Games and Economic Behavior 81 (2013) 145–164

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Shaping beliefs in experimental markets for expert services: Guilt aversion and the impact of promises and money-burning options ☆,☆☆

Adrian Beck [a], Rudolf Kerschbamer [a], Jianying Qiu [b], Matthias Sutter [c,d,*]

[a] Department of Economics, University of Innsbruck, Austria
[b] Department of Economics, Radboud University Nijmegen, The Netherlands
[c] Department of Public Finance, University of Innsbruck, Austria
[d] Department of Economics, University of Gothenburg, Sweden

ARTICLE INFO

Article history:
Received 14 July 2010
Available online 16 May 2013

JEL classification:
D03
D84
C72
C91
D82

Keywords:
Credence goods
Belief-dependent preferences
Guilt aversion
Promises
Money burning
Psychological forward induction
Experiments

ABSTRACT

In a credence goods game with an expert and a consumer, we study experimentally the impact of two devices that are predicted to induce consumer-friendly behavior if the expert has a propensity to feel guilty when he believes that he violates the consumer's payoff expectations: (i) an opportunity for the expert to make a non-binding promise; and (ii) an opportunity for the consumer to burn money. In belief-based guilt aversion theory the first opportunity shapes an expert's behavior if an appropriate promise is made and if it is expected to be believed by the consumer; by contrast, the second opportunity might change behavior even though this option is never used along the predicted path. Experimental results confirm the behavioral relevance of (i) but fail to confirm (ii).

1. Introduction

Goods and services where an expert seller knows more about the quality a consumer needs than the consumer herself are called credence goods. While they have an uncommon name, these goods are consumed frequently. Examples include car repair services, where the mechanic knows more about the type of service the vehicle needs than the owner; taxicab

☆ A previous version of this paper was circulated as IZA Discussion Paper 4827 under the title "Guilt from promise-breaking and trust in markets for expert services". We are indebted to an associate editor and three anonymous referees for their very detailed reports. Their constructive comments and suggestions helped to improve the paper a lot. Special thanks are due to Martin Dufwenberg for inspiring the forward-induction interpretation of the money-burning option. Financial support from the Austrian Science Fund (FWF) through grant number P20796 and from the Austrian National Bank (OeNB Jubiläumsfonds) through grant number 13602 is gratefully acknowledged.

* Corresponding author at: Department of Public Finance, University of Innsbruck, Universitätsstrasse 15, A-6020 Innsbruck, Austria.
E-mail address: matthias.sutter@uibk.ac.at (M. Sutter).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.05.002

rides in an unknown city, where the driver is better informed about the shortest route to the destination than the tourist; or medical treatments, where the doctor knows better which disease a patient has and which treatment is needed.

From the viewpoint of standard economic theory (relying on the assumption that all agents are rational, risk-neutral and exclusively interested in their own material payoffs) efficiency in markets for credence goods is expected to be low for the following reasons: If not restricted by institutional safeguards, such as liability clauses or ex post verifiability of actions, experts will always provide a low quality service, even when consumers need a higher quality; and experts will ask for a higher price than warranted by the provided service. The former type of fraud is known as *undertreatment* and the latter type as *overcharging*. When consumers can judge the quality of service they get (without knowing whether the quality received is the ex ante needed one, though), experts may also provide an unnecessarily high quality, which is referred to as *overtreatment* (see Dulleck and Kerschbamer, 2006, for a survey of the theoretical literature).

Despite the monetary incentives for experts to defraud consumers, the turnover in markets for credence goods is huge. For instance, the online site researchandmarkets.com reports that the U.S. auto repair industry comprises about 170,000 firms with combined annual revenues of $90 billion, of which 70% originates from mechanical repair. Likewise, health care expenditures account for approximately 15% of GDP in the U.S. and are still rising (WHO World Health Statistics 2009).

In this paper, we examine in an experiment the influence of belief-shaping devices on the efficient provision of credence goods. While institutional safeguards against fraud (like liability) and market forces (like competition and the possibility to build up a reputation) have been shown recently to increase efficiency on credence goods markets, the impact of "soft" factors such as non-binding promises has been ignored so far as possibly important for limiting undertreatment, overtreatment, and overcharging, and thus contributing to the efficient provision of credence goods.[1]

Soft factors such as non-binding promises have no effect on market behavior if all players are rational and only interested in their own material payoff, and if this fact is common knowledge among players. However, if agents have belief-dependent preferences, soft factors might become behaviorally relevant.

The point that belief-dependent motivations may be important for strategic decision making has first been made by Geanakoplos et al. (1989). They point out that traditional game theory is ill-equipped for studying such issues and develop an extension – psychological game theory – to investigate the impact of belief-dependent motivations on behavior. Battigalli and Dufwenberg (2009) generalize and extend the Geanakoplos et al. framework in several directions, thereby providing a sound theoretical basis for the study of belief-dependent motivations in dynamic games.

This paper focuses on a specific kind of belief-dependent motivation, guilt aversion. Psychological guilt aversion – as modeled by Battigalli and Dufwenberg (2007) – is the hypothesis that people feel guilty if they think that they violate others' payoff expectations. Its behavioral relevance is hard to prove empirically because a direct test of the theory requires elicitation of second-order beliefs. This has been done in several previous studies (by Dufwenberg and Gneezy, 2000; Guerra and Zizzo, 2004; Charness and Dufwenberg, 2006; Bacharach et al., 2007, for instance) and a typical finding has been that there is a significant positive correlation between stated second-order beliefs and behavior. This evidence is clearly consistent with psychological guilt aversion theory. However, as several researchers have recently pointed out, it is also consistent with other hypotheses, for instance with the consensus-effects hypothesis positing a reverse causality relationship (from preferences and behavior to beliefs).[2]

This paper investigates the relevance of belief-dependent guilt aversion for efficiency on credence goods markets without eliciting information about beliefs of players. Specifically, it tests the effects of two devices that are predicted to have no impact on behavior if agents have standard preferences, but have the power to manipulate beliefs and are therefore predicted to have an impact on behavior if players are guilt averse:

Device 1: Giving the expert an opportunity to make a promise. Though being cheap talk under the assumption of common knowledge that both parties are rational and exclusively interested in their own material payoffs, an opportunity to make a promise might increase efficiency on credence goods markets if experts are guilt averse and believe that appropriate promises affect the payoff expectations of consumers. In this case appropriate promises affect second-order beliefs and second-order beliefs affect behavior. The behavioral relevance of non-binding promises has been investigated before, of course, for instance by Ellingsen and Johannesson (2004), Charness and Dufwenberg (2006) and Vanberg (2008).[3] Our novelty here is that, due to the nature of credence goods, promises can take three natural forms, each implying a particular restraint on the expert's behavior. In particular, there is (i) a promise to provide a sufficient quality that – if kept – acts as a substitute for liability; (ii) a promise to charge the appropriate price for the quality provided that – if kept – acts as a substitute for verifiability; and (iii) a promise to provide a sufficient quality and to charge the appropriate price for

[1] See Dulleck et al. (2011) for a large experimental study on the effects of verifiability, liability, competition, and reputation on markets for credence goods, and Huck et al. (2007, 2010, 2012) on the role of competition, reputation, and information exchange on markets for experience goods. For a distinction of the different types of goods see Darby and Karni (1973).

[2] This possibility has already been discussed by Charness and Dufwenberg (2006, p. 1594). Ellingsen et al. (2010) test the guilt aversion hypothesis in a way that reduces the scope for consensus effects and find almost no evidence supporting the behavioral relevance of belief-dependent motivations. They conclude that previous findings to the contrary may be driven by consensus effects. Vanberg (2010) shows theoretically that rational belief formation implies a correlation between preferences and beliefs (of any order) as long as preferences of players in the population are correlated. He concludes that we should expect a correlation in experimental settings even if belief-dependent motivations are absent.

[3] In the context of social dilemmas – where both players have dominant strategies to defect rather than cooperate – promises have also been found to be efficiency-increasing. See Orbell et al. (1988), Ostrom (1998) or Dawes and Messick (2000), for instance.

that quality. This design feature allows investigating the endogenous selection among promises that differ in more than one dimension, and their effectiveness on trade in credence goods markets.

