\$15 \$ (15 − 10)

- We analyzed a **GA-model** in the CTSR implying that *Senders tend to lie the least in treatment 2 (a.s.) and the most in treatment 3 (h.s.).* Note: Gneezy was not trying to test this model, he only wanted to study the effects of payoff changes on deception.

Deception, the role of payoff consequences: results

- Recall: in the CTSR, $\Delta_i = \pi_i(B) - \pi_i(A)$ is the profit/loss to $i \in \{\mathrm{S}, \mathrm{R}\}$ for switching from A (best for Receiver, pl. 2) to B (best for Sender, pl. 1)

Freq. lies	$\Delta_{\mathrm{R}} = -1$	$\Delta_{\mathrm{R}} = -10$
$\Delta_{\mathrm{S}} = 1$	(l.s.) 36%	(a.s.) 17%
$\Delta_{\mathrm{S}} = 10$	———	(h.s.) 52%

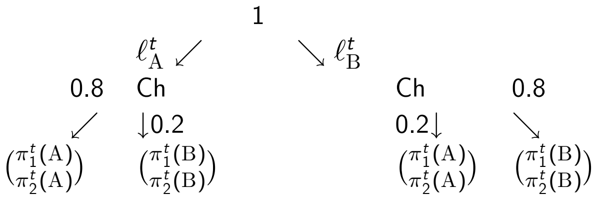
- As predicted by the GA model (due to expected disappointment being decreasing and convex in the Receiver's payoff), Gneezy (2005) finds the lowest frequency of lies (17%) in the asymmetric-stakes treatment, the highest in the high-stakes treatment (52%). See Fig. 1, p 388.
- A fixed cost of lying may play a role in reducing the number of lies in each treatment, but cannot drive the comparative results, because payoffs matter.

Deception vs distributional preferences: Dictator Game

- Are the results in the 3 CTSR's explained only by distributional preferences (or GA)? Gneezy compared with 3 noisy binary Dictator Games, that should be roughly equivalent given such preferences.
- About 80% of R's followed the message in the CTSR's. Thus, the following 3 noisy Dictator Games below were played:

$t \in \{\mathrm{ls}, \mathrm{as}, \mathrm{hs}\}$

$\pi_1^t(A) < \pi_1^t(B)$
$\pi_2^t(A) > \pi_2^t(B)$



- **Results** (See Fig. 2, p 389, compare with Fig. 1, p 388):
  - *The freq.s of choice of* ℓ_{B} *in the treatments* $t \in \{\mathrm{ls}, \mathrm{as}, \mathrm{hs}\}$ *are ordered as the freq.s of lies in the corresponding CTSR game forms:* $Fr(\ell_{\mathrm{B}}^{\mathrm{as}}) < Fr(\ell_{\mathrm{B}}^{\mathrm{ls}}) < Fr(\ell_{\mathrm{B}}^{\mathrm{hs}})$.
  - *But the freq.s in the DG's are much higher:* $\forall t$, $Fr(\ell_{\mathrm{B}}^{\mathrm{t}}) \gg Fr(\mathrm{lie}^{\mathrm{t}})$.

Dictator Game vs CTSR: discussion

- *The results in the DG are consistent with GA.* The driving force is the same as in the CTSR: expected disappointment is decreasing and convex in π_2, monotonicity explains why $Fr(\ell_{\mathrm{B}}^{\mathrm{as}}) < Fr(\ell_{\mathrm{B}}^{\mathrm{ls}})$, convexity explains why $Fr(\ell_{\mathrm{B}}^{\mathrm{ls}}) < Fr(\ell_{\mathrm{B}}^{\mathrm{hs}})$ (cf. Lecture 13; in this case the analysis is simpler, because we do not need auxiliary hypotheses concerning beliefs of receivers about lies and 2^{nd}-order beliefs of the Senders about such beliefs).
- The CTSR *vs* DG comparison can be explained by the existence of a cost of lying interacting with GA in the CTSR, but immaterial in the DG.
- The DG results are not explained by inequality aversion: $Fr(\ell_{\mathrm{B}}^{\mathrm{as}})$ (with $\ell_{\mathrm{B}}^{\mathrm{as}}$ giving the less unequal distribution—($6,$5) instead of ($5,$15)—with prob. 80%) should be higher, not lower than $Fr(\ell_{\mathrm{B}}^{\mathrm{ls}})$ and $Fr(\ell_{\mathrm{B}}^{\mathrm{hs}})$ (in ls and hs the alternative distributions are symmetrically unequal). They may be consistent with some form of partial altruism.

References: Theory

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., 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.

