An experimental analysis of gender discrimination in a credence goods market

Maxime Perodaud* Nobuyuki Hanaki† Takashi Yamada‡

Abstract

Field economic interactions always involve mutual knowledge about the distinctive personal characteristics of economic players. In particular, anecdotal evidence that female consumers being more likely defrauded in markets for car repair motivates the present study. In a 2×4 design, we experimentally investigated how (i) displaying information about the counterpart's gender and (ii) framing instructions that emphasize a specific car repair market context affect outcomes in markets for credence goods. A two-player game in four steps reproduced the conditions under which credence goods are provided. Overall, our analysis suggests that both market context and gender are of fundamental importance to explain the under-provision dimension of fraud in such markets.

Keywords: Credence goods, Discrimination, Fraud
JEL Codes: D82; J16

This version: December, 12 2021
Declarations of interest: none.
Acknowledgments: We thank the audiences at the Winter School on Credence Goods 2018 in Innsbruck, the ASFEE 2018 in Nice, the ASFEE 2019 in Toulouse and internal seminars at Université Côte d'Azur for helpful comments. Special thanks are due to Mira Toumi and Imen Bouhlel for assisting in the experiment. Financial support from the JSPS Fund for the Promotion of Joint International Research (Fostering Joint International Research, 15KK0123) and Grant-in-aid for Scientific Research (18K19954, 20H05631), as well as the Joint Usage/Research Center at ISER, Osaka University, are gratefully acknowledged.

*corresponding author: Université Côte d'Azur, GREDEG, CNRS, France. Email : maxime.david-perodaud@gredeg.cnrs.fr
†Institute of Social and Economic Research, Osaka University, Japan
‡Faculty of Global and Science Studies, Yamaguchi University, Japan

1 Introduction

Field interactions often – if not always – involve mutual knowledge of the distinctive personal characteristics of economic players such as ethnicity, age or gender. These characteristics can be observed or deduced at first glimpse. In this paper, we experimentally study the relevance of one of these characteristics, gender, in a credence goods environment.

The notion of credence goods was introduced by Darby and Karni (1973). Defined as an economic situation where an expert knows more about the type of good or service the consumer needs than the consumer himself, credence attributes make it difficult for consumers to assess the utility of the good or service even after consuming it. This is due to large informational asymmetries (generated by the specific expertise required or by the high cost of acquiring information) between the two sides of the market. As a result, after performing a diagnosis, experts can take advantage of the situation by exploiting the information asymmetry and thereby, can defraud customers. Dulleck and Kerschbamer (2006) identify several types of problems (commonly called "fraud") on markets for credence goods related to provision and charging behaviors: under-provision, over-provision, and overcharging. Table 1 provides a definition for each type of behavior described and its consequences. These situations arise because consumers have no choice but to rely on the experts' advice and provision.

This paper deals with gender discrimination in markets for credence goods. Discrim-

Table 1: Types of fraudulent behavior

Problem	**(In)Efficiency**	**Description**
Under-provision	Inefficient provision	Consumers receive products or services that are insufficient to satisfy their needs.
Over-provision	Inefficient provision	Consumers receive more expensive products or services than those sufficient to satisfy their needs.
Overcharging	Deters consumers from trading in the future	Consumers get charged for products or services they have not received.

ination can be defined as a phenomenon where a particular group is treated differently because of its salient characteristic (most of the time, this particular group is treated in a worse way, otherwise the term positive discrimination is used). For example, anecdotal evidence[1] suggests that female consumers are more likely to be defrauded by sellers in markets for car repair. Note that in many (off-line) market interactions, people usually know whom they are interacting with. Therefore, the gender characteristic of the economic partner is a variable that is known by both the seller and the consumer. However, the existing experimental studies of credence goods markets do not address the impact of gender pairing on the outcomes. Another empirical result related to credence goods markets can be found in Tsugawa et al. (2017). In their paper, the authors show that patients treated by female doctors have lower 30-day mortality and readmission rates than those treated by male doctors. If this type of behavior is effectively found to be caused by the gender of agents (sellers and consumers), the fundamental question of discrimination deserves attention to allow public policy makers to address it. This is what motivated our study.

In this paper we designed a controlled laboratory experiment to examine the effect of (i) the gender of participants and (ii) the market environment. In particular we address the following research question: *Is the extent and the type of fraud that occurs in experimental credence goods markets related to the agent's gender and the context of the transaction?* The experiment - which heavily borrows from Dulleck et al. (2011) - reproduces the conditions under which credence goods are provided. It consists of a four-step game in which two types of agents (sellers and consumers) are asked to interact with each other and information about the other participant is displayed. More specifically, in the within-subject variation, we let participants notice the gender of the other participant with whom they were matched. In the between-subject variation, we manipulated the market environment by providing neutral or contextual instructions to the participants.

The results of our experiment, taken together, suggest that context plays an important role in gender discrimination: males are less likely to face under-provision in the car-

[1]Belsky Gary and Llosa Luis Fernando "Watch out: Car-Repair Crooks Have Some New Tricks Up Their Grimy Sleeves," *Money*, June 1996, Periodical, Vol. 25, Issue 6, p. 172

repair context. We also found that the presence of a car-repair context makes female (male) sellers less (more) likely to under-treat compared to our baseline treatment without context. Such a context-dependent gender difference was not observed for over-treatment. These results suggest that the extent of gender discrimination in credence goods markets depends on both the specificity of the transaction and the type of fraud.

The remainder of this paper is organized as follows. In section 2, we briefly present the related literature and how we aim to contribute to it. Section 3 describes the experimental design and theoretical predictions. We present the results and the methodology employed in Section 4. Section 5 discusses the results and concludes.

2 Related works

The present study is related to two different branches of economic literature. First, it is closely linked to literature on (i) gender differences in preferences and their impact on economic outcomes and (ii) gender discrimination. Second, it is also related to the literature – and for the most part the experimental one – on credence goods markets. Below, we summarize these branches of the literature highlighting the novelty of our study.

As far as we are aware, only two studies have jointly dealt with these two topics. First, Agrawal et al. (2019) experimentally investigated the role played by the gender of experts in the choice of incentive structure and the provision of credence goods. Participants playing the role of experts faced a dilemma between maximizing their earnings (or minimizing their effort) and helping their customers by their actions in a real-effort task. Prior to performing the real-effort task, participants (playing the role of experts) had to choose from among three types of payment schemes: capitation, fee-for-service or salary. Controlling for the expert's performance, ability and self-confidence, the authors found no gender differences in preferences for variable rate payment schemes but did find differences in earnings. The choice of variable rate payment across genders departs from the prior literature, see for instance, Dohmen and Falk (2011). The differences in earn-

ings were due to the number of services provided. Under fee-for-service schemes, men listed more services than women and this accounted for higher earnings in male population. However, evidence presented in the Agrawal et al. (2019) paper only concerns the causal relationship between the gender of service provider, the selection of a payment and the provision of credence goods. Our investigation adds another dimension by providing information about the gender of both the service provider and the receiver.

Second and probably more closely related to our study, the paper of Balafoutas et al. (2017) addresses, among others, the question of gender discrimination in the market for taxi rides in Athens. Although the authors found no difference in over-treatment indices across gender, substantial differences were observed in charging behavior. The results of the experiment show that female passengers were more likely (about 23% more) to be overcharged than male passengers in the control treatment. However, when the passenger explicitly informed the taxi driver that the fare would be reimbursed by the employer, no gender difference was observed (overcharging rates were similar across gender in this case). In contrast to the above study, we consider not only the gender of the buyer but also that of the seller. Our case is thus more specific as gender pairing of players is central to our design and represents the within-subject variation of our treatments.

2.1 Gender in the economic literature

2.1.1 Gender differences

A considerable amount of economic literature has documented many differences in preferences between men and women. These results point out that women exhibit a greater aversion to financial risk (Booth and Nolen, 2012; Charness and Gneezy, 2012; Booth et al., 2014; Filippin, 2016) and tend to flee competitive environments by preferring deterministic payment schemes such as piece-rates (Niederle and Vesterlund, 2007; Andersen et al., 2013; List and Gneezy, 2014; Kuhn and Villeval, 2015). These experimental studies also identified that women had a more pro-social attitude (Andreoni and Vesterlund, 2001; Kamas and Preston, 2015; Balafoutas et al., 2012), a higher degree of belief in reciprocity (Chaudhuri and Gangadharan, 2007; Dittrich, 2015), and were considered to

be more trustworthy (Alesina and La Ferrara, 2002; Buchan et al., 2008). Nevertheless, although these studies investigate the role of the decision maker's gender, most of them do not investigate the effect of gender pairing which can produce hitherto unknown effects which can be different from gender effects. For instance, Sutter et al. (2009) analyzed the behavior of participants in a bargaining game called power-to-take game (Bosman and Van Winden, 2002) by varying the gender pairing. Their results indicate that the gender of participants has no effect alone, but it did when gender combinations were considered: the take and destruction rates were lower in mixed gender pairing than in same gender pairing.

2.1.2 Discrimination

This paper also builds on the literature on (gender) discrimination. This branch of literature is very well developed as it applies to the labor market. Two papers are especially relevant to our topic. First, the paper of Bertrand and Mullainathan (2004) investigated the prevalence of ethnicity-based discrimination in callbacks for a job application. The authors showed strong evidence of racial discrimination against CVs displaying Afro-American sounding names. Also, by varying the quality of the CVs they sent to employers, the authors showed that an increase in the quality of the CV had a far lower marginal effect on the number of callbacks received with Afro-American sounding names compared to white sounding names.

More recently, Gneezy et al. (2012) offered an overview of several case studies in which discrimination was highlighted. In each of the market situations examined, discrimination was prevalent. Questioning its underlying motivations, the authors suggested two complementary explanations depending on whether the object of discrimination could be perceived as controllable (e.g. obesity or homosexuality) or not (e.g. race, sex or physical disabilities) by the individual. In the former case, the data supported evidence for a taste-based discrimination (Becker, 1971) while in the second case when the object of discrimination was perceived to be out of control, statistical inference (Phelps, 1972) appeared to drive discrimination.

This strand of literature finds a significant association between gender of economics agents, gender pairing and economic decisions. It also provides insights into the motivations that drive decision-makers to discriminate. However, the extent of fraud as well as its determinants in the four possible gender constellations are still an open question. Thus, to complement these studies, we provide an analysis of how mutual observation of gender of agents affect fraudulent provisions of credence goods.

2.2 Credence goods studies

Regarding the literature on credence goods, the framework developed by Dulleck and Kerschbamer (2006) offers a comprehensive theoretical analysis of credence goods. In order to be efficient, these markets need to be regulated by such institutional factors such as liability[2] or verifiability[3] (although one of them is theoretically sufficient to ensure full efficiency). In a controlled laboratory experiment, Dulleck et al. (2011) explored the role of these institutional factors under different market conditions. The results show that institutional factors do not have the predicted role: while liability does indeed protect consumers against over-pricing and offers high efficiency rates (above 80%), verifiability is only partially effective in preventing over-provision. Under verifiability, efficiency rates were low and close to those observed in the treatment without it. According to the authors, the heterogeneity of distributional preferences explains this result: anti-social sellers do not post the predicted prices in accordance with standard selfish preferences. In a later paper, Kerschbamer et al. (2016) confirm that the level of efficiency in markets under verifiability is positively correlated with the presence of pro-social and efficiency-loving sellers within these markets. They conclude and report the following policy implications which suggest that differences in preferences contribute to determine fraudulent outcomes:

"Instead of choosing doctors, mechanics or computer specialists exclusively

[2]The expert must provide a sufficient quality, which therefore prevents under-provision.
[3]The expert must charge for the quality provided, which therefore prevents overcharging.

according to their training, customers or their representatives should also take into consideration the attitudes of these experts to their customers. [...] Public policy might step in here, for instance, by screening applicants for particular jobs [...]. A CV featuring an impressive track record of volunteer work might well act as a screening device" (Kerschbamer et al. (2016) p. 415).

However and despite the close connection that exists between preferences and the gender of economic agents, these approaches did not address the question of gender to explain the failure of verifiability.

Several other studies have explored the overpricing dimension of fraud (Sülzle and Wambach, 2005; Liu, 2011). More recently, surveys, laboratory and field experiments have contributed to the analysis of markets for credence goods. In these field experiments, some determinants – such as the informational distance (necessarily implying a lack of knowledge - Balafoutas et al. (2013)) or a low socio-economic status (Gottschalk et al., 2018) – have been found to be related to a higher level of fraud. See, Kerschbamer and Sutter (2017) and Balafoutas and Kerschbamer (2020), for a survey of the experimental literature.