Device 2: Giving the consumer an opportunity to burn money. Battigalli and Dufwenberg (2009, Proposition 12) have shown that the presence of a money-burning option might be quite powerful when agents have belief-dependent motivations. Their story is based on a psychological forward-induction argument and the idea that chosen and unchosen alternatives reveal first-order beliefs of the decision maker and thereby shape second-order beliefs of an observer. Since in psychological guilt aversion theory second-order beliefs influence behavior, unchosen alternatives of one person might have an effect on the behavior of another person. Here is a sketch of the argument (see Section 3 and Appendix A for details): Suppose the consumer of a credence good has the opportunity to burn money before interacting with the expert and the amount burned is communicated to the expert. Furthermore, suppose that the moderately guilt-averse expert knows that for a rational consumer burning money is only a best reply if he has at least somewhat optimistic beliefs about the behavior of the expert. Then, the observation that the consumer has burned money allows the expert to draw the inference that the consumer has at least somewhat optimistic beliefs and this inference might induce him to behave in a consumer-friendly manner. Moreover, by not burning money the consumer (anticipating consumer-friendly behavior in response to burning money) conveys the belief that he expects to get even more. Thus, in the presence of a money-burning option simply trading with the expert (without burning money) allows the expert to draw more exact inferences about the consumer's optimistic expectations than the same action allows in the absence of the money-burning option and might thereby induce even a moderately guilt-averse expert to behave in a consumer-friendly manner (while in the absence of this option only very guilt-averse experts would behave that way). Psychological forward-induction arguments are theoretically elegant (they rely only on the assumption of common belief in rationality without evoking equilibrium thinking) and quite powerful (in many games evoking equilibrium yields multiplicity while applying psychological forward induction yields uniqueness – game Γ_2 in Battigalli and Dufwenberg, 2009, is an example).

Our results from an experiment with 208 participants confirm the behavioral relevance of promises: most experts make the predicted promise and proper promises induce consumer-friendly behavior. However, we also obtain some data that is clearly inconsistent with the current version of the theory of 'simple guilt': A non-negligible fraction of experts make promises that should not be observed according to this theory and this fraction is increasing over time. We discuss this point further in the discussion and conclusion section. Turning to the effect of the money-burning option our results fail to confirm its behavioral relevance, thus confirming earlier studies. We conclude that promises are behaviorally relevant but that their impact cannot be explained by psychological guilt aversion theory alone, and that money burning is less powerful in practice than in theory.

The rest of the paper is organized as follows. Section 2 introduces three versions of our credence goods game, a baseline game, an extended game where the expert has an option to make a promise to the consumer before the consumer decides whether to interact with the expert, and a different extension in which the consumer has an option to burn money before interacting with the expert. The section ends with predictions for the three games for the case where it is common knowledge that all agents are rational and exclusively interested in their own material payoffs. Section 3 introduces belief-based guilt aversion and studies its impact on behavior in the three games. Section 4 explains the experimental design and derives predictions for our three treatments which correspond to the three games introduced in Section 2. Section 5 reports the experimental results, and Section 6 discusses alternative explanations and concludes. All proofs are in Appendix A.

2. Material games

2.1. Baseline game, Γ_B

We take a simplified version of Dulleck and Kerschbamer's (2006) model of credence goods as our starting point. In this game, there are two players, an expert (he) and a consumer (she). The consumer has a problem θ that is with an ex ante probability q_l minor ($\theta = l$) and with probability $1 - q_l$ major ($\theta = h$). If the consumer decides to visit the expert, the expert finds out the severity of the problem by performing a diagnosis. He then provides a service that is either of high (h) or of low (l) quality. In the following, we denote the index of the quality provided (the expert's "provision decision") by $\tau \in \{l, h\}$. The high quality solves both types of problems, while the low quality is only sufficient for the minor problem. Different qualities come at different costs for the expert. Denoting the cost to the expert for providing the quality $\tau \in \{l, h\}$ by C_τ, it holds that $C_l < C_h$.[4] For the quality he has provided the expert can charge one of two exogenously given prices, either P_l or P_h, with $P_l > C_l$, $P_h > C_h$, and $P_h \geqslant P_l$. In what follows, we denote the index of the quality charged for (the expert's "charging decision") by $\eta \in \{l, h\}$. The price the expert charges does not need to correspond to the quality he has actually provided. That is, η can be different from τ. If the quality is sufficient, i.e. if $\tau = h$ for $\theta = h$, or $\tau \in \{l, h\}$ for $\theta = l$, the consumer receives a value of $V > 0$, otherwise she receives a value of zero. In the sequel we assume that V satisfies

$$V > \max\left\{C_h, \frac{C_h - C_l}{1 - q_l}\right\}. \tag{1}$$

[4] For simplicity we will often denote both the quality provided and the cost of the quality provided by C_τ. No confusion should result.

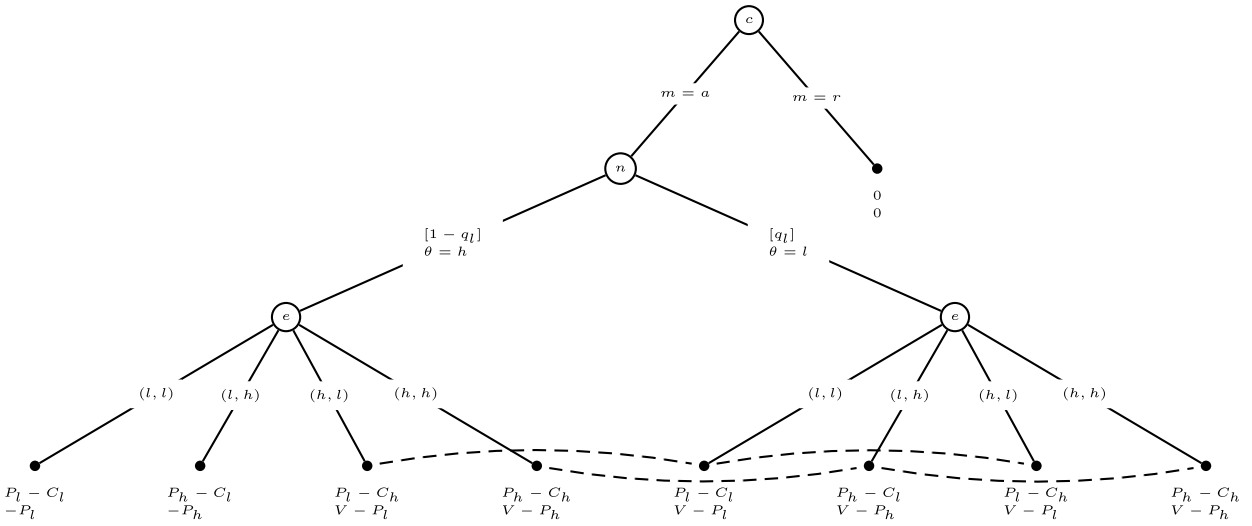


Fig. 1. The baseline game with material payoffs, Γ_B, where c denotes the consumer, e the expert, and n nature. The expert's choice is represented by (τ, η), where τ is the index for the provided quality and η is the index for the charged price. Non-singleton information sets, that is the dashed lines, represent the consumer's terminal information. The expert's material payoff is the top value of the payoff vector, and the consumer's payoff is the bottom value of the payoff vector.

That V is larger than the first element in $\{\cdot\}$ makes sure that solving the major problem with the high-quality service is more efficient than leaving the major problem unsolved and that V is larger than the second element implies that, from an ex ante perspective, providing the high-quality service is more efficient than providing the low-quality one.

If an interaction takes place the material payoff of the consumer, denoted $\pi_c(\tau, \eta|\theta)$, is the value (of V or 0) minus the price for the quality charged for:

$$\pi_c(\tau, \eta|\theta) = \begin{cases} -P_\eta & \text{for } \tau = l,\ \eta \in \{l, h\} \text{ and } \theta = h, \\ V - P_\eta & \text{for any other } (\tau, \eta, \theta) \in \{l, h\}^3. \end{cases}$$

The material payoff of the expert in the case of interaction, $\pi_e(\tau, \eta)$, corresponds to the price charged minus the cost of the quality provided:

$$\pi_e(\tau, \eta) = P_\eta - C_\tau \quad \text{for any } (\tau, \eta, \theta) \in \{l, h\}^3.$$

If no trade takes place both parties end up with a material payoff of zero. To make things interesting we assume that

$$V < P_h/q_l. \tag{2}$$

This condition implies that if both players are exclusively interested in their own material payoffs and if this fact is common knowledge, then the consumer never accepts to be served, causing a breakdown of the market. See Subsection 2.4 for details.

The order of moves is shown in Fig. 1. Prices are exogenously given and observed by both players. Based on this information the consumer decides whether to visit the expert or not. We denote this decision by $m \in \{a, r\}$, where a stands for acceptance and r for rejection. If the consumer interacts with the expert, a random move by nature determines the severity of her problem. The expert observes the severity and then decides which quality to provide and which price to charge. We define *overcharging* as charging for the high quality while providing the low one ($\tau = l,\ \eta = h|\theta \in \{l, h\}$), *undertreatment* as providing the low quality when the consumer has the major problem ($\tau = l,\ \eta \in \{l, h\}|\theta = h$), and *overtreatment* as providing the high quality when the consumer has the minor problem ($\tau = h,\ \eta \in \{l, h\}|\theta = l$).

Notice that in this game neither is the expert obliged to provide a quality that solves the consumer's problem, nor can the expert's action be (perfectly) verified.[5] We denote this baseline game by Γ_B. In the sequel we modify it to two extended games, Γ_P and Γ_M.