References: experiments [optional]

- Attanasi G., P. Battigalli, and R. Nagel (2013): "Disclosure of Belief-Dependent Preferences in the Trust Game," IGIER W.P. 506, Bocconi University.
- Azar, O. (2019): "The Influence of Psychological Game Theory," *Journal of Economic Behavior & Organization*, **167**, 445-453.
- Bellemare, C., A. Sebald, M. Strobel (2011): "Measuring Willingness to Pay to Avoid Guilt: Estimation using Equilibrium and Stated Belief Models," *Journal of Applied Economics*, **26**, 437-453.

References: experiments [optional]

- Cartwright, E. (2019): "A Survey of Belief-Based Guilt Aversion in Trust and Dictator Games," *Journal of Economic Behavior & Organization*, **167**: 430-444.
- Charness, G., and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, **74**, 1579-1601.
- Gneezy, U. (2005): "Deception: the Role of Consequences," *American Economic Review*, **95**, 384-394.
- Dufwenberg, M., and U. Gneezy (2000): "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, **30**, 163-182.

We find it natural to focus on such messages, because previous work has indicated that promises can induce commitments to cooperate (cf., e.g., Kerr and Kaufman-Gilliland (1994), Ellingsen and Johannesson (2004)).

4. RESULTS

We consider the effect of communication in Section 4.1, beliefs and behavior in Section 4.2, and the effect of promises on beliefs and behavior in Section 4.3.

4.1. *The Effect of Communication*

Figure 3 summarizes choices with and without B messages in our payoff calibrations. In the (5, 5) treatment without B messages, 20 of 45 (44%) B's chose Roll and 25 of 45 (56%) A's chose *In*. When B could send a message to A, we observe considerably more cooperation: 28 of 42 (67%) B's chose Roll and 31 of 42 (74%) A's chose *In*. The (*In*, *Roll*) profile occurred 20% of the time (9 of 45 pairs) without communication, compared to 50% (21 of 42 pairs) with messages possible from B's.

We observe similar effects in the (7, 7) treatment. Without B messages, 12 of 48 (25%) B's chose Roll and 11 of 48 (23%) A's chose *In*. When B could send a message to A, we once again observe considerably more cooperation: 24 of 49 (49%) B's chose Roll and 23 of 49 (47%) A's chose *In*. The (*In*, *Roll*) profile

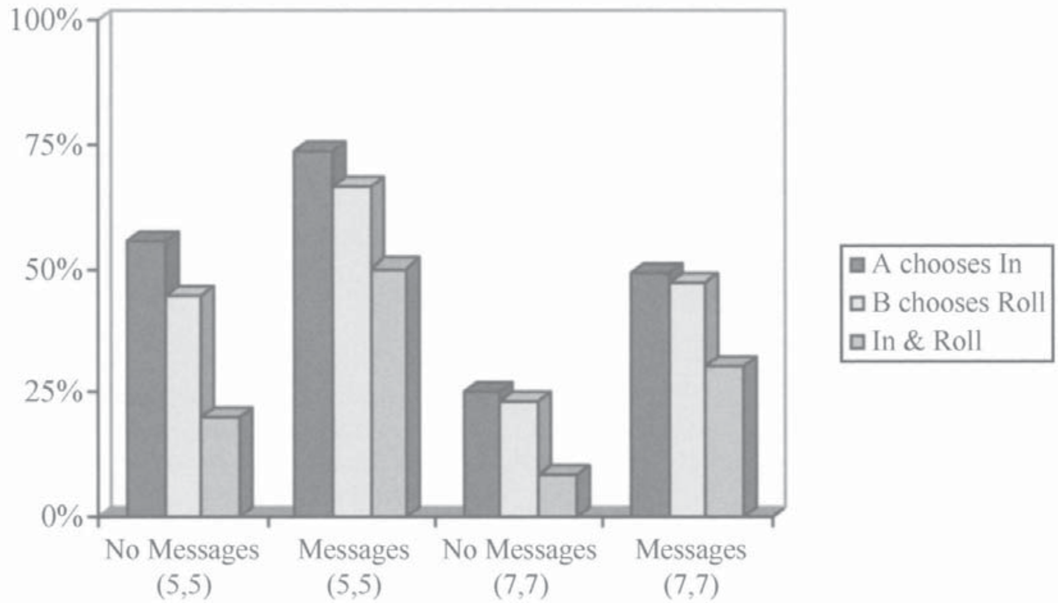


FIGURE 3.—The effect of messages from B.