We add to this literature on credence goods with a laboratory experiment on the mutual gender observation of the other players' gender in neutral and contextual frameworks.

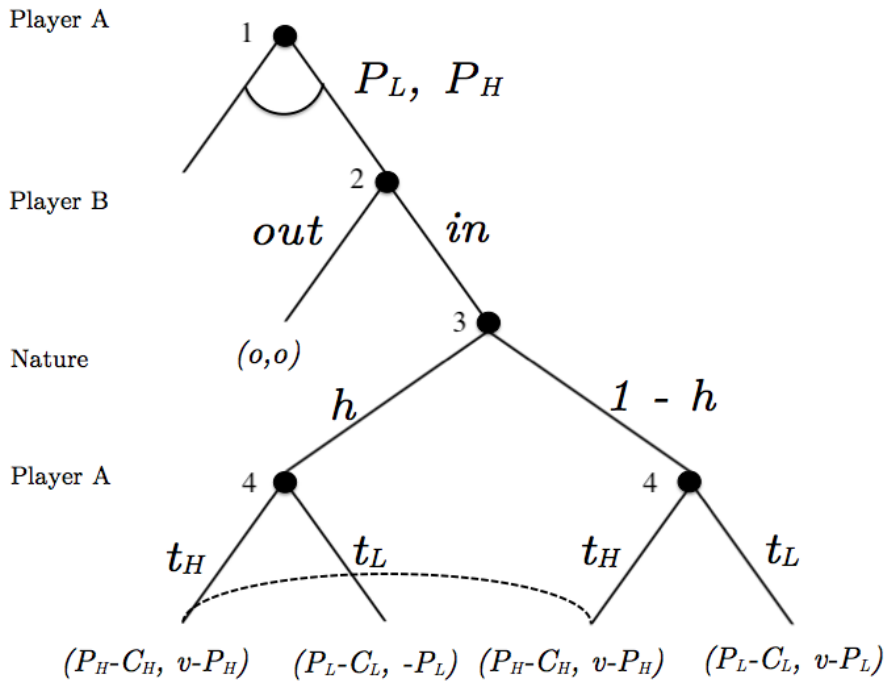
3 Experimental design

3.1 Experimental Market

In the main task of our experiment, participants were asked to interact in a fictitious market. This specific market reproduces the conditions under which credence goods are provided. The market was developed in a theoretical analysis by Dulleck and Kerschbamer (2006) and first studied experimentally by Dulleck et al. (2011). Considering

the interaction between verifiability and social preferences previously indicated, we decided to study how the market functions under the institutional factor of verifiability. Making actions verifiable means that consumers are able to observe *ex post* the quality they received. Hence, they are protected against overcharging.[4] Figure 1 contains an extensive-form representation of the game by Dulleck et al. (2011).

Figure 1: Market and sequence of the game



As in Dulleck et al. (2011), each consumer (he) knows that he has a problem but he does not know how severe it is (it can be major or minor with probability h and $(1-h)$, respectively). The problem must be solved by a visit to an expert (a seller), a recommendation, and a provision. The seller has to provide one of two types of quality (respectively denoted as t_L and t_H for low and high quality treatments). The low quality only solves the minor problems while the high one solves both minor and major ones. If the quality is sufficient to solve the problem, the consumer receives a payoff of v, otherwise a payoff of 0, from which the price paid must be subtracted. The two qualities come at

[4]Although verifiability should theoretically work as a safeguard, sellers are still allowed to provide a quality that is insufficient or useless to solve the problem.

two different costs for the expert (with $C_L < C_H$). Costs are fixed and observable by all players before the first move, i.e., before the seller posts a pair of prices. A game proceed as follows:

Node 1 - Player A (seller): Each seller posts a price for each of two qualities ($P_L \leq P_H$).

Node 2 - Player B (consumer): Each consumer observes the prices posted by the seller, and the characteristics of the seller (age, gender, and field of study) he is matched with, and decides whether to interact with the seller or not. If the consumer chooses not to interact (out), both the consumer and the seller get the payoff o and the game ends. If the consumer chooses to interact (in), the game continues (nodes 3 and 4).

Node 3 - Nature: Nature randomly determines the severity of the problem each consumer faces (H or L) according to the probability h and $(1-h)$.

Node 4 - Player A (seller): Each seller (she) observes the severity of the problem and the characteristics (age, gender, and field of study) of the consumer. She decides which quality to provide (and because of the verifiability assumption, the price of the chosen quality is charged to the consumer).

Payoff and feedback - Players A & B: At the end of each period, both the seller and the consumer receive feedback about their own payoff. From this, the consumer can infer whether the provision received is sufficient.

Except for the fact that participants were informed of the characteristics of their counterpart, which was new to our experiment, we followed Dulleck et al. (2011) including as regards the parameter values. Sellers could post integer prices between 1 to 11 ($P_L \leq P_H$). The ex-ante probability of a consumer having a major and a minor problem was the same ($h = 1 - h = 0.5$). The cost for providing the minor (major) quality was C_L= 2 (C_H= 6). The consumer derived a utility of $v = 10$ when the quality received was

sufficient. If no trade took place, both the consumer and the seller received a payoff of $o = 1.6$.

We used the perfect Bayesian equilibrium as the solution concept. To derive it, we assumed a risk neutral consumer and a risk neutral seller. We considered three different types of price vectors:

- An under-provision price markup, i.e., $P_L - C_L > P_H - C_H$.
- An over-provision price markup, i.e., $P_H - C_H > P_L - C_L$.
- An equal price markup, i.e., $P_L - C_L = P_H - C_H$.

In the case of an under-provision (over-provision) price markup, the seller provides the low (high) quality once a consumer decides to interact. Under an equal price markup, the seller is indifferent between the two qualities, and, in this case, we assume that she provides the appropriate one.

Knowing the costs and observing the prices posted, the consumer can infer the seller's incentives to provide one quality over another. Anticipating the quality to be provided, the consumer is willing to interact with the seller only if he can derive a higher utility from doing so, than not interacting. With a risk neutral consumer, this means that he enters the market if he can expect his payoff to be no less than $o = 1.6$. Therefore, for a consumer to enter given the posted price vectors, we need:

- With an under-provision price markup: $0.5(-P_L) + 0.5(10 - P_L) = 5 - P_L \geq 1.6$, i.e., $P_L \leq 3.4$ (and $P_H < P_L + 4$).
- With an over-provision price markup: $(10 - P_H) \geq 1.6$, i.e., $P_H \leq 8.4$ (and $P_L < P_H - 4$).
- With an equal price markup: $0.5(10 - P_H) - 0.5(10 - P_L) = 10 - (P_H + P_L)/2 \geq 1.6$. Because with the equal price markup $P_L = P_H - 4$, we have $P_H \leq 10.4$.

Assuming that the consumer enters the market in a case of indifference, the seller posts the maximum prices that satisfy the above conditions. The expected payoff for the seller is

then 1.4 for under-provision, 2.4 for over-provision, and 4.4 for equal mark up. Therefore, under the perfect Bayesian equilibrium, the seller posts $P_H = 10.4$ and $P_L = 6.4$: the consumer enters the market, and the seller provides the right quality. The experimental evidence described above, however, shows that participant behaviors deviate substantially from these equilibrium predictions.

3.2 Treatments

We varied the gender (of sellers and buyers) and contextualization in our experiment by employing both a within and between-subject design. In the within-subject dimension, the gender of the players was revealed to the participants. Since it was a two-player game, we were able to analyze the behavior of the subjects according to the four possible gender constellations. Instead of revealing names or providing pictures of counterparts, we decided to directly reveal the gender by a clear binary information: he or she is either male or female. This allows us to draw inferences solely based on gender and avoid any confusion with the effects of variables such as ethnicity, attractiveness, visual expression, beauty, etc. In the between-subject design, contextual instructions, including the context of car-repair were provided to half of the participants while the remaining half received context-free neutral instructions. We chose to use terms related to the car repair market for the following two reasons. First, since our aim was to simulate conditions under which credence goods are provided, the market for car repair perfectly falls within this category. Second, the choice of this context was also motivated by its gender orientation. This context was chosen to better capture behavior in the field by reproducing a real-life economic transaction. The within-subject variations of gender were implemented as follows: each session included a total of 24 participants. These 24 participants were divided into 3 matching groups of 8 participants. Within a matching group, half of the participants were assigned the role of consumers and the remaining half were assigned the role of sellers during the session. We repeated the game 16 times, randomly re-matching a seller and a buyer each time within the matching group. An overview of the 2×4 experimental design is presented in Table 2.

Table 2: Experimental design

F-F	F-M	M-F	M-M
Condition without framing (Baseline)			
Seller is a female	Seller is a female	Seller is a male	Seller is a male
Buyer is a female	Buyer is a male	Buyer is a female	Buyer is a male
Condition with framing (Context)			
Seller is a female	Seller is a female	Seller is a male	Seller is a male
Buyer is a female	Buyer is a male	Buyer is a female	Buyer is a male

Without framing	With framing
Player A	Mechanic
Player B	Consumer
Type 1	Light breakdown
Type 2	Heavy breakdown
Action 1	Light repair
Action 2	Important repair

For instance, later in the paper the abbreviation Baseline-F-F stands for female seller and female buyer in the baseline condition (without framing).

3.3 Control tasks

The literature has documented many differences in preferences between men and women. As control variables, we considered social and risk preferences because (i) findings are robust[5] (Croson and Gneezy, 2009) and (ii) we believe they have economic implications in provision decisions considering the behavioral mechanisms described hereafter. On the one hand, risk-averse sellers are expected to provide an efficient quality to avoid impairing the trade relationship (that costs in the long run) and they post lower prices in order to avoid the risk of not interacting with a consumer for a profitable price[6]. Likewise, pro-social sellers are also expected to provide an appropriate quality because an inefficient quality decreases consumer payoffs that in turn decrease their own payoff. On the other hand, risk-averse consumers are expected to be reluctant to interact.

We measured social preferences using the SVO Slider Measure (random matching procedure, Murphy et al., 2011). Attitudes toward risk were elicited with the Holt and Laury (2002) multiple price list.

3.4 Experimental procedures

The experiment was conducted in the Laboratory for Experimental Economics (LEEN) in April and November 2018. It was computerized using z-Tree (Fischbacher, 2007). Participants were recruited among those registered in the ORSEE (Greiner, 2015) database.

[5]Notably, Kamas and Preston (2015) show that social preferences are of importance to explain economic outcomes. The authors show that women choose egalitarian payoffs more often than men. In a broader perspective, Balafoutas et al. (2012) reported no significant difference between men and women in the willingness to choose a competitive rate payment to enter a tournament after controlling for social preferences and risk attitudes. This shows the crucial direct link between gender and preferences.

[6]This will also be mentioned in the formulation of our hypotheses.

In total, 8 sessions were conducted, each with 24 participants. None of them participated more than once. Four sessions implemented the condition with neutral instructions and another four sessions implemented the condition with framed instructions. All 192 participants were the students in economics and between 18 and 26 years old. The absence of variation in the Field of study and Age variables reduced potential experimenter demand effects (Zizzo, 2010). The experiment was designed to determine whether the gender of participants was relevant information to the seller's likelihood to defraud consumers. Although *Gender* was the key variable we were interested in, we decided to add two more variables to avoid an emphasis on gender. During the experiment and for each buyer-seller pair, the opponent's age and field of study were displayed on the screen (see Figure 2 for a screenshot in the treatment without framing (the original was in French)). However, these two additional variables – thanks to strict control of the recruitment process – did not vary. It was therefore possible for us to make comparisons based on the *Gender* variable with all things equal otherwise.