2.2. Game with promise option, Γ_P

In Γ_P at the start of the game the expert has the option to make a non-binding promise p from the set $P = \{NO, SQ, AP, SQ\&AP\}$ to the consumer, where the elements of P are defined as follows[6]:

[5] Only when the consumer has a major problem and the expert provides the low quality, the consumer can indirectly infer her problem type and the quality provided since then she receives no positive value.

[6] A capital P stands for two different things in this paper: When it comes with a subscript (as in P_l and P_h) it stands for a price, when it comes without a subscript (as in $p \in P$) it denotes the set of available promises in game Γ_P and in experimental treatment T_P.

NO ("no promise"): an irrelevant message;
SQ ("sufficient quality"): a promise to provide sufficient quality;
AP ("appropriate price"): a promise to charge the price for the quality provided;
SQ&AP ("sufficient quality and appropriate price"): a promise to provide sufficient quality and to charge the price for the quality provided.

Promises have neither an effect on the available alternatives in the subgame starting after a promise has been made, nor an effect on the material payoffs players receive for any $(\tau, \eta, \theta) \in \{l, h\}^3$.[7]

2.3. Game with money-burning option, Γ_M

In Γ_M at the start of the game the consumer has the opportunity to choose a money-burning option m out of a finite set $M = \{0, \epsilon, 2\epsilon, \ldots, x\epsilon\}$, where $\epsilon > 0$ is some smallest currency unit and x some strictly positive integer satisfying $x \leqslant (V - P_h - \epsilon)/\epsilon$.[8] That is, instead of deciding whether to visit the expert or not, $(m \in \{a, r\})$, the consumer is now allowed to choose a message $m \in M \cup \{r\}$, where r stands again for rejecting interaction. In case of participation $(m \in M)$, the sum m is deducted from the consumer's material payoff (but is not transferred to the expert), and the expert is informed about the magnitude of the sum before making his provision and charging decision. Accordingly, the consumer's material payoff in case of participation changes to $\pi_c(\tau, \eta|\theta) - m$, while the expert's material payoff stays the same (as in Γ_B).[9]

2.4. Standard predictions

Consider game Γ_B. Suppose that both parties are rational, risk-neutral and care only for their own material payoffs, and that this is common knowledge. Then in any Perfect Bayesian Equilibrium (PBE) the expert always provides the low quality and charges the price for the high quality if the consumer accepts; anticipating this, the consumer decides against the visit of the expert (this follows from condition (2)), which leads to a market breakdown. The prediction for Γ_P is basically the same: In this game the expert has a much larger strategy space – he now chooses a promise p from the set P at the start, and for each p and each θ he then chooses $\tau \in \{l, h\}$ and $\eta \in \{l, h\}$ – but this does not have any effect on the predicted path: The consumer knows that she will always receive the low quality and pay the high price whatever the promise of the expert is. So she will decide against participation, in case of participation p is arbitrary, and for any p we get undertreatment and overcharging just as in Γ_B. The argument for Γ_M is similar, also leading to a market breakdown.

3. The impact of guilt aversion on market outcomes

The predictions for our three games change significantly if the expert has a disposition to feel guilty when he believes that he violates the consumer's payoff expectation. To show this we modify the expert's utility function to incorporate a guilt sentiment a là Battigalli and Dufwenberg (2007). To translate their 'simple guilt' hypothesis to the present context consider the baseline game and let α_a denote the consumer's initial (first-order) belief about the expert's provision and charging behavior in case of acceptance, and β_a the expert's expectation of α_a, *conditional* on the consumer's acceptance. Then the expert's ex post utility for the case where the consumer has accepted, the expert has provided the quality C_τ and charged the price P_η, given the consumer has problem θ and the expert the belief β_a, is assumed to be given by $U_e(\tau, \eta|\theta, \beta_a) = P_\eta - C_\tau - \gamma \max\{E_{\beta_a}(\pi_c(\tau', \eta'|\theta)) - \pi_c(\tau, \eta|\theta), 0\}$, where E_{β_a} is the conditional expectation operator given the belief β_a and $\gamma \geqslant 0$ is a psychological guilt-sensitivity parameter. We start by securing some facts that are true for any expectation the expert might hold and any value of the guilt-sensitivity parameter. To avoid a lot of conditional statements we strengthen for the rest of the paper our condition (1) by imposing in addition the requirement

$$V > \max\left\{\frac{P_h - P_l}{q_l}, \frac{C_h - C_l}{q_l}\right\}, \tag{3}$$

which is fulfilled by the parameter constellations implemented in the experiment. Given this condition we get (all formal statements and proofs are in Appendix A):

Observation 1. *If the consumer has the minor problem, then the expert strictly prefers providing the low over providing the high quality service.*

[7] The game Γ_P resembles the stochastic trust game studied by Charness and Dufwenberg (2006). Our novelty here is to transfer their ideas and framework to the richer credence goods context where different types of promises are related to specific problems in such markets, and where the stochastic structure of the game allows for some discrimination between different explanations for promise-keeping (as discussed in the concluding section).

[8] This restriction simplifies the presentation without qualitatively affecting the results – see Appendix A for details.

[9] The game Γ_M is a close relative of the "Generalized Trust Game with Guilt Aversion" discussed in Battigalli and Dufwenberg (2009, p. 23).

Observation 1 tells us that the expert has never an incentive to overtreat the consumer. The reason is that overtreatment reduces the expert's material payoff but has no effect on the amount of guilt as it leaves the consumer's material payoff and her payoff expectation unaffected.

Observation 2. *Generically, an expert who is willing to solve the consumer's problem in both states will charge the same price in both states.*

Observation 2 implies that a guilt-averse expert will never provide the appropriate quality and charge the price for the quality provided in both states.[10] This follows (i) from the observation that the consumer's payoff expectation is necessarily an average over the payoffs in the two states and as such depends neither on the severity of her problem nor on the expert's provision behavior; and (ii) from the fact that – for the case where the consumer's problem is solved – her actual payoff does not depend on those variables either. Thus, for the case where the consumer's problem is solved, charging the low instead of the high price involves exactly the same trade-off for the expert in both states: The cost is material and amounts to $P_h - P_l$ in both states, and the benefit consists in reducing the amount of guilt which is also the same in both states.

Observation 3. *The expert will never undertreat the consumer and at the same time charge honestly. Moreover, if the expert prefers charging the low over charging the high price when he solves the major problem, then he also prefers solving the consumer's major problem over letting it unsolved.*

Observation 3 is important because it implies that undertreatment is easier to prevent than overcharging. The reason is that overcharging involves a pure (1:1) transfer of money from the customer to the expert while undertreatment involves a transfer that is associated with an efficiency loss. Thus, for an expert not to have an incentive to overcharge, the material payoff of the customer must receive at least the same weight in the expert's utility function as the expert's own material payoff. A necessary condition for this to be the case is a value of the guilt-sensitivity parameter γ of at least 1. Together with the belief that the consumer has a high payoff expectation this condition is also sufficient. By contrast, if the expert undertreats the customer then his material payoff increases by $C_h - C_l$, while the material payoff of the customer decreases by $V > C_h - C_l$. Thus, for a given (high) payoff expectation of the customer a lower value of γ is sufficient to prevent undertreatment.[11]

Observations 1–3 together imply that generically the expert's behavior in the subgame starting after the consumer has accepted is of one of the following three types:

- **type-1 behavior:** the expert provides the low quality and charges the price for the high quality in both states;
- **type-2 behavior:** the expert provides the low quality if the consumer has the minor problem and the high quality if she has the major problem and he charges the price for the high quality in both states;
- **type-3 behavior:** the expert provides the low quality if the consumer has the minor problem and the high quality if she has the major problem and he charges the price for the low quality in both states.

3.1. Guilt aversion in the baseline game, Γ_B

We explore the impact of guilt aversion in Γ_B under two conditions. In the first condition, we assume that players coordinate on a Perfect Bayesian Equilibrium (hereafter PBE). The notion of PBE implies that players never change their beliefs about the beliefs of their opponents along the predicted path. As Battigalli and Dufwenberg (2009) argue in their pioneering paper on dynamic psychological games, however, this implication might be considered as questionable in a psychological game. Hence, in our second condition we relax the consistency condition of PBE to allow the expert to change his second-order belief when he observes an unexpected move by the consumer. Specifically, we allow the expert to draw inferences about the consumer's unobservable first-order belief from her observable participation decision via forward-induction reasoning.