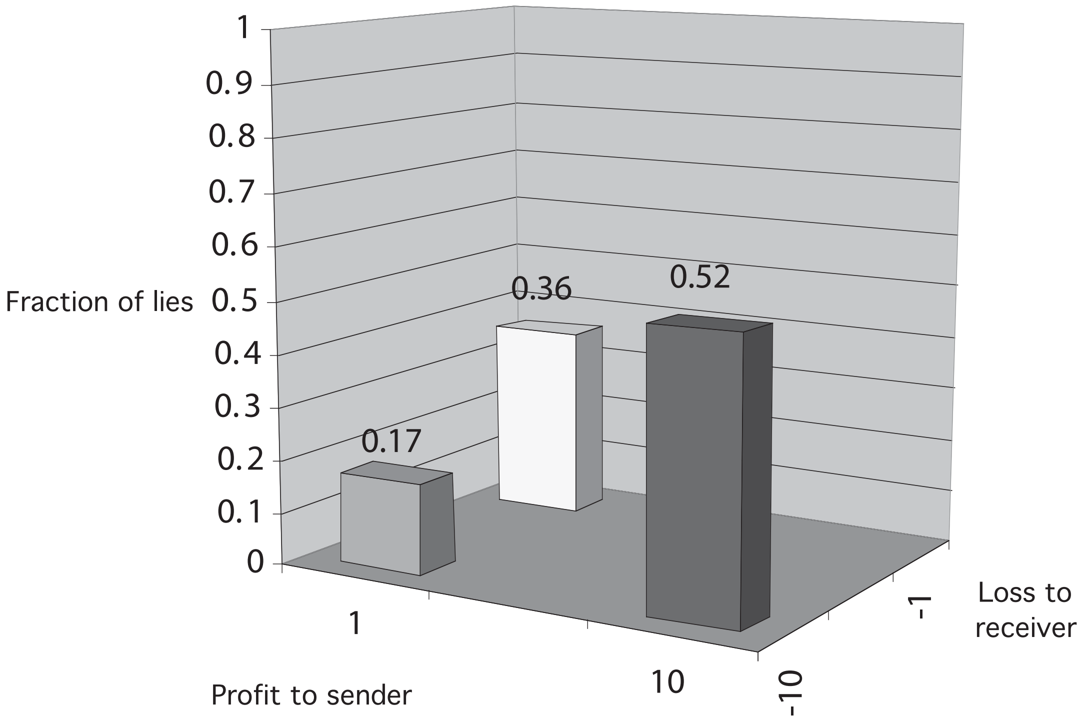


FIGURE 1. FRACTION OF PARTICIPANTS WHO LIED IN THE DECEPTION GAME

Note: The horizontal axis represents the gains from lying for player 1 and the associated loss for player 2.

receiver were both $10, that number rose to 39 (52 percent). A statistical comparison of these differences shows that they are all significant.[10]

Deception versus Choices between Allocations.—In order to determine the extent to which these results reflect an aversion to lying as opposed to preferences over distributions of payoffs, I used a control dictator treatment in which player 1 chose between two monetary allocations, just as in the deception game. Player 2 has no choice in this control treatment. The probability of executing player 1's choice was 80 percent, while in the other 20 percent the alternative allocation was implemented. Since approximately 80 percent of player 2s followed player 1s' recommendation in the deception game, this results in a treatment that is equivalent, in payoff, to the deception game. If player 1s had chosen the materially advantageous allocation more often in this control treatment, it would be direct evidence of lie-aversion (and against consequentialist preferences). The results are presented in Figure 2, with $N = 50$ in each of the cells of the dictator game.

The results presented in Figure 2 display the same pattern we observed in the deception game, but to a much greater degree. The results of the two games are compared in Table 2.

From these results I conclude that it is not only care for others that motivates behavior, but also aversion to lying. People's choices reflect nonconsequentialist preferences since, for example, they treat the choice between (5, 6) and (6, 5) differently, depending on whether it was a simple choice or a lie that led to the final outcome.

B. *The Questionnaires*

What do people think about the role of consequences in lying, and what do they say about the relative fairness of different lies? I studied these issues with a set of questionnaires whose items referred to an empirically realistic scenario. The participants in this study were students at the University of Chicago who volunteered to fill out the questionnaires and were paid $1 for their participation. They were asked to judge the following scenario:

[10] The p-values are approximated to three decimal places and calculated from a one-tailed test of the equality of proportions, using normal approximation to the binomial distribution. For the comparison of treatment 1 and 2, $Z = 2.58$, and $p = .005$. For treatment 1 versus 3, $Z = 1.97$, and $p = .024$, and for treatment 2 versus treatment 3, $Z = 4.48$ and $p = .001$.