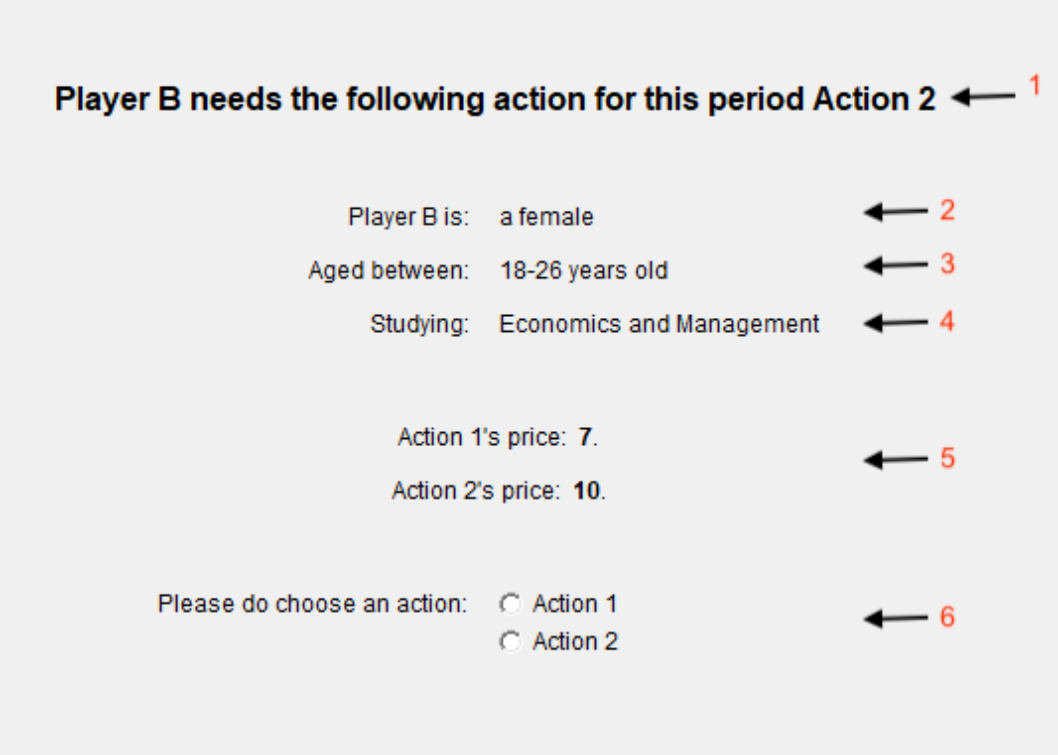
Table 3 provides summary statistics of the participants' characteristics. Just under half the sample was female (48%) reflecting our equal gender-based recruitment.[7] Figure 3 presents the participants' SVO scores.[8]

On their arrival at the laboratory, the participants were seated randomly at each computer. At the beginning of each session, they were asked to fill out an idiosyncratic questionnaire (including questions about gender, age and degree). Answers from this questionnaire were extracted and displayed on the participants' screen in the credence goods game. Then, they completed the SVO and Holt & Laury risk elicitation control tasks. Once matching groups of 8 participants were formed, instructions of the main

[7]Despite of our efforts, due to no-show, we could not achieve the complete balance in the gender in every session. Note that because participants were randomly divided into three matching groups of 8 participants (4 sellers and 4 buyers) in each session, we could generate various combinations of gender pairing across 16 repetitions within each matching group. However, as one can observe in Table 5, this resulted in a varying number of observations across four gender pairings in both Baseline and Context.

[8]The bars indicate the proportion of subjects expressing a corresponding SVO angle. This histogram suggests a multimodal distribution of SVO types. Two large clusters can be observed: one in the prosocial category with a shift to the left (toward individualist) and a second in the individualist region with a shift to the right (toward pro-social).

Figure 2: A z-Tree interface screenshot of the game's node 4 in the Baseline treatment



Screenshot for player A (the seller), after player B has accepted to interact with him. On the right-hand side, numbers from 1 to 6 correspond to the sequence in which information appears on the screen. At this point in the game, player A has to make a decision (Action 1 or 2 – ⑥). If we go back to the top of the figure, before player A makes his decision, he can observe the following information about player B: the action needed to solve the problem (①), gender (②), age (③), field of study (④). A reminder of player A's posted prices is also displayed (⑤). Players A sees a similar screen at node 2 when making the decision to enter the market (except for information ①).

task were read aloud. Participants answered a control questionnaire to ascertain their understanding of the game.[9] Finally, the credence goods game described in the market subsection was played. A session lasted one hour on average including instructions, questionnaire, control tasks, and payment. During the experiment, we used the fictitious ECU currency that was converted to euros at the rate of 1 ECU = 0.20 euro. Participants were paid confidentially at the end of the experiment. Average payment was 16.86 euros (5 euro show-up fee included).[10] A difference in profit close to 4 euros to the advantage of the seller was observed, which can be explained by the fact that once they decided to interact, consumers had no choice but to rely on the seller's decision and thus could be

[9]Appendix F provides an English translation of the instructions and the control questionnaire.

[10]One participant earned 2.65 euros in the experiment. He received 5 euros in accordance with the LEEN remuneration policy and was not informed that he had gained less than 5 euros.

Table 3: Participant characteristics

	Mean	S.d.	Min	Max	# Independent obs.
Female	0.48	0.5	0	1	192
Female (seller role)	0.44	0.5	0	1	96
Altruist	0.005	0.07	0	1	192
Prosocial	0.51	0.5	0	1	192
Individualist	0.48	0.5	0	1	192
Competitor	0	0	0	0	192
Risk aversion	4.2	1.97	0	10	192
Profit (euros)	16.86	3.73	2.65	24.90	192
Profit seller role (euros)	18.79	2.52	13.24	24.90	96
Profit consumer role (euros)	14.93	3.75	2.65	24.43	96

exploited.

4 Hypotheses and Results

4.1 Hypotheses

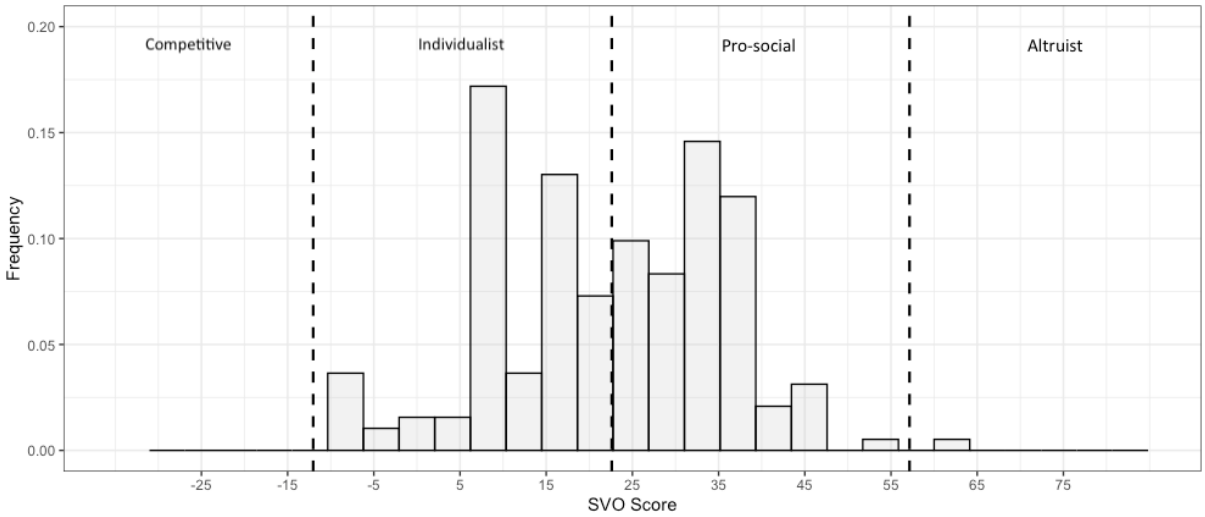
The following hypotheses were developed on the basis of the previous discussion on (i) gender effects (that may influence player beliefs) and (ii) discrimination phenomena.

Seller behavior

As noted above, the literature suggests that women are financially more risk averse than men. Thus our first hypothesis was related to gender differences in the price posting step. Given that sellers were free to post the prices they wished (as long as $P_L \leq P_H$), a risk averse seller could try to convince the consumer to enter by posting lower prices that satisfy $\Pi_H = \Pi_L$.

H_1: A risk averse seller will try to avoid the risk of not interacting with a consumer for a profitable price. To do so, he will post lower prices.

Figure 3: Participant SVO scores



Moreover, one should expect that the more prosocial a seller is, the more the provision behavior is honest (Kerschbamer et al., 2016).

H_2: A seller with pro-social distributional concerns is less likely to defraud consumers.

As mentioned above, the literature emphasizes the correlated role of distributional preferences and gender in determining economic outcomes (Kamas and Preston, 2015; Balafoutas et al., 2012). Moreover, the literature suggests that women are considered to be more trustworthy (Alesina and La Ferrara, 2002; Buchan et al., 2008).

H_3: Because women tend to behave more pro-socially and tend to prefer egalitarian payoff, women sellers are less likely to defraud buyers than male sellers.

We expected the specific car-repair market environment to induce a belief associated with gender stereotypes. Given the belief that women are less knowledgeable about car repair than men, it was hypothesized that, in this particular case, women would suffer more from fraudulent behavior upon deciding to interact with sellers. In the treatment without framing, i.e., in the absence of oriented beliefs, no discrimination was expected occur.

H_4: In the framed condition, female buyers are more likely to be defrauded than male buyers.

Buyer behavior

According to the H_3 hypothesis, buyers expect to be defrauded less by female sellers than by male sellers. Anticipating this, buyers could therefore be more willing to enter the market if they are informed that the seller is a woman, and less willing if they are informed that the seller is a man.

H_5: Buyers are less willing to enter the market when they are informed that the seller is a man rather than a woman.

Expecting both H_3 and H_4, female buyers should be less likely to interact with male sellers than with female sellers under the framed condition compared to the non-framed condition while there should be no such difference for male buyers. The empirical study of Ellingsen et al. (2013) complementing the work of Croson and Gneezy (2009) argues for a gender-dependent sensitivity to the context. More precisely, they found that the behavior of women was considerably affected by the social framing and the context of the transaction which justifies our hypothesis H_6.

H_6: The difference between male buyers and female buyers as regards their willingness to interact based on the sellers' gender is greater in the treatment with framing than in the one without. More precisely, since women are more responsive to the context of the experiment, they are less likely to enter the market in this condition than in the situation without.

Referring back to the research aim, which is to investigate the effect of the participants' gender and the market environment on economic outcomes, this section presents the statistical procedures and describes the results.

4.2 Variables

Table 4 defines and provides an overview of the variables used in the following analyses and the way they were built.

Table 5 displays summary statistics of these variables. The table is divided into

Table 4: List of variables

Variables	**Type**	**Description**
Under-provision*	Categorical	1 if the quality provided is lower than the quality needed, 0 otherwise
Over-provision*	Categorical	1 if the quality provided is higher than the quality needed, 0 otherwise
Efficient provision*	Categorical	1 if the seller provides the right quality, 0 otherwise
Interaction	Categorical	1 if the buyer enters the market, 0 otherwise
{6,8}	Categorical	1 if the seller posts the price vector {6,8}, 0 otherwise
P_L	Continuous	Low-quality price posted by the seller
P_H	Continuous	High-quality price posted by the seller
Π_L	Continuous	Seller's profit from providing the low quality : $P_L - C_L$
Π_H	Continuous	Seller's profit from providing the low quality : $P_H - C_H$
$\mid \Pi_H - \Pi_L \mid$	Continuous	Deviation for the incentives to provide the appropriate quality
$\Pi_L - \Pi_H$	Continuous	Incentives encouraging under-provision: $(P_L - C_L) - (P_H - C_H)$
$\Pi_H - \Pi_L$	Continuous	Incentives encouraging over-provision: $(P_H - C_H) - (P_L - C_L)$
Context	Categorical	1 if instructions are contextualized, 0 otherwise
Male	Categorical	1 if the seller is male, 0 otherwise
Buyer Male	Categorical	1 if the consumer is male, 0 otherwise
SVO angle	Continuous	SVO ring, the higher the more prosocial (from -7 to 61)
Risk aversion	Continuous	H & L risk elicitation method, the higher the more risk averse (from 0 to 10)

*: Conditional on *Interaction* = 1

4 panels to present the overall summary statistics and those broken down to various treatments. Panel (a) corresponds to the between-subject dimension of our experiment. It presents the mean values for the whole experiment and separately for the baseline condition (without framing) and context condition (with framing). Second, panel (b) corresponds to our within-subject dimension and presents the results according to the four possible gender couple constellations. Panel (c) presents the effect of gender pairing in the baseline condition and panel (d) in the context condition.

As can be seen in panel (a) of Table 5, all the variables seem to be sensitive to our context condition: compared to the baseline condition, buyers enter the market more (+10 pp on average)[11] and are on average less likely to suffer from mis-provision (mainly due to the large decrease in the frequency of under-provision decisions from 0.58 to 0.35). This, in turn, translates into higher profits for consumers (nearly twice as much) and thus lower profits for sellers. However, panel (b) shows no difference in the decision to enter the market (to interact) between the four gender constellations: ratios are very similar. Interestingly, we find that almost two-thirds of the consumers (from 0.60 to 0.66) were provided with the right quality. This result, however, hides wide variations between under and over-provision. For instance, the M-M gender pairing simultaneously combines higher ratios for under-provision (0.61) and lower ratios for over-provision (0.19), compared to the F-M gender pairing (0.37 for under-provision *vs.* 0.44 for over-provision). Regarding profits, further differences can be observed between the experimental conditions. In M-M and M-F configurations (i.e., where sellers are male) earnings are higher for sellers than in F-M and F-F ones (i.e., where sellers are female). But we find the opposite regarding consumer profits: consumers earn more from trading with women (although the difference is much smaller: on average 1.55 ECU in F-F and 1.48 ECU in F-M *vs.* 1.40 ECU in M-F and 1.18 ECU in M-M)).