Equilibrium analysis in Γ_B. Given the findings above and the requirement that in equilibrium beliefs of all orders have to be consistent with actual behavior, it is straightforward to prove that generically the baseline game admits three types of PBE: (i) *type*-1 *equilibria* in which the consumer rejects and the expert exhibits type-1 behavior in case of acceptance; (ii) *type*-2 *equilibria* in which the consumer accepts and the expert exhibits type-2 behavior in case of acceptance; and (iii) *type*-3 *equilibria* in which the consumer accepts and the expert exhibits type-3 behavior in case of acceptance. The bounds for existence of each of the three types of PBE are as shown in Fig. 2. As can be seen from this figure, for relatively low values

[10] An exception is the case where the expert is exactly indifferent between charging the low and charging the high price and when he is indifferent then he is so in both states. However, for any belief the expert might hold, this case is non-generic as it is consistent with at most a single realization of the guilt-sensitivity parameter.

[11] Note that invoking "guilt from blame" (as defined in Battigalli and Dufwenberg, 2007) would qualitatively yield the same prediction: A player suffering "guilt from blame" cares about others' inferences regarding the extent to which he is willing to let them down. Since undertreatment is observable while overcharging is not (except for the case where it comes together with undertreatment; see the information sets on terminal nodes in Fig. 1) an expert suffering guilt from blame has more incentives to avoid undertreatment than overcharging.

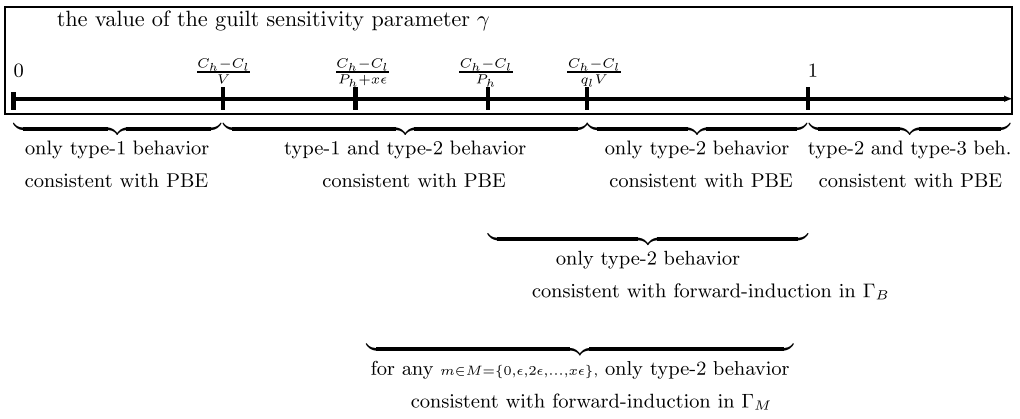


Fig. 2. Predictions for the baseline game, Γ_B, and for the game with money-burning option, Γ_M.

of the guilt-sensitivity parameter γ there is a unique PBE and the behavior in that PBE implies market breakdown, just as in the selfish benchmark of Subsection 2.4. In an intermediate range both market breakdown (supported by the expectation that the expert exhibits type-1 behavior in case of acceptance) and acceptance (supported by the belief that the expert exhibits type-2 behavior in case of acceptance) are consistent with equilibrium. And for relatively high values of γ the market necessarily works well: the consumer accepts and she gets the appropriate treatment in any PBE.

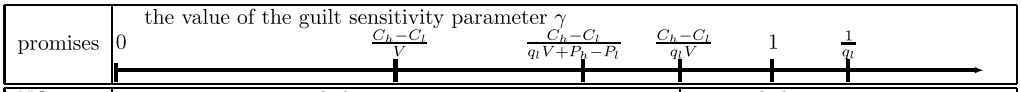
Forward-induction reasoning in Γ_B. Following Battigalli and Dufwenberg (2009) consider the following psychological forward-induction argument: Suppose the consumer accepts. If she is rational, she must believe that the expert behaves in a way that implies a material payoff of at least 0 for her – since this is what she receives if she rejects. Thus, even if the expert is initially uncertain regarding the customer's expectation, if he believes that the consumer is rational, then he infers $E_{\alpha_a}(\pi_c) \geqslant 0$ from her acceptance. Thus, $E_{\beta_a}(\pi_c) \geqslant 0$, which in turn implies that for values of the guilt-sensitivity parameter γ above the threshold $(C_h - C_l)/P_h$ type-1 behavior is dominated by type-2 behavior. It follows that for any $\gamma > (C_h - C_l)/P_h$ only type-2 and type-3 equilibria are consistent with forward-induction reasoning in Γ_B.

3.2. Guilt aversion in the game with promise option, Γ_P

Again, we explore the impact of guilt aversion under two conditions, first under the (equilibrium) assumption that promises are believed (and believed to be believed) if and only if believing in the promise is consistent with equilibrium; and then under the (non-equilibrium) assumption that promises are believed (and believed to be believed) whatever their content is.

Equilibrium analysis in Γ_P. Under the assumptions that promises are believed (and believed to be believed) if and only if believing in the promise is consistent with equilibrium, promises *NO* and *SQ*&*AP* never have any effect, the former because it has no content and the latter because generically there is no PBE in which the expert provides sufficient quality and charges the price for the quality provided in both states. By contrast, the impact of promises *SQ* and *AP* on the expert's behavior in the subgame starting after the consumer has accepted depends on the realization of γ. For values of γ in the range $[(C_h - C_l)/V, (C_h - C_l)/q_l V]$, promise *SQ* selects the type-2 equilibrium, while *AP* selects the type-1 equilibrium. Since type-1 equilibrium involves rejection while type-2 equilibrium entails profitable trade, the model predicts that the expert makes promise *SQ* and that he exhibits type-2 behavior after acceptance. Our model makes no prediction for other values of γ.

Non-equilibrium analysis in Γ_P. Suppose now that promises are believed (and believed to be believed) whatever their content is. Specifically, assume that for any promise $p \in P$ the consumer expects that, in case of acceptance, the expert maximizes lexicographically, first his own material payoff subject to the constraint that his promise is kept, and secondly (in case of a tie in own material payoffs) the payoff of the consumer (over the options that yield him the same material payoff). Further assume that the expert expects that the consumer has such beliefs. For this expectation it is straightforward to show that the expert's behavior for the case where the consumer has accepted after promise $p \in P$ is as shown in Fig. 3. First notice that the promise *NO* yields rejection. Thus, the expert will never use this promise if he expects that promises are believed. Next notice that *SQ* strictly dominates *SQ*&*AP* for any $\gamma > 0$. To see this, first note that both promises yield acceptance by the consumer. Furthermore, for $\gamma > 1/q_l$ the promise *SQ*&*AP* implies a lower price for the consumer than *SQ* at a material cost for the expert, but exactly the same amount of guilt; and for $\gamma \leqslant 1/q_l$ both promises yield exactly the same behavior (and thereby also the same material payoff for the expert), but *SQ*&*AP* yields a higher payoff expectation and therewith a higher amount of guilt than *SQ*. Thus, promise *SQ*&*AP* will never be made if it is expected to be believed. So we are left with the promises in the set $\{SQ, AP\}$ and conclude that the expert will make one of these promises and that the promise made by the expert is accepted by the consumer. Thus, promises are predicted to increase the frequency of trade under this condition.



NO	type-1 behavior	type-2 behavior	
AP_{UT}	type-1 behavior	type-2 behavior	
SQ	type-1 behavior	type-2 behavior	
SQ&AP	type-1 behavior	type-2 behavior	type-3 behavior

Fig. 3. The consequences of promises in the game with promise option, Γ_P, with the assumption that promises are believed to be believed. In the figure AP_{UT} stands for the promise *AP* under an undertreatment price-vector (i.e. $P_h - C_h < P_l - C_l$). The behavior after the promise *AP* under an overtreatment price-vector (defined by $P_h - C_h > P_l - C_l$) is the same as after promise *SQ*, and the behavior under an equal mark-up vector (defined by $P_h - C_h = P_l - C_l$) is the same as after *SQ*&*AP*.

3.3. Guilt aversion in the game with money-burning option, Γ_M

Equilibrium analysis in Γ_M. It is easily verified that the type of behavior consistent with equilibrium in the subgame starting after the consumer has sent message $m \in M$ does not depend on the amount of money burned by the consumer. Whether actual money burning is consistent with equilibrium depends on the extent to which the expert is guilt averse. For $\gamma < (C_h - C_l)/V$ there is no equilibrium in which the consumer decides for an $m \in \{\epsilon, \ldots, x\epsilon\}$ (she knows that, in case of acceptance, the expert always provides the low quality and charges the price for the high quality for any $m \in M$, and therefore decides for $m = r$); for $\gamma \in ((C_h - C_l)/q_lV, 1)$ there is no PBE in which she decides for such an m either (she knows that, in case of acceptance, the expert always provides the appropriate quality and charges the price for the high quality for any $m \in M$, and therefore decides for $m = 0$). For the remaining ranges of γ there is equilibrium multiplicity and actual money burning is consistent with PBE.