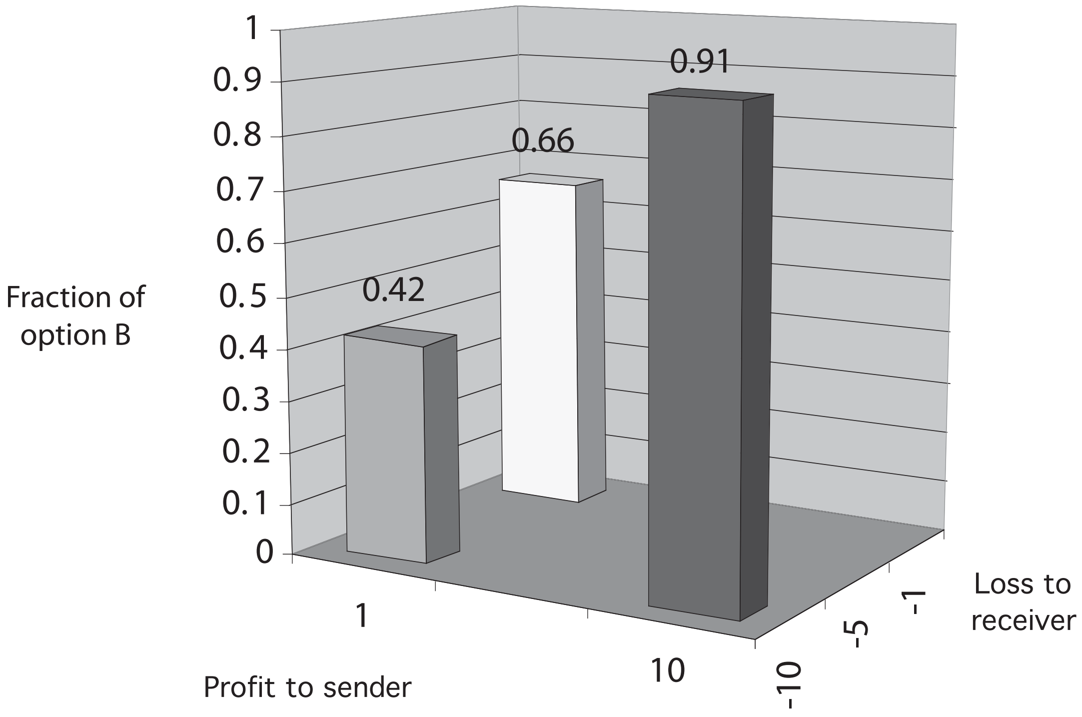


FIGURE 2. FRACTION OF PLAYER 1S WHO CHOSE OPTION B IN THE DICTATOR GAME

Note: The horizontal axis represents the gains from choosing B for player 1 and the associated loss for player 2.

TABLE 2—THE FRACTION OF PLAYER 1S WHO CHOSE ALLOCATION B

	Allocations		
Game	5, 6 versus 6, 5	5, 15 versus 6, 5	5, 15 versus 15, 5
Deception	0.36	0.17	0.52
Dictator	0.66	0.42	0.90

Notes: All differences between the dictator game and the deception game for a given distribution of payoffs are statistically significant at $P < 0.01$. Differences between the different allocations within the dictator game are also statistically significant at the 0.01 level.

> Mr. Johnson is about to close a deal and sell his car for $1,200. The engine's oil-pump does not work well, and Mr. Johnson knows that if the buyer learns about this, he will have to reduce the price by $250 (the cost of fixing the pump). If Mr. Johnson doesn't tell the buyer, the engine will overheat on the first hot day, resulting in damages of $250 for the buyer. Being winter, the only way the buyer can learn about this now is if Mr. Johnson were to tell him. Otherwise, the buyer will learn about it only on the next hot day. Mr. Johnson chose not to tell the buyer about the problems with the oil pump. In your opinion, Mr. Johnson's behavior is (please circle one): completely fair; fair; unfair; very unfair.
>
> What would your answer be if the cost of fixing the damage for the buyer in case Mr. Johnson does not tell him is $1,000 instead of $250? Mr. Johnson's behavior is (please circle one): completely fair; fair; unfair; very unfair.

Although there was no difference between the two scenarios in terms of the seller's payoffs, the buyer's cost increases from $250 to $1,000. I used both a between-subjects design (i.e, "what would be. . . "), with $N = 50$ students answering each question, and a within-subjects design (i.e., the participants answered the question for both parameters as they are presented above); again $N = 50$. The students' responses are presented in Figure 3.

The difference between the answers to the first and second question in the between-subjects design is significant ($p < .05$).[11] Inspection of the within-subjects design shows a large difference in choices. In the $250 cost question, 70 percent of the participants chose "unfair" and 18 percent chose "very unfair." In

[11] Using both Kolmogorov-Smirnov and Wilcoxon rank-sum tests.