Let us now turn to the mixed effect of gender pairing according to the two different

[11] Figure 4 presented in Appendix A shows that the average relative frequency of trade (all treatments) is around 70 percent at the beginning of the experiment and that it gradually decreases to just under 50 percent. This graph corroborates the findings of Dulleck et al. (2011) on the failure of verifiability to ensure efficiency and full trade volume as well as the learning process occurring during the experiment due to the experience of under-treatment.

Table 5: Descriptive statistics by treatment

Panel (a)

	All	**Baseline**	**Context**
Interaction[1]	0.53	0.48	0.58
Efficient provision[1]	0.63	0.58	0.67
Under-provision[1]	0.46	0.58	0.35
Over-provision[1]	0.29	0.26	0.31
Actually charged price[2]	6.64	6.88	6.45
Seller profit[2]	2.35	2.51	2.19
Consumer profit[2]	1.40	0.93	1.85
	{6,8}	{6,8}	{6,8}
Most chosen $\{P_L, P_H\}$	{5,8}	{7,8}	{5,8}
	{4,8}	{6,10}	{4,8}
# Observations	1536	768	768

Panel (b) **All**

	M-M	**M-F**	**F-M**	**F-F**
Interaction[1]	0.51	0.55	0.52	0.53
Efficient provision[1]	0.61	0.65	0.60	0.66
Under-provision[1]	0.61	0.45	0.37	0.41
Over-provision[1]	0.19	0.28	0.44	0.29
Actually charged price[2]	6.34	6.82	6.75	6.62
Seller profit[2]	3.2	3.22	2.61	2.90
Consumer profit[2]	1.18	1.40	1.48	1.55
	{6,8}	{6,8}	{6,8}	{4,8}
Most chosen $\{P_L, P_H\}$	{5,8}	{5,8}	{4,8}	{5,8}
	{6,6}	{7,8}	{5,8}	{5,10}
# Observations	396	468	340	332

panel (c) **Baseline**

	M-M	**M-F**	**F-M**	**F-F**
Interaction[1]	0.47	0.53	0.44	0.46
Efficient provision[1]	0.55	0.68	0.44	0.60
Under-provision[1]	0.67	0.42	0.7	0.52
Over-provision[1]	0.21	0.24	0.4	0.23
Actually charged price[2]	6.64	7.05	6.49	7.28
Seller profit[2]	3.56	3.48	3.11	3.80
Consumer profit[2]	0.78	1.31	0.80	0.78
	{6,8}	{6,8}	{6,8}	{6,7}
Most chosen $\{P_L, P_H\}$	{6,6}	{7,8}	{6,10}	{6,10}
	{7,8}	{5,8}	{11,11}	{7,8}
# Observations	212	220	172	164

Panel (d) **Context**

	M-M	**M-F**	**F-M**	**F-F**
Interaction[1]	0.54	0.57	0.60	0.60
Efficient provision[1]	0.67	0.62	0.71	0.69
Under-provision[1]	0.53	0.47	0.10	0.28
Over-provision[1]	0.17	0.31	0.46	0.32
Actually charged price[2]	6.04	6.61	6.94	6.12
Seller profit[2]	2.84	3	2.25	2.22
Consumer profit[2]	1.63	1.48	2.18	2.31
	{6,8}	{6,8}	{4,8}	{4,8}
Most chosen $\{P_L, P_H\}$	{5,8}	{5,8}	{6,8}	{5,8}
	{4,8}	{5,7}	{5,8}	{5,10}
# Observations	184	248	168	168

Couples' appellations are shortened according to Table 2. The first (second) letter refers to seller's (buyer's) gender. The letter "M" ("F") stands for Male (Female). Differences must be read vertically.

[1] Relative frequency. These frequencies are computed from different totals, thus they do not add up to one. First, they are conditional on consumer's choice to interact. Second, if a consumer is a high type, over-provision is not possible by definition and conversely if the consumer is a low type, under-provision is not possible by definition.

[3] In experimental currency units, mean per period.

market conditions. When performing a cross-comparison across gender pairing between treatments, we note that under the context condition (panel (d)) and compared to the baseline (panel (c)), female sellers more often provide the appropriate quality to male buyers (+27 pp, from 0.44 to 0.71). Moreover, the average frequency of over-provision is larger in the F-M configuration compared to others. Intriguingly, we observe that the treatment with context reduces under-provision from female sellers and especially toward male buyers (from 0.70 to 0.10). Regarding paid prices, even though some differences can be noticed in prices across the composition of pairs, no clear pattern is observed.[12]

In what follows, we consider first the sellers' behavior (section 4.3.1) and second, the consumers' behavior (section 4.3.2), by testing the hypotheses presented above. Since we have 192 subjects, this gives us 96 buyer-seller pairs (i.e. 48 in each condition) interacting over 16 rounds generating an unbalanced panel dataset of 1536 observations in total (see Table 5).[13] In our econometrics analysis, we first used a panel data random effects linear specification to analyze the determinants of posted prices. We then used panel data random effects probit regressions to investigate the likelihood of occurrence of the following: *under-provision*, *over-provision*, *efficient provision* and *interaction*. Given the endogenous nature of financial incentives in the choice of provisions (prices are posted by each seller at each period), we employed a two-stage probit least squares (2SPLS) estimator.[14] We chose our instruments with the objective of eliminating the correlation between these financial incentives and our variables of interest (i.e., treatments). The first stage consists in exclusively using variables from the price posting stage (seller characteristics and market environment) to predict markup difference. In the second stage, the dependent variables *under-provision*, *over-provision*, and *efficient provision* were estimated from the sellers' point of view.[15] The *interaction* decision was estimated from

[12]Other results and evidences related to prices will be presented in Section 4.3.1.

[13]Unbalanced panel data structure is due to (i) either the unwillingness of consumers to enter the market or (ii) the consumer's type that is determined by nature. We provide examples of (i) and (ii).
(i) : If a consumer decides to not enter the market, no quality can be given at this period and therefore no behavior can be observed.
(ii) : If a consumer is a high type, over-provision is not possible by definition.

[14]In Appendix E, we provide results of regressions for the *under-provision*, *over-provision*, and *efficient provision* variables using a probit analysis without instrumenting for financial incentives in predicting the occurrence of behaviors.

[15]These three estimations deal with unbalanced panel data structure. See footnote 13.

Table 6: Data analysis methodology

	Model	**Instruments**	**Unit of observation**
P_L	Linear panel data random effects	*Not applicable*	Seller's decision
P_H	Linear panel data random effects	*Not applicable*	Seller's decision
Under-provision	Probit panel data random effects	$\Pi_L - \Pi_H$	Seller's decision
Over-provision	Probit panel data random effects	$\Pi_H - \Pi_L$	Seller's decision
Efficient provision	Probit panel data random effects	$\mid \Pi_H - \Pi_L \mid$ and $\{6, 8\}$	Seller's decision
Interaction	Probit panel data random effects	*Not applicable*	Consumer's decision

the consumers' point of view. For each model, we provide two goodness-of-fit measures: the McFadden R-squared value[16] and the log-likelihood ratio (LLR) value.[17] Pairings of female sellers and buyers in the non-framed treatment constitute the benchmark. Table 6 summarizes the methodology used for analyzing the data.[18]

4.3 Findings

4.3.1 Seller behavior

Prices and markup

Let $\{P_L, P_H\}$ be a price vector posted by sellers in our experiment. Our sample is characterized by a very high prevalence of price vector {6,8} and {5,8} capturing about quarter of all observations (respectively 14 and 8 percent). The equilibrium equal markup price vector {6,10} is posted in only 4 percent of cases.[19] Given that each quality has a fixed cost ($C_L = 2$ and $C_H = 6$), these prices vectors reflect financial incentives that encourage under-provision ($6 - 2 > 5 - 2 > 8 - 6$ i.e., $\Pi_L > \Pi_H$). About 62 percent of

[16] Computed as the $\frac{\text{Log-likelihood value intercept-only model} - \text{Log-likelihood value model considered}}{\text{Log-likelihood value intercept-only model}}$.

[17] Computed as 2(Log-likelihood unconstrained model − Log-likelihood constrained model). We used the LLR value for joint hypothesis testing.

[18] As noted in footnote 7, in each session, participants were divided into matching groups consisting of 8 subjects with equal number of buyers and seller. Thus, there are 24 independent matching groups. Our analyses, however, do not explicitly take this data structure into consideration because of the two-stage estimation procedure we employ as discussed above.

[19] Figures 5, 6, and 7 and Table 13 in Appendix B plot actual posted prices in each treatment in order to give the reader a more precise overview of the distribution of prices.

Table 7: Panel random effects estimations: prices determinants

	P_L	P_H
(Intercept)	6.35^{***}	9.15^{***}
	(0.16)	(0.13)
Male	-0.30^{**}	-0.54^{***}
	(0.12)	(0.10)
Context	-2.09^{***}	-0.43^{***}
	(0.12)	(0.10)
Male × Context	1.05^{***}	0.27^{*}
	(0.17)	(0.14)
SVO angle	-0.00	-0.01^{***}
	(0.00)	(0.00)
Risk aversion	0.04^{*}	-0.03
	(0.02)	(0.02)
R^2	0.20	0.04
Adj. R^2	0.20	0.04
Num. obs.	1536	1536

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

price vectors are of the under-provision type. It should be noted that under-provision is worse for consumers than over-provision in our experiment, because the former always results in a negative payoff for them.

However, the fact that the price vector {6,8} is the most prominent confirms the idea that sellers are concerned about relative payoffs in accordance with Dulleck et al. (2011) and Kerschbamer et al. (2016). Even though this price vector gives sellers an incentive to under-treat and thus always provide the low quality in the equilibrium analyses, assuming they are selfish material payoff maximizers, if sellers provide the right quality, it results in an equal sharing of the gains from trade. Note that, with $P_L = 6$ and $P_H = 8$, the profit sellers and buyers make after providing/receiving the low (high) quality is 4 (2). We consider this possibility in our further analyses.

How do different treatments affect these pricing strategies? We find that context decreases prices, especially P_L: on average prices for low and high quality are respectively 6.25 and 8.52 in the non-framed condition *vs.* 4.76 and 8.22 in the framed condition.

Table 7 takes a closer look at the determinants of these prices. It clearly shows that context has a negative effect on the prices female sellers charge, both P_L and P_H, while

the effect is larger on P_L. For male sellers, context also has a negative effect, especially on P_L, but this negative effect is significantly smaller in magnitude compared to female sellers as one can observe by the positive and statistically significant coefficient of Male $\times$ Context. Thus, in the framed condition (with context), male sellers post higher P_L than female sellers.

Table 8: Panel random effects estimations

	$\Pi_L - \Pi_H$	$\mid \Pi_H - \Pi_L \mid$	$\Pi_H - \Pi_L$
	(1)	(2)	(3)
(Intercept)	0.17	1.69^{***}	-0.88^{***}
	(0.28)	(0.15)	(0.27)
Male	0.52^{**}	0.30^{**}	-0.73^{***}
	(0.22)	(0.12)	(0.23)
Context	-1.32^{***}	-0.67^{***}	1.22^{***}
	(0.23)	(0.12)	(0.23)
Male $\times$ Context	0.41	0.30^{*}	-0.20
	(0.31)	(0.16)	(0.30)
SVO angle	0.03^{***}	0.00	-0.00
	(0.01)	(0.00)	(0.01)
Risk aversion	0.13^{***}	-0.05^{**}	-0.06
	(0.04)	(0.02)	(0.04)
R^2	0.20	0.09	0.18
Adj. R^2	0.19	0.09	0.17
Num. obs.	377	811	434
RMSE		1.13	

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Before moving to the provision decisions, let us first comment on the results of price markups regressions shown in Table 8. These regressions serve as our first stage regressions for the provision decisions discussed later. For this purpose, we consider three separate regressions: (1) incentives to under-provide, (2) deviation from the equal markup and (3) incentives to over-provide. Therefore, for the regression (1), we consider only the cases where, at the provision stage, the consumer needed H to be treated (and to receive a positive payoff). Similarly, for regression (3), we consider only the cases where the consumer needed L. Regression (2) is based on all the cases where consumers decided to interact with the sellers. The baseline is a female seller interacting with a female consumer.