Forward-induction reasoning in Γ_M. We now show that iterated forward-induction reasoning has quite some power in our money-burning game. Specifically, we show that the path predicted by (iterated) forward induction is unique for any $\gamma \in ((C_h - C_l)/(P_h + x\epsilon), 1)$ and that the predicted path involves full efficiency (interaction followed by appropriate service) *without* money burning. We start by showing that – for guilt-sensitivity parameters in this range – for any $m \neq r$ only type-2 behavior is consistent with the assumption of common knowledge of rationality. To see this, first suppose the consumer burns the maximal feasible amount, $x\epsilon$. If the expert is rational and believes in the rationality of the consumer, then he infers from this move that the consumer expects at least $x\epsilon$. Therefore, $E_{\beta_{x\epsilon}}(\pi_c) \geqslant x\epsilon$. In Appendix A we show that an expert with such an expectation necessarily provides the high quality service to a consumer who has the major problem for any $\gamma > (C_h - C_l)/(P_h + x\epsilon)$. Thus, for guilt-sensitivity parameters in this range, burning $x\epsilon$ gives the consumer a payoff of $V - P_h - x\epsilon$ for sure. Taking the same logic one step further, suppose now the consumer burns the amount $(x-1)\epsilon$. Then the expert infers from that that the consumer expects at least $V - P_h - x\epsilon$. This is so because – if both the expert and the consumer accept the above reasoning – they believe that by burning the amount $x\epsilon$ the consumer can guarantee a material payoff of at least $V - P_h - x\epsilon$. Hence, burning less money is consistent with common knowledge of rationality only if it does not yield less. Now, we can apply the same logic to show that the expert treats the consumer appropriately, even when she burns only the amount $(x-1)\epsilon$. Proceeding this way further we get to the prediction that for any $m \geqslant 0$ (even for $m = 0$!) the expert treats the consumer appropriately. Anticipating appropriate treatment for any $m \geqslant 0$, the consumer decides not to burn any money (that is, to choose $m = 0$). Thus, actually burning money is not necessary to shape second-order beliefs and behavior; knowing that a money-burning option was available to the consumer is sufficient to induce consumer-friendly behavior.

4. Experimental design

At the beginning of our experiment, subjects were informed about their (fixed) role as either an expert or a consumer, and they received an initial endowment of 200 experimental currency units (ECU), equivalent to 2.5 Euro. Each session was run with 16 subjects who were split into two matching groups of 8 subjects each, yielding two independent observations per session. In each matching group, four subjects had the role of an expert, and the other four that of a consumer. Experts only interacted with consumers in the same matching group. The consumer's probability of having the minor problem was $q_l = 0.5$ and the value for receiving a sufficient quality was $V = 100$ ECU. The cost of providing the low quality was $C_l = 0$ ECU, and the cost of providing the high quality was $C_h = 30$ ECU. Each expert was exposed to each price-vector (P_l, P_h) in the set $\{(30, 50), (30, 60), (30, 70), (30, 65), (40, 65), (50, 65)\}$ four times, two times with the consumer having a minor problem and two times with the consumer having a major problem.[12] In total this sums up to 24 rounds for each subject, and the profits from all rounds were added up to yield the total payoffs. The sequence of (price-vector, problem)-pairs was randomized on the individual expert's level.

[12] In the theory section we have distinguished between "undertreatment", "overtreatment" and "equal mark-up" price-vectors (see the text to Fig. 3). In the set of tested price-vectors (30, 50), (40, 65), and (50, 65) are undertreatment vectors, (30,60) is an equal mark-up vector, and (30, 65) and (30, 70) are overtreatment vectors.

We had three different treatments – denoted T_B, T_P and T_M – which correspond to our three games Γ_B, Γ_P and Γ_M, as described in Section 2. More details on treatments follow below. The experiment was run in June 2009. All sessions were computerized using z-Tree (Fischbacher, 2007) and recruiting was done with ORSEE (Greiner, 2004). Four sessions were conducted for treatments T_B, and T_P, and five for treatment T_M. In total we had 208 undergraduate students from the University of Innsbruck participating in the experiment. At the beginning of each session, the instructions (see Appendix B) were read aloud to make them common knowledge. Subjects were also given about 10 minutes to read through the instructions alone and ask questions. Before the experiment started, subjects had to answer a set of control questions, and the experiment proceeded only after all control questions were answered correctly. Each session, including instructions and control questions, lasted on average 1 hour and 15 minutes, and subjects' average earnings, including a show up fee of 5 Euro, were 15 Euro.

4.1. Description and prediction for baseline treatment, T_B

In treatment T_B, which corresponds to game Γ_B, at the beginning of each round, an expert was randomly paired with a consumer, and both got to know the price-vector (P_l, P_h). Then the consumer could decide whether she would like to interact with the expert or not. If the consumer decided to interact with the expert, the expert got to know the problem of the consumer and decided which quality to provide and which price to charge. At the end of each round both the consumer and the expert were informed of their own payoffs. Based on the theoretical analysis in Section 3 our prediction for T_B is[13]:

Prediction 1 (*Behavior in baseline treatment*). **P1a:** Conditional on acceptance ($m \neq r$), the expert exhibits either (1) type-1 behavior: he provides the low quality and charges the price for the high quality in both states; or (2) type-2 behavior: he provides the low quality if the consumer has the minor problem and the high quality if she has the major problem, and he charges the price for the high quality in both states; or (3) type-3 behavior: he provides the low quality if the consumer has the minor problem and the high quality if she has the major problem and he charges the price for the low quality in both states. Overtreatment (provision of the high quality if the consumer has the minor problem) is rarely observed for any price constellation. **P1b:** If psychological forward induction works, type-2 behavior becomes more and type-1 behavior becomes less frequent when P_h increases; by contrast, the frequency of type-3 behavior does not depend on P_h.

4.2. Description and prediction for treatment with promise option, T_P

In the treatment T_P, which corresponds to game Γ_P, the expert was given an opportunity to send one of the following four messages after having observed the price-vector[14]:

NO: "Hello".
SQ: "I promise I will provide a sufficient quality".
AP: "I promise I will charge the low price if I provide the low quality, and I will charge the high price if I provide the high quality".
SQ&AP: "I promise I will provide the low quality and charge for it if you have the minor problem, and I will provide the high quality and charge for it if you have the major problem".

Based on the theoretical analysis in the previous section our prediction for T_P is:

Prediction 2 (*The effect of giving the expert an option to make a promise*). **P2a:** Promise *SQ* is frequently observed in T_P, while promises *NO* and *SQ&AP* are rarely observed. **P2b:** Compared to the baseline treatment T_B consumers are more likely to accept in T_P after promises *SQ* and *SQ&AP* and less likely to accept after promise *NO*. **P2c:** Conditional on acceptance ($m \neq r$), the expert is more likely to behave in a consumer-friendly manner in T_P than in T_B. Furthermore, in T_P the expert is more likely to behave in a consumer-friendly manner after promises *SQ* and *SQ&AP* than after promises *NO* and AP_{UT}.[15]

4.3. Description and prediction for treatment with money-burning option, T_M

In the treatment T_M, which corresponds to game Γ_M, the consumer could decide whether she would like to interact with the expert or not. If not, the game ended and both got zero payment for this particular round. If yes, the consumer

[13] While introducing psychological guilt aversion into our simple credence goods model would allow for sharp point predictions, we formulate weaker statements to illustrate the qualitative effects that we expect from guilt aversion. For instance, we translate the point prediction (from Observation 1) that an expert never overtreats a customer into the weaker statement that overtreatment is rarely observed.

[14] The exact wording of the experimental instructions (see Appendix B) refers to the consumer's two problems as "Problem I" (minor problem) and "Problem II" (major problem), and to the two qualities of the good as "Solution I" (low quality) and "Solution II" (high quality). In the text we rephrase the wording so that it matches the terminology used in the theory part of the paper.

[15] Remember that AP_{UT} stands for the promise *AP* under an undertreatment price-vector (i.e., a price-vector satisfying $P_h - C_h < P_l - C_l$). Such a price-vector makes providing the low quality more profitable than providing the high quality if the appropriate price is charged for.

Table 1
Relative frequencies of behavior in T_B, conditional on price-vector.

Price-vectors (P_l, P_h)	$\theta = l$				$\theta = h$			
	(l,l)	(l,h)	(h,l)	(h,h)	(l,l)	(l,h)	(h,l)	(h,h)
all	0.05	0.94	0.01	0.01	0.03	0.73	0.00	0.24
(30, 50)	0.05	0.93	0.02	0.00	0.02	0.80	0.00	0.18
(30, 60)	0.04	0.94	0.00	0.02	0.03	0.65	0.00	0.32
(30, 65)	0.02	0.97	0.00	0.02	0.02	0.75	0.00	0.23
(30, 70)	0.02	0.98	0.00	0.00	0.00	0.82	0.00	0.18
(40, 65)	0.04	0.94	0.02	0.00	0.08	0.71	0.00	0.21
(50, 65)	0.14	0.86	0.00	0.00	0.00	0.64	0.03	0.33

could voluntarily pay an "interaction price" m from the discrete set $M = \{0, 5, 10, \ldots, 30\}$ to the experimenter. That is, in the notation introduced in the theory section, we had $\epsilon = 5$ and $x = 6$ in our treatment T_M.[16] Based on our theoretical analysis our prediction for T_M would be that the consumer either rejects, or accepts without burning money. This extreme prediction is based on iterated forward-induction reasoning which requires – for each additional iteration – that players' knowledge of each others' rationality is one level deeper. Results from previous experiments – e.g., on Nagel's guessing game (see Nagel, 1995, or Ho et al., 1998) – suggest that assuming so much sophistication probably means assuming too much. How does the prediction for T_M change if we apply forward-induction reasoning (requiring only two layers of mutual knowledge of rationality) without iteration? Then, if the consumer burns the amount y, the expert infers from this that the consumer expects at least y. That is, in this case we get an increasing relationship between the amount of money the consumer burns and the lower bound on the second-order belief of the expert. Based on this reasoning (and the arguments in the theory section) our prediction for T_M is:

Prediction 3 (*The effect of giving the consumer an option to burn money*). **P3a:** Compared to the baseline treatment T_B consumers are more willing to accept (that is, to choose an $m \in \{0, 5, 10, \ldots, 30\}$) in T_M. **P3b:** Message $m = 0$ is frequently observed in T_M, while messages in $\{5, 10, \ldots, 30\}$ are rarely observed. **P3c:** Conditional on acceptance ($m \neq r$), the expert is more likely to behave in a consumer-friendly manner in T_M than in T_B. Furthermore, in T_M the expert is more likely to behave in a consumer-friendly manner when P_h is higher and when the consumer burns a higher amount of money.

5. Experimental results

In reporting the experimental results we proceed in the same order as in previous sections. That is, we start with the baseline treatment, then discuss the effect of introducing an opportunity to make a promise, and finally report the effects of allowing for a money-burning option.[17] Given that our games are rather complex when being exposed to them for the first time (in particular in treatments with promise options or money-burning options) we concentrate in the subsequent analysis on the final 20 rounds (i.e., rounds 5–24), thus considering rounds 1–4 as an opportunity for subjects to gather experience with the game. When we have an ex ante directional hypothesis we use a one-sided test, otherwise a two-sided test.

5.1. Results for the baseline treatment, T_B

Testing predictions **P1a** and **P1b** directly is difficult, because they refer to complete strategies, while we can only observe an expert's provision and charging decision in one of the two states. We therefore proceed as follows. First, we check whether the aggregate data is qualitatively consistent with the theoretical predictions that (i) a consumer with the minor problem will always receive the low quality service (no overtreatment), and that (ii) undertreatment always comes together with overcharging. Next we exploit the fact that the major problem allows for a neat identification of the behavioral type, while the minor problem does not. Specifically, in analyzing the data we use the fact that for the case where the consumer has the major problem, type-1 behavior yields undertreatment and overcharging, type-2 behavior implies appropriate quality and charging for the quality provided, and type-3 behavior yields appropriate quality and charging the price for the low quality. Consistent with the theoretical prediction overtreatment is rarely observed – overall, it occurred in less than 2% of the possible cases. Also, undertreatment almost always came together with overcharging – undertreatment with honest charging occurred in only about 3% of possible cases. Identifying the relative frequencies of type-1, type-2, and type-3 behavior by looking at experts' choices when they faced a major problem, we see (in Table 1) that 73% of the choices are consistent with type-1 behavior, 24% of the choices are consistent with type-2 behavior, and less than 1% of the choices are consistent with type-3 behavior. Thus, type-3 behavior is practically absent while the rest of **P1a** is largely confirmed.

[16] Note that an "interaction price" of 0 results in trade; thus, $m = 0$ is different from opting out ($m = r$).

[17] In the working paper version of this paper we also present a treatment that combines promises (by the expert) and money burning (by the consumer). For reasons of succinctness, we do not report this treatment here but refer the interested reader to Beck et al. (2010).

Table 2
Relative frequencies of promises in T_P, conditional on price-vector.

Price-vectors (P_l, P_h)	Promises			
	NO	*SQ*	*AP*	*SQ&AP*
all	0.07	0.53	0.14	0.26
(30, 50)	0.09	0.52	0.15	0.23
(30, 60)	0.07	0.55	0.12	0.26
(30, 65)	0.05	0.52	0.16	0.28
(30, 70)	0.06	0.55	0.13	0.26
(40, 65)	0.08	0.52	0.12	0.28
(50, 65)	0.04	0.52	0.19	0.26

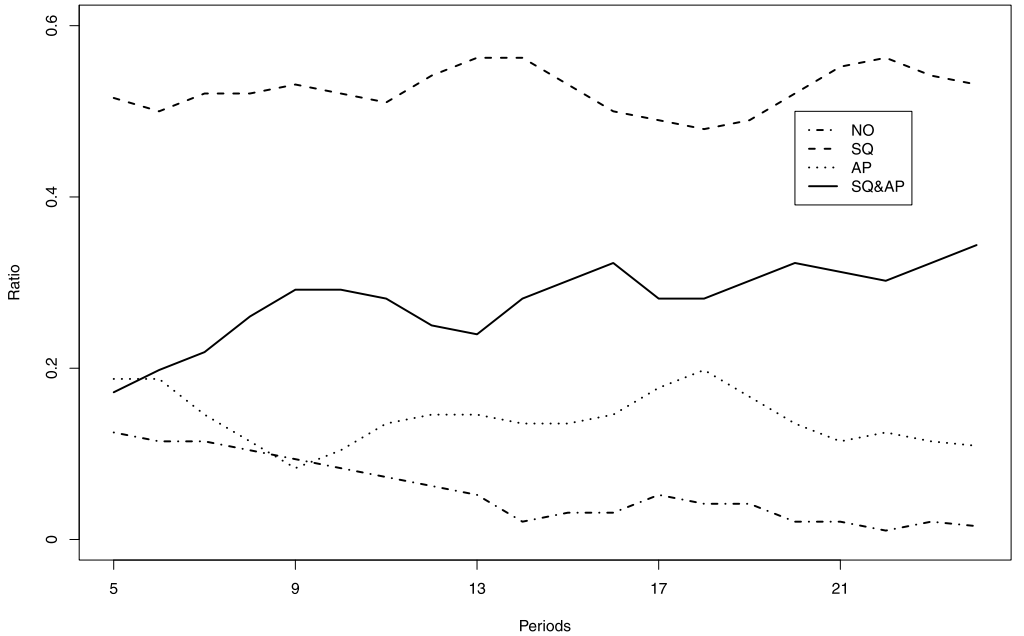


Fig. 4. Development of promises *NO*, *SQ*, *AP*, and *SQ&AP* over time in treatment T_P (smoothed by calculating 3-period moving averages).

In our experiment, P_h can take the value of 50, 60, 65, and 70. Prediction **P1b** says that if psychological forward induction works then type-2 behavior should become more and type-1 behavior less frequent when P_h increases. To see the effects of P_h on provision behavior, we computed the relative frequencies of type-1, type-2, and type-3 behavior for the four values of P_h, again using choices when experts faced a major problem for identification. As we can see from Table 1, there is no evidence supporting **P1b** (one-sided Wilcoxon tests are insignificant).[18]

5.2. Results for treatment with promise options, T_P

Table 2 reports the relative frequencies of promises in treatment T_P. Consistent with **P2a**, we see that promise *SQ* is (with 53% of possible observations) frequently made, while *NO* is (with 7%) rarely made. The distribution of promises is also significantly different from a random one (χ^2 test: $p < 0.01$). As we can see from Table 2, the kind of promise made does not depend on the price-vector. A surprising observation is that a significant proportion of promises (26%) were *SQ&AP* which is a dominated promise for an expert suffering from 'simple guilt'. Even more surprising, this message becomes more popular in later rounds, as can be seen in Fig. 4. This figure displays the development of the four promises over time and shows that while the relative frequency of *SQ* and *AP* remains rather constant, it is increasing for *SQ&AP* and decreasing for *NO*. One possible explanation for this observation is that experts – instead of suffering from 'simple guilt' – suffer

[18] We also searched for evidence in support of **P1b** at the individual level. For instance, defining an expert to be guilt averse if he provides the high quality in at least *x*% of cases where he faces a consumer with the major problem, we asked – for different values of *x* – whether guilt-averse experts behaved in a more consumer-friendly manner when P_h is higher. We found no evidence in support of **P1b** either.

Table 3
Probit regression of interaction with random effects on consumer subjects. The reference is treatment T_B. T_M denotes the treatment with money-burning option, *NO*, *SQ*, *AP*, and *SQ&AP* are the four types of promises in treatment T_P.