In column (1) of Table 8, the positive and significant coefficient of Male means that male sellers post different price vectors from female sellers such that it gives them significantly higher positive incentives to under-treat consumers ($\Pi_L - \Pi_H$) than female sellers. As already seen in Table 7 where context reduced P_L significantly more than P_H, context significantly reduces the incentives to under-treat. The effect of context on incentives to under-treat is not significantly different between male and female sellers as the non-significant coefficient of Male $\times$ Context shows.

Moving to column (3) of Table 8, we find that women post larger incentives to over-treat than men in our baseline situation without the context, as the negative and significant coefficient of Male shows. The positive and significant coefficient of Context indicates that female sellers post larger incentives to over-treat compared to the baseline situation. As in the case of under-provision price markup, the effect of context is not significantly different between men and women with respect to the size of over-provision price markup as indicated by the insignificant coefficient of Male $\times$ Context.

Finally, looking at column (2) of Table 8, contrary to the theoretical prediction that sellers post equal markup prices, a positive difference (in absolute value) between Π_H and Π_L is on average observed. The conditional mean (controlling for the SVO and degree of risk aversion) for F-F pairs in the benchmark treatment is 1.69 ECU. This deviation from the equal markup increases for male sellers to reach an average of 2 ECU in the benchmark. But it becomes significantly lower for women in the presence of context, with an average decrease of 0.67 ECU. For men, the positive and significant coefficients of Male $\times$ Context show that the effect of context is smaller on male sellers (average decrease of 0.37 ECU for them).

Moreover, we find that our control variables related to risk and distributional preferences appear to be predictive an under-provision price markup only. We find the positive and significant effect of risk and SVO in column (1). This indicates that the more pro-social and the higher risk averse a subject is the greater the incentives are to under-treat. In columns (2) and (3), it can be seen that SVO has at best a moderate impact on the way other types of incentives are posted.

In order to identify potentially more subtle connections between distributional preferences and outcomes, we present additional econometric specifications related to the impact of distributional preferences in Appendix C. To show that an inequality averse seller can provide the exact quality a consumer needs with the price vector {6.8}, we built a dummy variable taking the value 1 if the sellers posted these two prices, 0 otherwise. Since this variable is a dummy, a non-linear method would have been the most appropriate method of estimation. However, a two-stage probit least squares requires, in the first stage, a linear estimation. We therefore adopted the following method. We first used a random effects linear model for panel data to predict the the determinants of posting the price vector {6,8}. We then proposed a probit specification to ensure the robustness of the results found in the first step of the regression. The results discussed will be only those that appeared significant in both types of regression. Then, conditional on the linear estimation of this price vector, we used, in the second stage, a random effects probit regression to predict the likelihood of the *efficient provision* variable. Tables 14 and 15 jointly identify the positive relation between the Male variable and posting {6,8}. That is, males are more likely to post the {6,8} price vector. Conditional on this first stage, we did not observe substantial changes in the second stage of the estimation which we discuss further below.

Result 1: Incentives to provide one quality over another are gender and context-dependent. More precisely,

Result 1-a: In our baseline condition, men are more likely to post price vectors leading to larger under-provision incentives than women.

Result 1-b: In our baseline condition, women are more likely to post price vectors leading to larger over-provision incentives than men.

Result 1-c: Moreover, in the presence of context incentives to under-provide and deviation from the equal markup are lower (with a larger absolute deviation in the female seller's population than in the male population), incentives to over-provide are nevertheless greater.

Result 1 does not provide total support for H_1 and H_2. Let us now turn to the pro-

vision decisions. The results of the second stage regressions are summarized in Table 9.

Efficient provision – Regression results presented in column (2) of Table 9 suggest that male buyers are significantly less likely to receive the appropriate quality from female sellers than female buyers (negative and significant coefficient of Male Buyer). Moreover, the estimated coefficient of Male Buyer × Context is positive and marginally insignificant (p-value = 0.11). However, a joint-hypothesis test indicates that male buyers are marginally significantly (at 10%) more likely to receive an appropriate quality under the framed treatment compared to the baseline non-framed treatment.[20] Neither additional significant interaction effect with the seller's gender nor with the context was observed. At the moment, this result is not sufficient to provide full support for our third hypothesis but serves us as an intuition. However, as we have observed above, because there are two types of non-efficient provisions, under- and over-provision, these may be impacted differently by gender and context. We now analyze them separately.

Result 2: Male buyers are less likely to receive the appropriate quality than female buyers in our baseline treatment. However, there is suggestive evidence that male buyers are marginally more likely to receive the appropriate quality when the context is provided than without. For female buyers, we observe no significant effect of the context.

Under-provision – Column (1) of Table 9 identifies the determinants of under-provision. Although we accounted for the (negative) effect of context on $\Pi_L - \Pi_H$ in our first stage regression, context has an additional significant negative effect on the likelihood to under-treat in the case of female sellers. That is, context makes female sellers less likely to under-treat female buyers than when no context is provided. At the same time, the positive coefficient on Male × Context indicates that male sellers are significantly

[20] We used the likelihood ratio test. The reference estimations presented in Table 9 are unconstrained in the sense that we allowed coefficients to differ according to the gender of the seller, buyer and the market environment. Alternative constrained estimations where parameters are forced to be the same are provided in Appendix D. Estimation (3) of Table 16 is constrained in that the effect of context is assumed to be the same across male and female buyers and, thus, does not include the interaction term Male Buyer × Context. We reject the null hypothesis that the effect of context is the same between male and female buyers at the 10% level (chi-square score = 5.21, df = 2, p-value = 0.074).

Table 9: Panel probit estimations: Provision decisions

	Under-provision	Efficient provision	Over-provision
	(1)	(2)	(3)
(Intercept)	0.05	0.22	−0.56
	(0.57)	(0.35)	(0.44)
Male	−0.37	0.22	−0.09
	(0.49)	(0.22)	(0.42)
Male Buyer	0.34	−0.43**	0.49
	(0.37)	(0.22)	(0.36)
Male × Male Buyer	0.42	0.12	−0.76
	(0.53)	(0.29)	(0.50)
Context	−1.27**	0.30	0.27
	(0.55)	(0.23)	(0.41)
Male × Context	1.36*	−0.42	−0.12
	(0.72)	(0.30)	(0.53)
Male Buyer × Context	−1.01*	0.46	−0.01
	(0.57)	(0.29)	(0.46)
Male × Male Buyer × Context	0.38	−0.03	−0.15
	(0.77)	(0.39)	(0.65)
$\Pi_L - \Pi_H$	0.08		
	(0.12)		
$\mid \Pi_H - \Pi_L \mid$		−0.07	
		(0.15)	
$\Pi_H - \Pi_L$			−0.34***
			(0.12)
SVO angle	0.00	0.00	−0.01
	(0.01)	(0.00)	(0.01)
Risk aversion	0.02	0.02	−0.08
	(0.08)	(0.03)	(0.06)
Residual standard deviation σ	1.58***	0.46***	1.03***
	(0.27)	(0.11)	(0.19)
Log-Likelihood	−204.81	−519.24	−234.22
Mc Fadden R^2	0.21	0.03	0.11
Num. obs.	377	811	434

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

more likely to under-treat female buyers in the presence of context.[21] Thus the effect of the context on the under-treatment decision differs between male and female sellers: it substantially reduces under-provision for female sellers while slightly encouraging it for male sellers. Furthermore, the negative and significant coefficient associated with Male Buyer × Context indicates that the context decreases the likelihood of under-provision toward male buyers compared to under-provision by female sellers without the context[22]

Result 3: Under-provision is gender and context-dependent. Hence, our experiment partially supports H_3 and H_4. In particular,

Result 3-a: Context decreases the likelihood of under-provision for female sellers but increases it for male sellers.

Result 3-b: The presence of context makes male buyers less likely to suffer from under-treatment.

Over-provision – Column (3) of Table 9 deals with *over-provision*. None of our treatment and control variables are significant. This suggests that under-provision and over-provision have different determinants.[23]

Result 4: From the point of view of our study, under-provision and over-provision dimensions of fraud have different determinants. Specifically over-provision is not sensitive to any of our treatment conditions. Thus, H_3 and H_4 are not supported as regards to over-provision in our experiment.

[21]Estimation (1) from Table 16 is constrained in that the effect of context is the same across sellers. The likelihood ratio test (chi-square score = 7.71 , df = 2, p-value = 0.021) indicates the rejection at the 5 percent level of the null hypothesis of independence between seller's gender and market's environment. Thus, the effect of the context on the under-treatment decision differs between male and female sellers.

[22]Estimation (2) from Table 16 in Appendix D is constrained in that the effect of context is the same across buyers. The likelihood ratio test (chi-square score = 4.77, df = 2, p-value = 0.092)indicates the rejection at the 10 percent level of the null hypothesis of independence between buyer's gender and market's environment. Thus, the presence of context makes male buyers less likely to suffer from undertreatment.

[23]The negative and significant coefficient of $\Pi_H - \Pi_L$ is very puzzling for us, and we do not have a good explanation for it.

4.3.2 Consumer behavior

Interaction – Table 10 in Appendix A shows that the decision to enter the market does not significantly vary across our experimental conditions.[24] That is, consumer decisions to enter are very similar across all treatments. Therefore, H_5 and H_6 regarding buyer behavior are not supported by our data. This result corroborates the findings of Bonein and Serra (2009) who reported no significant change in trust behavior according to the counterpart's gender in an investment game.

Result 5: Consumer decisions to enter the market are very similar across all the treatments. Hence, H_5 and H_6 are not supported.

Summing up our findings, we find partial support for two of our hypotheses H_3 and H_4. Indeed, our data suggest that a form of positive discrimination toward males occurs in specific context in which buyers are subject to under-treatment. We also find that male sellers are more likely to receive an appropriate quality in the framed treatment. This indicates an important role played by the market environment, and shows that the prevalence of gender discrimination depends on the context of the transaction. In the next section, we discuss the results and the underlying mechanisms which might explain these findings.

5 Discussion and conclusion

In this paper, we conducted a controlled laboratory economic experiment from which we analyzed the behavior of 192 subjects. The starting point of this project was to

[24] All joint-hypothesis tests imposing restrictions on this regression regarding our variables of interest are insignificant. Furthermore, in Appendix A we show additional results of probit regression analyses for the interaction decision. Regarding the effect of past outcomes, Table 11 presents the results including two new variables: *Sufficient in* t_{-1} and *Profit in* t_{-1} that respectively account for (i) whether the quality provided in the previous period was sufficient to solve the problem and (ii) consumer's profit during the previous period. Then, investigating the effect of learning during the experiment, Table 12 presents the results of the regression analysis for periods 1 to 8 (column (1)) and periods 9 to 16 (column (2)). The results reveal that interaction decisions are still solely significantly negatively affected by the prices posted by the seller. Indeed, we do not observe any significant change in behavior due to past outcomes nor learning effect (no significant change between the last rounds compared the first ones).

explore the causal link between the gender of economic actors and discrimination in the provision of credence goods. The latest Global Gender Gap report (2020) of the World Economic Forum states: "in many countries, women are significantly disadvantaged in accessing credit, land or financial products, which prevents opportunities for them to start a company or make a living by managing assets" (p. 5). This points to the need for an understanding of the gender discrimination phenomenon that occurs in markets with strong informational asymmetries (such as financial advisory markets) and that might impact others spheres of economic life.[25] Thus, this issue deserves attention.

To the best of our knowledge, no other studies have been conducted to understand the effect of mutual observation of the other players' gender within these specific markets. Prior studies reviewed have noted the importance of gender differences in preferences, but none of them have assessed its impact on provision behavior in credence goods markets. These markets are characterized by huge informational asymmetries and expose consumers to a wide range of fraudulent behaviors that are of significant importance.[26] To address this question, we proposed a new experiment extending the one conducted by Dulleck et al. (2011), in which we dealt with the effects of a market environment and participant gender in a 2×4 design. We made several hypotheses with respect to discrimination and trust. Our design allowed us to assess the influence of gender information on (fraudulent) behavior and address the following issues: Is gender a driver variable in the choice of (i) provision and (ii) interaction with an expert? Moreover, are these effects sensitive to the context of the transaction?

As predicted, we found, even in our highly stylized experiment, evidence of a context-dependent gender discrimination. This could be the case in the market for car repair where women are subject to negative gender discrimination. We identify two main channels that may lead to gender discrimination. First, it might be the result of a statistical discrimination (Phelps, 1972). This result supports the hypothesis that the strategic de-

[25]That said, markets for health advice are not exempt as shown by the title of the international conference held in 2016 and hosted by CERES "Gender-Sensitive Health Literacy – A Future Concept for Public Health?". Experts argued that there remains much to be done to avoid "under-, over- and mis-provision of health services" due to gender-related inequalities.

[26]In this paper we only considered under-provision and over-provision issues.

cision to provide one quality over another is based on the consumer's knowledge about the good or service provided – knowledge that is hypothesized by sellers. The rationale is the following. Independently from the level of expertise required to provide a good or service, each transaction is rooted in a specific context including - but not limited to - historic, social, cultural, or geographical factors. The smaller the informational distance hypothesized, the less likely sellers feel free to discriminate consumers. Second and in addition to that, principle of taste-based discrimination (Becker, 1971) provides an alternative explanation. According to this principle, discrimination is rooted in preferences and driven by a sort of animus toward women as members of an out-group. Thus, the discrimination occurring in our experiment would results from male sellers who dislike and exhibit anti-social preferences against women by favoring male buyers. Put another way and as particularly evidenced in our framed treatment, the taste-based explanation could be interpreted as a form of sexism that sellers exhibit against women. However, since in our experiment we did not elicit the sellers' beliefs about the consumer's knowledge, nor did we ask them how these beliefs were formed, nor even did we invite them to reveal their strategies, we are not able to disentangle between the two channels. Considerably more work will be needed to clearly distinguish these different drivers of gender discrimination: biased beliefs and taste-based discrimination.

While gender information seems to have no effect on the decision to enter (in any of our treatments), we unexpectedly find that the two dimensions of fraud – over-provision and under-provision – have different determinants. Even if we identify gender discrimination as a key driver of under-provision, it is not the reason for over-provision. Note that in our experiment, under-provision hurts the consumers much more than over-provision. Thus, although as we have noted that our experiment does not clearly distinguish between the two channels of discrimination, taste-based or statistical, our result may be better explained by the former than the latter.

What is clear from our experiment is that female consumers should be simply aware that the quality of some services provided by men could be lower than required due to specific malevolent targeting. Our results provide a possible rationale to the general

perception that women are discriminated in markets for car repair even when the quality of the service provided can be verified. However, there is an important mechanism that is absent in our study but may prevent gender discrimination and reduce the level of fraud, reputation. This is especially relevant in the age of digitalization with the general access to consumer's online reviews: it might be a powerful tool creating strong incentives for experts to change their behavior.[27]

Public policies implemented with a view to reducing the contextual weight could also be useful to overcome some types of gender-based discrimination. Notably, the gender stereotypes that girls and boys acquire in early childhood still seem to affect their choices (although unconsciously) as it is the case in our experiment. It hinders progress toward equality. It seems difficult to eliminate these deep-rooted attitudes by offering a legal framework to promote equality. As Naves and Wisnia-Weill (2014) explain in their report for the French government, a change in the micro-incentives that are at play from early childhood to adolescence could prevent from unfounded differentiation and thus in the long run from inequality. Since it seems that these subjective beliefs that are shaped from childhood determine the provision choice, it may be in the interest of public institutions to address this issue. Further research would be needed to shed more light on this issue.

Our findings complement the existing literature on credence goods. It is encouraging to compare our results with those of Soraperra et al. (2019) who recently found that under-provision is sensitive to the sellers' social image concerns. The fear of being considered by consumers as an opportunistic seller who always provides high-costs treatments (that may be unnecessary) may induce decision-makers to undertreat. Given that our design does not allow us to address this question, this highlights the need for future investigation to examine the discrimination phenomenon in light of different social image concerns which is one of the several explanations proposed for differences in fraudulent behavior.

[27]Experimental results of Dulleck et al. (2011) and Mimra et al. (2016), however, show limited results on the effect of reputation itself in ensuring efficiency of markets for credence goods.

References

Agrawal, A., E. P. Green, and L. Lavergne (2019). Gender effects in the credence goods market: An experimental study. *Economics Letters 174*, 195–199.

Alesina, A. and E. La Ferrara (2002). Who trusts others? *Journal of public economics 85*(2), 207–234.

Andersen, S., S. Ertac, U. Gneezy, J. A. List, and S. Maximiano (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics 95*(4), 1438–1443.

Andreoni, J. and L. Vesterlund (2001). Which is the fair sex? gender differences in altruism. *The Quarterly Journal of Economics 116*(1), 293–312.

Balafoutas, L., A. Beck, R. Kerschbamer, and M. Sutter (2013). What drives taxi drivers? a field experiment on fraud in a market for credence goods. *Review of Economic Studies 80*(3), 876–891.

Balafoutas, L. and R. Kerschbamer (2020). Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions. *Journal of Behavioral and Experimental Finance*, 100285.

Balafoutas, L., R. Kerschbamer, and M. Sutter (2012). Distributional preferences and competitive behavior. *Journal of economic behavior & organization 83*(1), 125–135.

Balafoutas, L., R. Kerschbamer, and M. Sutter (2017). Second-degree moral hazard in a real-world credence goods market. *The Economic Journal 127*(599), 1–18.

Becker, G. S. (1971). The economics of discrimination, seconde édition (chicago.

Bertrand, M. and S. Mullainathan (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American economic review 94*(4), 991–1013.

Bonein, A. and D. Serra (2009). Gender pairing bias in trustworthiness. *The Journal of Socio-Economics 38*(5), 779–789.

Booth, A., L. Cardona-Sosa, and P. Nolen (2014). Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of economic behavior &*

Organization 99, 126–154.

Booth, A. L. and P. Nolen (2012). Gender differences in risk behaviour: Does nurture matter? *The Economic Journal 122*(558), F56–F78.

Bosman, R. and F. Van Winden (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal 112*(476), 147–169.

Buchan, N. R., R. T. Croson, and S. Solnick (2008). Trust and gender: An examination of behavior and beliefs in the investment game. *Journal of Economic Behavior & Organization 68*(3), 466–476.

Charness, G. and U. Gneezy (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization 83*(1), 50–58.

Chaudhuri, A. and L. Gangadharan (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, 959–985.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic literature 47*(2), 448–74.

Darby, M. R. and E. Karni (1973). Free competition and the optimal amount of fraud. *The Journal of law and economics 16*(1), 67–88.

Dittrich, M. (2015). Gender differences in trust and reciprocity: evidence from a large-scale experiment with heterogeneous subjects. *Applied Economics 47*(36), 3825–3838.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–90.

Dulleck, U. and R. Kerschbamer (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic literature 44*(1), 5–42.

Dulleck, U., R. Kerschbamer, and M. Sutter (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *The American Economic Review 101*(2), 526–555.

Ellingsen, T., M. Johannesson, J. Mollerstrom, and S. Munkhammar (2013). Gender differences in social framing effects. *Economics Letters 118*(3), 470–472.

Filippin, A. (2016). Gender differences in risk attitudes. *IZA World of Labor* .(.), .

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments.

Experimental economics 10(2), 171–178.

Gneezy, U., J. List, and M. K. Price (2012). Toward an understanding of why people discriminate: Evidence from a series of natural field experiments. Technical report, National Bureau of Economic Research.

Gottschalk, F., W. Mimra, and C. Waibel (2018). Health services as credence goods: A field experiment.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *American economic review 92*(5), 1644–1655.

Kamas, L. and A. Preston (2015). Can social preferences explain gender differences in economic behavior? *Journal of Economic Behavior & Organization 116*, 525–539.

Kerschbamer, R. and M. Sutter (2017). The economics of credence goods–a survey of recent lab and field experiments. *CESifo Economic Studies 63*(1), 1–23.

Kerschbamer, R., M. Sutter, and U. Dulleck (2016). How social preferences shape incentives in (experimental) markets for credence goods. *The Economic Journal 127*(.), 393–416.

Kuhn, P. and M. C. Villeval (2015). Are women more attracted to co-operation than men? *The Economic Journal 125*(582), 115–140.

List, J. and U. Gneezy (2014). *The why axis: hidden motives and the undiscovered economics of everyday life*. Random House.

Liu, T. (2011). Credence goods markets with conscientious and selfish experts. *International Economic Review 52*(1), 227–244.

Mimra, W., A. Rasch, and C. Waibel (2016). Price competition and reputation in credence goods markets: Experimental evidence. *Games and Economic Behavior 100*, 337–352.

Murphy, R. O., K. A. Ackermann, and M. Handgraaf (2011). Measuring social value orientation. *Judgment and Decision making 6*(8), 771–781.

Naves, M.-C. and V. Wisnia-Weill (2014). Lutter contre les stéréotypes filles-garçons. *Rapport du Commissariat général à la stratégie et à la prospective 2*.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The american economic review 62*(4), 659–661.

Soraperra, I., A. Suvorov, J. Van de Ven, and M. C. Villeval (2019). Doing bad to look good. *Revue économique 70*(6), 945–966.

Sülzle, K. and A. Wambach (2005). Insurance in a market for credence goods. *Journal of Risk and Insurance 72*(1), 159–176.

Sutter, M., R. Bosman, M. G. Kocher, and F. van Winden (2009). Gender pairing and bargaining—beware the same sex! *Experimental Economics 12*(3), 318–331.

Tsugawa, Y., A. B. Jena, J. F. Figueroa, E. J. Orav, D. M. Blumenthal, and A. K. Jha (2017). Comparison of hospital mortality and readmission rates for medicare patients treated by male vs female physicians. *JAMA internal medicine 177*(2), 206–213.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics 13*(1), 75–98.

A Interaction decision

Figure 4: Relative frequency of interaction

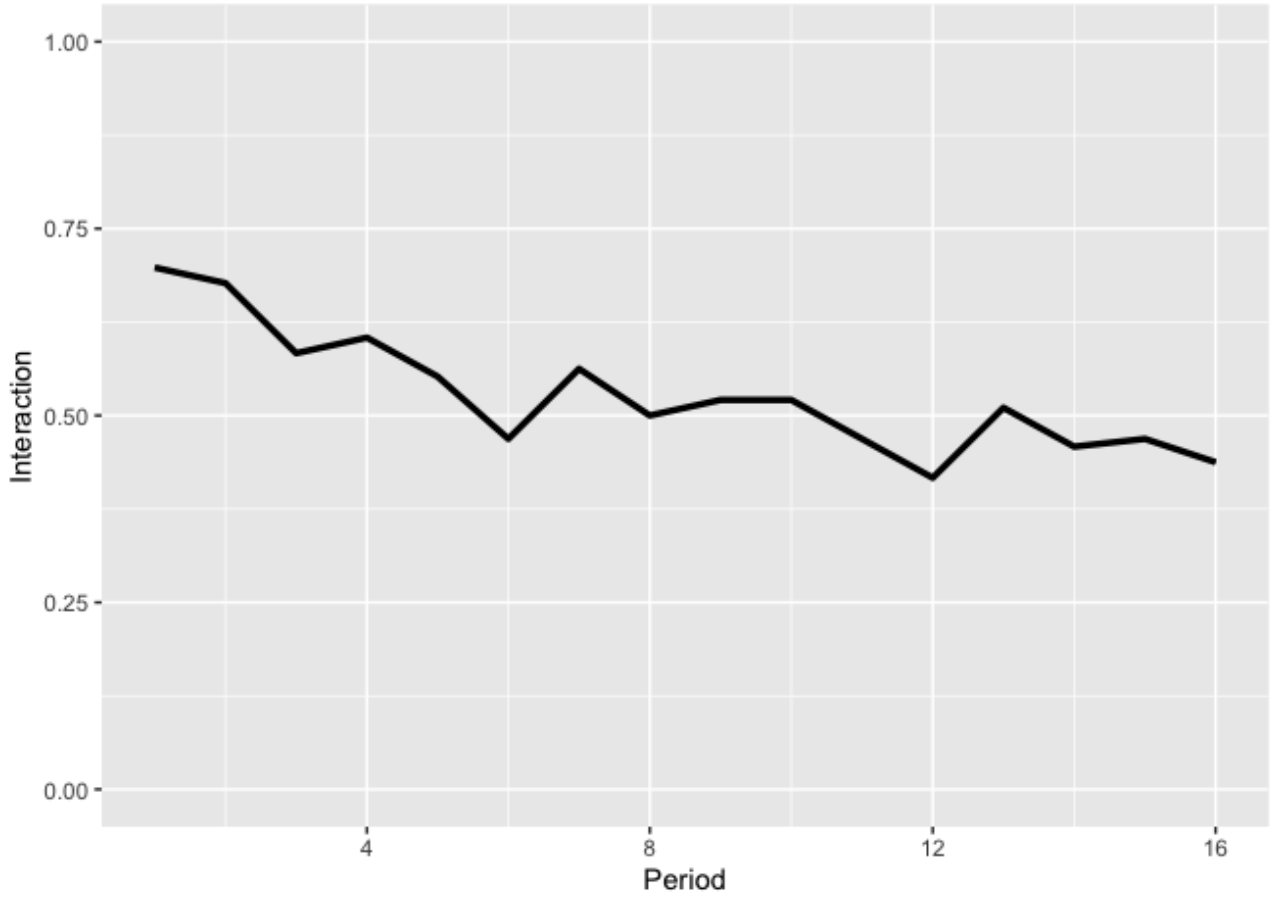


Table 10: Panel probit estimation

	Interaction
	(1)
(Intercept)	3.52^{***}
	(0.39)
Male	0.05
	(0.15)
Male Buyer	-0.30
	(0.24)
Male × Male Buyer	0.02
	(0.22)
Context	-0.18
	(0.24)
Male × Context	0.00
	(0.22)
Male Buyer × Context	0.37
	(0.33)
Male × Male Buyer × Context	-0.27
	(0.31)
P_L	-0.19^{***}
	(0.03)
P_H	-0.26^{***}
	(0.03)
SVO angle	-0.00
	(0.01)
Risk aversion	-0.00
	(0.04)
Residual standard deviation σ	0.83^{***}
	(0.09)
Log-Likelihood	-913.89
Mc Fadden R^2	0.14
Num. obs.	1536