	Value	Std. error	t-value	p-value
Intercept	1.175	0.086	13.727	< 0.01
Period	−0.019	0.004	−4.624	< 0.01
T_M	0.171	0.058	2.923	< 0.01
NO	−1.147	0.162	−7.067	< 0.01
SQ	0.483	0.083	5.843	< 0.01
AP	0.088	0.124	0.714	0.475
SQ&AP	0.214	0.097	2.217	0.027

Std. Error of the random effects on individual expert subjects: 0.3453.

from 'guilt from blame'.[19] Also, if consumers believe in a promise whatever its content is, then they are more willing to accept after *SQ&AP* than after *SQ*, giving (selfish) experts an incentive to make the promise *SQ&AP*. This yields a second explanation for the prevalence of *SQ&AP*. We will discuss alternative explanations in Section 6.

To test prediction **P2b** we compare interaction rates. The interaction rate is defined as the relative frequency with which consumers accept to trade with an expert. In treatment T_P it is with 0.85 weakly significantly larger than in T_B with 0.81 (one-sided Wilcoxon test based on 8 independent observations, $p = 0.09$), see Table 4. One reason for the weak effect of promises in T_P against T_B is perhaps that the interaction rate is already rather high in the baseline treatment. The scope of a further increase in the interaction rate is therefore limited. Another possibility is that messages, despite our efforts of making them rich and endogenous, might still be perceived as 'bare'. Charness and Dufwenberg (2010) argue that bare promises are virtually ineffective. When calculating the interaction rates after promise *NO*, *SQ*, *AP*, and *SQ&AP* separately, we see that it is 0.47 after promise *NO*, 0.90 after promise *SQ*, 0.82 after promise *AP*, and 0.85 after promise *SQ&AP*. Thus, by making the right sort of promise (*SQ*) the probability of an interaction not taking place is halved (relative to the baseline treatment), while by making the wrong promise (*NO*) this probability is more than doubled. Comparing the interaction rates with promises to the corresponding value in the baseline treatment with one-sided Wilcoxon tests we find that the difference is significant for promise *NO* (0.47 after *NO* vs 0.81 in the baseline, $p = 0.001$), but not significant for the other promises ($p > 0.10$).[20] Since some matching groups only had very few observations for specific promises, we also ran a Probit regression with random effects in order to see more clearly the effects of promises and to compare them directly to the baseline treatment. The dependent variable is the binary interaction value, 0 or 1. The independent variables include a dummy variable for the treatment with money-burning options, T_B (to be discussed below), and four dummy variables for the promises *NO*, *SQ*, *AP*, and *SQ&AP*. Random effects are on the level of the individual consumer. The results of the regression are reported in Table 3. As we can see from this table, compared to the baseline treatment T_B consumers are significantly more likely to accept after promises *SQ* and *SQ&AP* and less likely to accept after promise *NO*. The interaction rate after promise *AP* is not significantly different from that of the baseline treatment T_B. Thus, consumers' acceptance behavior is largely consistent with **P2b**. Table 3 also shows that the consumer's acceptance is significantly decreasing over time (by about 2 percentage points per period), indicating that consumers shy away more often from interaction the more experience they have gained in the market.

To test prediction **P2c** we shall use two proxies to capture how consumer-friendly an expert is: the undertreatment rate and the honesty rate. The undertreatment rate is defined as the ratio of cases where the consumer actually got undertreated (having the major problem and receiving the low quality) over all cases where the consumer agreed to interact and had the major problem. The honesty rate is defined as the ratio of honest behavior (getting the needed quality and paying accordingly) over all cases with interaction.[21]

Table 4 compares treatments T_B and T_P. The aggregate picture emerging from the first two columns suggests that the possibility to make promises decreases the undertreatment rate and increases the honesty rate. However, none of these differences between T_P and T_B is statistically significant (one-sided Wilcoxon rank-sum tests: $p > 0.10$). Compared to experts' behavior after promise *NO*, experts were significantly more friendly after promise *SQ*: the undertreatment rate decreases from 1.00 to 0.49 and the honesty rate increases from 0.00 to 0.24 ($p < 0.01$ for both). Experts' behavior after promise *SQ&AP* is weakly significantly more friendly than after promise *NO*: the undertreatment rate decreases from 1.00 to 0.71 and the honesty rate rises from 0.00 to 0.16 ($p < 0.10$ for both). Column AP_{UT} of Table 4 displays the undertreatment rate and the honesty rate after promise *AP* when experts faced undertreatment price-vectors ($(30, 50)$, $(40, 65)$, $(50, 65)$). In line with prediction **P2c** experts behave significantly more friendly after promise *SQ* than after promise AP_{UT}: the undertreatment rate decreases from 0.76 to 0.49 ($p < 0.01$) and the honesty rate rises from 0.14 to 0.24 ($p = 0.016$). However,

[19] An expert suffering guilt from blame cares about how strongly the consumer thinks that the expert intended to harm her. Since overcharging remains unobservable, the consumer might hold the belief that the expert always charges appropriately. Under this belief, she would assign blame for the high price to nature having chosen the major problem and the expert would therefore never feel guilty from being blamed for overcharging.

[20] The effects of promises on experts' payoffs are largely the same as the effects of promises on interaction rates.

[21] The overcharging rate, defined as the ratio of cases where consumers actually got overcharged (paying the high price while receiving the low quality) over all cases where consumers agreed to interact with the expert and received the low quality could also be used as a proxy for consumer-friendliness. We dropped it from the analysis because it is very close to 1.00 in all treatments, leaving the treatment effects untestable.

Table 4
The effects of promises, comparing T_B and T_P. The numbers of observations with which the relative frequencies are calculated are reported in brackets.

Treatment	T_B	T_P					
		Overall	*NO*	*SQ*	*AP*	*SQ&AP*	AP_{UT}
Interaction rate	0.81	0.85	0.47***	$0.90_{\diamond\diamond\diamond}$	$0.82_{\diamond\diamond}$	0.85	0.76
	(516/640)	(546/640)	(17/36)	(303/335)	(74/90)	(152/179)	(35/46)
Undertreatment rate	0.77	0.60	1.00***	$0.49^{\triangleright\triangleright\triangleright}_{\diamond\diamond\diamond}$	0.64	$0.71_{\diamond}$	0.76
	(202/264)	(164/275)	(8/8)	(70/140)	(25/39)	(61/86)	(13/17)
Honesty rate	0.14	0.21	0.00***	$0.24^{\triangleright\triangleright}_{\diamond\diamond\diamond}$	0.20	$0.16_{\diamond}$	0.14
	(72/516)	(114/546)	(0/17)	(74/303)	(15/74)	(25/152)	(5/35)

*** / ** / * significantly different from treatment T_B at the 1% / 5% / 10% level.
$\diamond\diamond\diamond$ / $\diamond\diamond$ / $\diamond$ significantly different from promise *NO* at the 1% / 5% / 10% level.
$\triangleright\triangleright\triangleright$ / $\triangleright\triangleright$ / $\triangleright$ significantly different from promise AP_{UT} at the 1% / 5% / 10% level.

experts do not behave significantly differently after promise AP_{UT} than after promise *SQ&AP* ($p > 0.10$ for the comparison of the undertreatment rate and the honesty rate).

5.3. Results for treatment with money-burning option, T_M

We now turn to the effects of giving the consumer a money-burning option. The evidence for **P3a** is, at best, weak. Compared to the baseline treatment T_B, the interaction rate is slightly higher in treatment T_M. The difference is not significant when using one-sided Wilcoxon test with 8 independent observations (0.81 in T_B vs 0.84 in T_M, $p > 0.10$), but is significant in a Probit regression when controlling for heterogeneity of consumers (see Table 3). Consistent with **P3b**, we find that among the cases where consumers accept, message $m = 0$ is observed much more frequently (with 76% of cases) than messages in $\{5, 10, \ldots, 30\}$ (with 24%). Recall that, for guilt-sensitivity parameter $\gamma \in [\frac{C_h - C_l}{P_h + x\epsilon}, \frac{C_h - C_l}{P_h})$, both type-1 and type-2 behavior is consistent with psychological forward-induction reasoning in the baseline game, while only type-2 behavior is consistent with it in T_M for any $m \in \{0, 5, \ldots, 30\}$. Thus, if the mass of experts with $\gamma \in [\frac{C_h - C_l}{P_h + x\epsilon}, \frac{C_h - C_l}{P_h})$ is non-negligible and if multiplicity in predicted paths results in some players coordinating on one and others on the other path, we would expect significantly more consumer-friendly behavior in T_M than in T_B. Comparing T_M with T_B we find, however, that experts did not behave in a more consumer-friendly way: the undertreatment rate even rises from 0.77 to 0.83 while the honesty rate decreases from 0.14 to 0.10, though both comparisons are not significant ($p > 0.10$). Finally, **P3c** suggests that in T_M experts are more likely to behave in a consumer-friendly manner when P_h is higher and when the consumer burns more money. To test the first part of this hypothesis, we compared the undertreatment rate and the honesty rate when P_h is low ($P_h = 50, 60$) to the corresponding values when P_h is high ($P_h = 65, 70$). In line with the prediction a higher P_h leads to more consumer-friendly behavior: When P_h is low, the average undertreatment rate is 0.89 and the average honesty rate is 0.05, while when P_h is high the corresponding values are 0.78 and 0.12. One-sided Wilcoxon rank-sum tests show that the differences are significant ($p = 0.06$ for undertreatment rate and $p = 0.04$ for honesty rate), lending some support to prediction **P3c** (see Table 5). To test the second part of the hypothesis (derived under the assumption that forward induction works, but iterated forward induction does not) we compared the undertreatment rate and the honesty rate for $m = 0$ to the corresponding values for $m \geqslant 5$. Qualitatively in line with the prediction, burning money leads to more consumer-friendly behavior: For $m \geqslant 5$ the average undertreatment rate is 0.78 and the average honesty rate is 0.12, while the corresponding values for $m = 0$ are 0.83 and 0.02. None of these differences is statistically significant, however.[22] The forward-induction argument is tricky, and arguably takes some time to figure out. We therefore also searched for interesting time trends, but failed to find any. Thus, our overall conclusion regarding the money-burning option is that giving the consumer such an option has at best a minor effect.