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 11: Panel probit estimation

	Interaction
	(1)
(Intercept)	3.63***
	(0.37)
Male	0.01
	(0.17)
Male Buyer	−0.34
	(0.29)
Male × Male Buyer	0.02
	(0.29)
Context	−0.21
	(0.16)
Male × Context	0.01
	(0.22)
Male Buyer × Context	0.37
	(0.36)
Male × Male Buyer × Context	−0.23
	(0.33)
Price L	−0.19***
	(0.03)
Price H	−0.27***
	(0.03)
Profit in t_{-1}	0.01
	(0.02)
Sufficient in t_{-1}	0.06
	(0.18)
SVO angle	−0.01
	(0.01)
Risk aversion	0.01
	(0.04)
Residual standard deviation σ	0.88***
	(0.10)
Log-Likelihood	−853.14
Num. obs.	1440

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 12: Panel probit estimations

	Interaction periods 1 to 8 (1)	Interaction periods 9 to 16 (2)
(Intercept)	3.57^{***}	3.95^{***}
	(0.49)	(0.59)
Male	-0.21	0.38
	(0.21)	(0.23)
Male Buyer	-0.43	-0.14
	(0.27)	(0.32)
Male × Male Buyer	0.36	-0.44
	(0.30)	(0.33)
Context	-0.47	0.20
	(0.28)	(0.32)
Male × Context	0.29	-0.39
	(0.30)	(0.33)
Male Buyer × Context	0.63	0.04
	(0.39)	(0.45)
Male × Male Buyer × Context	-0.57	0.19
	(0.43)	(0.46)
Price L	-0.18^{***}	-0.22^{***}
	(0.03)	(0.04)
Price H	-0.24^{***}	-0.32^{***}
	(0.04)	(0.05)
SVO angle	0.00	-0.01
	(0.01)	(0.01)
Risk aversion	-0.01	0.00
	(0.04)	(0.05)
Residual standard deviation σ	0.75^{***}	1.01^{***}
	(0.11)	(0.13)
Log-Likelihood	-456.11	-452.03
Num. obs.	768	768

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

B Most frequently chosen price vectors broken down by treatment

Table 13: Prices vectors

Treatment	Most prominent price vectors (P_L, P_H)
No frame	(6;8) : 14% (7;8) : 8% (6;10) : 6%
With frame	(6;8) : 13% (5;8) : 12% (4;8) : 11%
Female sellers	(4;8) : 10% (6;8) : 8% (5;8) : 7%
Male sellers	(6;8) : 18% (5;8) : 9% (7;8) : 7%
Female sellers no frame (B-F)	(6;8) : 7% (6;10) : 6% (6;7) : 6%
Male sellers no frame (B-M)	(6;8) : 20% (7;8) : 10% (6;6) : 7%
Female sellers with frame (C-F)	(4;8) : 16% (5;8) : 10% (6;8) : 10%
Male sellers with frame (C-M)	(6;8) : 16% (5;8) : 13% (5;7) : 8%

Figure 5: Distribution of prices according to framing

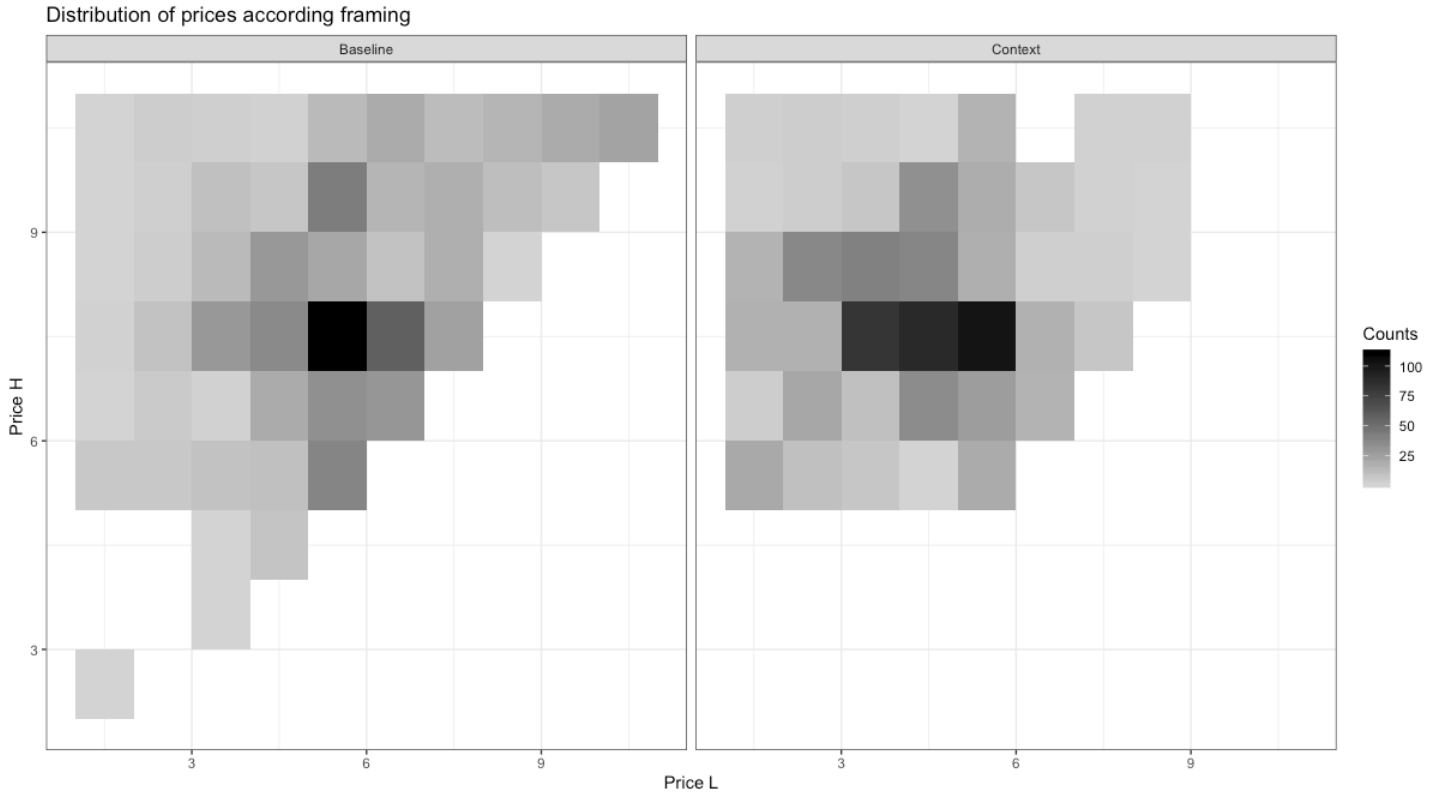


Figure 6: Distribution of prices according to seller's gender

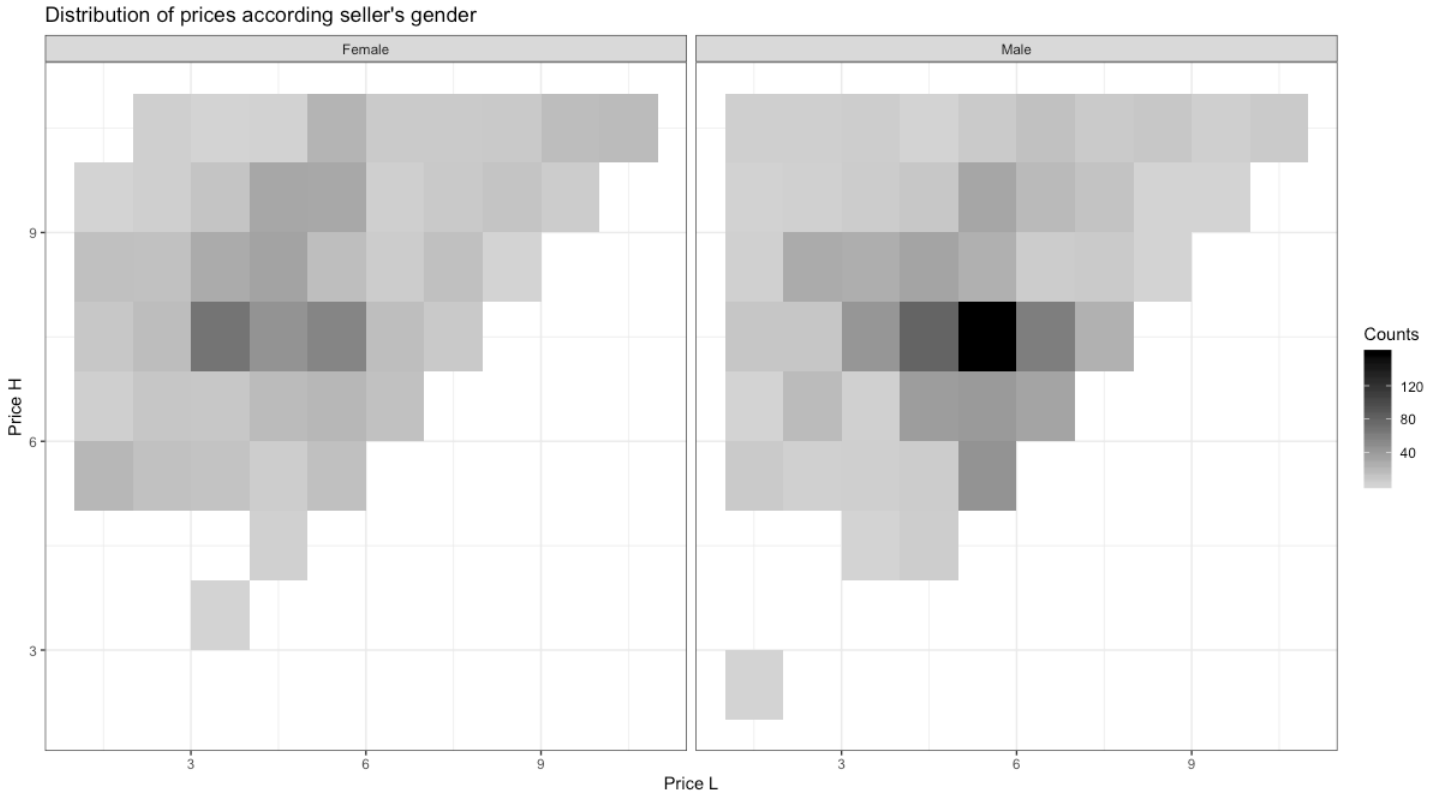
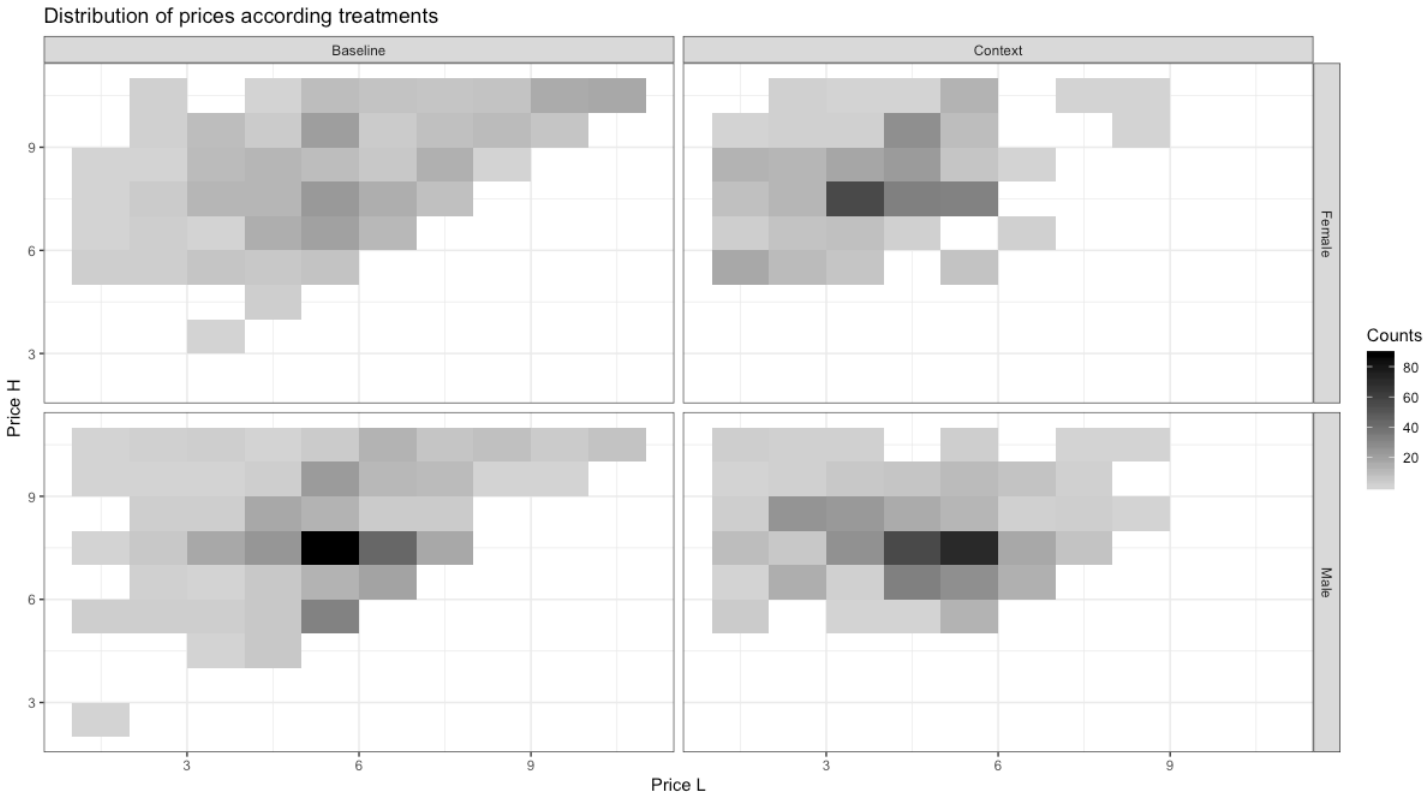


Figure 7: Distribution of prices according to treatment



C Inequality aversion analysis

Table 14: Efficient provision conditional on different price vectors

	$1^{st} stage$ {6,8}	$2^{nd} stage$ Efficient
(Intercept)	0.04	0.03
	(0.05)	(0.27)
Male	0.22^{***}	0.23
	(0.04)	(0.22)
Context	0.01	0.29
	(0.04)	(0.23)
Male × Context	-0.13^{**}	−0.42
	(0.05)	(0.30)
Male Buyer		-0.42^{*}
		(0.22)
Male × Male Buyer		0.12
		(0.29)
Context × Male Buyer		0.46
		(0.29)
Male × Context × Male Buyer		−0.02
		(0.39)
{6,8}		0.16
		(0.35)
SVO angle	0.00^{***}	0.00
	(0.00)	(0.00)
Risk aversion	−0.01	0.02
	(0.01)	(0.03)
Residual standard deviation σ		0.46^{***}
		(0.11)
R^2	0.06	
Adj. R^2	0.05	
Num. obs.	811	811
Log-Likelihood		−519.25

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 15: Probit $1^{st}stage$

	$1^{st}stage$ {6,8}
(Intercept)	-2.57^{***}
	(0.91)
Male	1.09^{*}
	(0.58)
Context	-0.67
	(0.71)
Male $\times$ Context	-0.24
	(0.71)
SVO angle	0.03
	(0.02)
Risk aversion	-0.05
	(0.10)
Residual standard deviation σ	2.06^{***}
	(0.48)
Log-Likelihood	-246.99
Num. obs.	811

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

D Joint hypotheses tests: constrained estimations

Table 16: Panel probit estimations

	Under-provision	Under-provision	Efficient provision
	(1)	(2)	(3)
(Intercept)	−0.39	0.21	0.07
	(0.56)	(0.56)	(0.34)
Male	0.31	−0.41	0.24
	(0.37)	(0.46)	(0.20)
Male Buyer	0.25	−0.10	−0.17
	(0.32)	(0.28)	(0.15)
Male × Male Buyer	0.61	0.50	0.10
	(0.38)	(0.38)	(0.20)
Context	−0.50	-1.77^{***}	0.53^{**}
	(0.36)	(0.48)	(0.18)
Male × Context		1.57^{*}	−0.46
		(0.62)	(0.24)
Male Buyer × Context	-0.83^{*}		
	(0.38)		
$\Pi_L - \Pi_H$	0.06	0.10	
	(0.12)	(0.12)	
$\mid \Pi_H - \Pi_L \mid$			−0.06
			(0.15)
SVO angle	0.00	0.00	0.00
	(0.01)	(0.01)	(0.00)
Risk aversion	0.03	0.02	0.02
	(0.08)	(0.08)	(0.03)
Residual standard deviation σ	1.69^{***}	1.61^{***}	0.48^{***}
	(0.28)	(0.27)	(0.11)
Log-Likelihood	−208.07	−207.20	−521.85
Mc Fadden R^2	0.20	0.21	0.02
Num. obs.	377	377	811

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

E Provision decisions without instrumenting for financial incentives

Table 17: Panel probit estimations

	Under-provision	Efficient provision	Over-provision
	(1)	(2)	(3)
(Intercept)	−0.18	0.17	0.21
	(0.58)	(0.24)	(0.41)
Male	−0.78	0.24	0.33
	(0.52)	(0.22)	(0.41)
Male Buyer	0.04	−0.42	0.39
	(0.40)	(0.22)	(0.40)
‘ Male × Male Buyer	0.76	0.12	−0.86
	(0.58)	(0.29)	(0.54)
Context	−0.61	0.26	−0.49
	(0.56)	(0.23)	(0.40)
Male × Context	1.17	−0.39	0.14
	(0.74)	(0.30)	(0.50)
Male Buyer × Context	−0.64	0.46	0.30
	(0.60)	(0.29)	(0.49)
Male × Male Buyer × Context	0.07	−0.04	−0.66
	(0.82)	(0.39)	(0.69)
$\Pi_L - \Pi_H$	0.71***		
	(0.10)		
$\mid \Pi_H - \Pi_L \mid$		−0.04	
		(0.04)	
$\Pi_H - \Pi_L$			0.60***
			(0.07)
SVO angle	−0.01	0.00	−0.00
	(0.01)	(0.00)	(0.01)
Risk aversion	−0.03	0.02	−0.05
	(0.08)	(0.03)	(0.05)
Residual standard deviation σ	1.49***	0.46***	0.73***
	(0.27)	(0.11)	(0.19)
Log-Likelihood	−167.49	−518.86	−192.24
Mc Fadden R^2	0.35	0.03	0.26
Num. obs.	377	811	434

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

F Experimental instructions and questionnaire

The instructions (Originally in French and adapted from Dulleck et al. (2011)) are presented here without framing. They are identical for the two treatments (except for some words, see Table 2).

F.1 Experimental instructions

INSTRUCTIONS OF THE EXPERIMENT

2 roles and 16 rounds

This experiment consists of 16 rounds, each of which consists of the same sequence of decisions. This sequence of decisions is explained in detail below.

There are 2 kinds of roles in this experiment: player A and player B. At the beginning of the experiment you will be randomly assigned one of these two roles. On the first screen of the experiment you will see which role you are assigned to. Your role remains the same throughout the experiment.

Players interact with each other in a market composed of 4 players A and 4 players B.

All participants get the same information on the rules of the game, including the costs and payoffs of both players.

Short Overview of the Sequence of Decisions in a Round

1. Player A chooses one price for action 1 and one price for action 2.
2. Player B sees the prices chosen by player A. Then player B decides whether he/she wants to interact with player A or not. If not, this round ends for him/her.
3. Each player A is informed if player B decided to interact with him/her. Then each player A is informed about the type of player B who decided to interact with him/her. There are two possible types of player B: he/she is either type 1 or type 2. This type is not necessarily identical for all players B. Player A has to choose an action for each player interacting with him: either action 1 or action 2.

Detailed Illustration of the Decisions and Their Consequences Regarding Payoffs

Each period is at most composed of 3 decisions. These decisions are carried out sequentially.

Player A makes decisions 1 and 3, while player B makes decision 2.

Decision 1: Player A

At decision 1 each player A has to set the prices for both actions, one these prices will then be charged to player B in decision 3.

In decision 3, in case of interaction, each player A has to choose between 2 actions (action 1 and action 2). Each action has a different cost for player A:

- Action 1 costs 2 points.
- Action 2 costs 6 points.

To cover these costs, player A must set prices for each of these actions. So, in decision 1, player A sets the price for each action. Only (strictly) positive integer numbers are possible, i.e., only 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 are valid prices. Note that the price for action 1 must not exceed the price for action 2.

Decision 2: Player B

Player B is informed of the prices set by player A for the two actions in decision 1. Then player B decides whether he/she wants to interact with player A. A variety of information about player A is displayed.

Player B decides whether he/she wants to interact with player A:

- If he/she wants to do so, player B has to click on "Yes" and confirm this by clicking on "OK". Then player A chooses an action in decision 3.
- If he/she does not want to interact, then this round ends and both players get a payoff of 1.6 points for this round.

Decision 3: Player A

Before Decision 3 is made (if player B chose 'Yes' in Decision 2) a type is randomly assigned to player B. Each player B can be of one of two types: type 1 or type 2. This type is determined for each player B and is new in each round. With a probability of 50% player B is of type 1, and with a probability of 50% player B is of type 2. This probability is completely random and independent from the type of other players B.

Player A gets to know the type of player B before he/she makes Decision 3. Then player A chooses an action, either action 1 or action 2, and receives the corresponding price. Player A can choose any action but depending on the type of player B, the chosen action could be insufficient.

An action is sufficient under the following conditions:

(i) Player B is a type 1 player and player A chooses either action 1 or action 2.

(ii) Player B is a type 2 player and player A chooses action 1.

An action is insufficient if player B is of type 2 and player A chooses action 1.

At no time will player B be informed of his own type. He will therefore not know if he is a type 1 or type 2 player.

Player A must choose an action (1 or 2). A series of information about player B will be displayed.

Payoffs

No interaction

If player B chooses not to interact with player A (decision 2 is "No") both players A and B get 1.6 points for this round.

Interaction

In case of interaction (decision 2 is "Yes"), profits are determined as follows:

- For player A: the price set in period 1 corresponding to the chosen action minus the cost (that is fixed and exogenously determined) of the action chosen in decision 3.
    - **Profit A = Price posted for the chosen action - Cost of the action**

- For player B: if the action chosen by player A is sufficient, then player B receives 10 points. Otherwise, player B receives nothing. From this amount, he has to remove the price set by player A for the chosen action.

  Then:

  - Player B is type 1:
    - **Player A chooses action 1 (sufficient): Profit B = 10 – $price_1$**
    - **Player A chooses action 2 (sufficient): Profit B = 10 – $price_2$**

  - Player B is type 2:
    - **Player A chooses action 1 (insufficient): Profit B = 0 – $price_1$**
    - **Player A chooses action 2 (sufficient): Profit B = 10 – $price_2$**

At the beginning of the experiment you receive an initial endowment of 6 points. In addition you received 5 euros corresponding to your show-up fee. With this endowment you are able to cover losses that might occur in some rounds. Losses can also be compensated by gains in other rounds. If your total payoff sums up to a loss at the end of the game your total profit for this step of the experiment will be reduced to 0. Please note that there is always a possibility to avoid losses in this experiment.

To calculate the final payoff, the initial endowment and the profits of all rounds are added up. This sum is then converted into cash using the following exchange rate:

1 point = 0.25 euros

(4 points = 1 euro).

F.2 Understanding questionnaire