6. Discussion and conclusion

Credence goods markets are characterized by informational asymmetries between consumers and experts, which can lead to various forms of fraud, such as undertreatment, overtreatment, or overcharging. Dulleck et al. (2011) have investigated the role of institutions and market conditions on experts' behavior and have shown that liability is important for efficiency on credence goods markets, while ex post verifiability of the expert's actions is not. The impact of soft factors, such as giving the expert an option to make a promise to consumers, has been ignored so far as potentially important for improving efficiency on credence goods markets.

Soft factors such as non-binding promises have no effect on market behavior if all players are rational and only interested in their own material payoff, and if this fact is common knowledge. However, if agents have belief-dependent preferences, soft factors might become behaviorally relevant. This paper has focused on a specific kind of belief-dependent motivation,

[22] We also tested whether it pays for a customer to burn money; on average, it does not.

Table 5
The effects of money-burning option in T_M, compared to T_B. The numbers of observations with which the relative frequencies are calculated are reported in brackets.

Treatment	T_B	T_M				
P_h	all	all	50 or 60	65 or 70	all	all
Money burning	n.a.	all	all	all	0	$\geqslant 5$
Interaction rate	0.81	0.84	0.91	0.81	–	–
	(516/640)	(672/800)	(238/262)	(434/538)	–	–
Undertreatment rate	0.77	0.82	0.89	0.78*	0.83	0.78
	(202/264)	(274/335)	(101/114)	(173/221)	(213/257)	(61/78)
Honesty rate	0.14	0.10	0.05	0.12**	0.09	0.12
	(72/516)	(67/672)	(13/238)	(54/434)	(48/510)	(19/162)

*** / ** / * significantly different between $P_h = 50, 60$ and $P_h = 65, 70$ at the 1% / 5% / 10% level.
n.a. not available in T_B.

aversion against 'simple guilt' – as formally modeled by Battigalli and Dufwenberg (2007). Specifically, it has tested the effects of two devices that are predicted to have no impact on behavior if agents have standard preferences, but have the power to manipulate beliefs and are therefore predicted to have an impact if players are guilt averse: (i) an opportunity for the expert to make a non-binding promise; and (ii) an opportunity for the consumer to burn money. In belief-based guilt aversion theory the first opportunity shapes an expert's behavior if an appropriate promise is made and if it is expected to be believed by the consumer; by contrast, the second opportunity might change behavior even though this option is never used along the predicted path.

The results from an experiment with 208 participants confirm the behavioral relevance of promises. Most experts make the predicted promise and proper promises induce more consumer-friendly behavior. Also, by making the right sort of promise (*SQ*) the probability of an interaction not taking place is halved (relative to the baseline treatment), while by making the wrong promise (*NO*) this probability is more than doubled. While our data are largely in line with the predictions derived from guilt aversion theory, it is important to mention that 'simple guilt' cannot explain every detail of our data. Notably, making the promise *SQ&AP* is inconsistent with the hypothesis that experts suffer from simple guilt, yet this message is fairly popular and becomes even more frequent over time, and is associated with high acceptance rates. A possible explanation is that experts – instead of suffering from 'simple guilt' – have an aversion against 'guilt from blame' (as defined in Battigalli and Dufwenberg, 2007), and that consumers anticipate that. An alternative explanation is that experts dislike disappointing consumer's behavior expectations. While the hypothesis that players are inclined not to hurt others relative to their *payoff* expectations and the alternative story that players are inclined not to hurt others relative to their *behavior* expectations are indistinguishable in the simple environments usually studied in lab experiments, they yield different predictions in our richer credence good game. The empirical relevance of guilt based on behavior expectations (which allows for state-dependent payoff expectations while the current version of guilt aversion does not) seems worth exploring in future research.[23] A further explanation for the prominence of promise *SQ&AP* is that a promise has (and is expected to have) a commitment value which is independent of its impact on the expectation of the consumer (as in Ellingsen and Johannesson, 2004, for instance), or that experts have an expectations-unrelated aversion against lying (Gneezy, 2005). As it should be expected, our data cannot give a clear-cut answer to the question which of these alternative explanations for promise-keeping is responsible for the prominence of the promise *SQ&AP*. More generally, we consider it unlikely that a single theory is able to explain all instances of promise-keeping across contexts and subjects and we think that more research is needed to find out which of the mentioned motivational stories is empirically relevant for which fraction of subjects in which context.[24]

Concerning the second device to influence the behavior of experts, we fail to confirm the behavioral relevance of allowing for a money-burning option of the consumer. The theoretical prediction here would have been that giving the consumer an opportunity to burn money changes the behavior of the expert even when the option is not used. That this prediction is not supported by our data is perhaps not surprising – after all it relies on iterated forward induction and thereby requires many layers of mutual knowledge of rationality. Perhaps more surprising is the observation that the milder hypothesis (requiring only two layers of mutual knowledge of rationality) that consumers who burn more money are treated better is not supported by the data either.[25]

Our overall conclusion is that promises are behaviorally relevant but that their impact cannot be explained by the theory of 'simple guilt' alone, and that money burning is less powerful in practice than in theory.

[23] A previous version of this paper contains a model of guilt based on behavior expectations.
[24] Some important steps in that direction have already been done – for instance, by Vanberg (2008) who presents a nice experimental design intended to discriminate between the expectation- and the commitment-based explanation for promise-keeping, and by Ellingsen et al. (2010) who directly test the expectation-based explanation.
[25] This is, however, somewhat in line with earlier findings that the size of player 1's outside option in 'lost wallet games' does not affect player 2's response behavior; see Dufwenberg and Gneezy (2000), Cox et al. (2010).

Appendix A. Proofs

A.1. Formal results for the baseline game, Γ_B

Consider the baseline game, Γ_B. Let $\sigma_e = (\sigma_{\tau,\eta|\theta})_{(\tau,\eta,\theta)\in\{l,h\}^3}$ denote a mixed strategy of the expert, where $\sigma_{\tau,\eta|\theta}$ stands for the probability with which the expert provides quality C_τ and charges price P_η, given the consumer has accepted and has problem θ. Similarly, let $\alpha_a = (\alpha_{\tau,\eta|\theta})_{(\tau,\eta,\theta)\in\{l,h\}^3}$ denote the consumer's initial (first-order) belief about the expert's behavior in case of acceptance, where $\alpha_{\tau,\eta|\theta}$ stands for the probability the consumer assigns to the event that the expert provides quality C_τ and charges price P_η, given she has problem θ. Finally, let $\beta_a = (\beta_{\tau,\eta|\theta})_{(\tau,\eta,\theta)\in\{l,h\}^3}$ denote the expert's expectation of α_a, *conditional* on the consumer's acceptance.

We start by proving formal versions of the three observations in the main text. For this purpose, let $(\tau,\eta|\theta) \succcurlyeq (\tau',\eta'|\theta)$ stand for the statement "the expert prefers (τ,η) over (τ',η') in state θ", and define the symmetric ($\sim$) and the asymmetric ($\succ$) relation similarly. Then:

Observation 1. *For any β_a and any γ, $(l,l|l) \succ (h,l|l)$ and $(l,h|l) \succ (h,h|l)$.*

Proof. The proof is by contradiction. First suppose $(l,l|l) \preccurlyeq (h,l|l)$. Then

$$P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_l), 0\} \leqslant P_l - C_h - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_l), 0\} \quad \Leftrightarrow \quad C_h - C_l \leqslant 0,$$

which is in violation of the assumption $C_h > C_l$. The proof for $(l,h|l) \succ (h,h|l)$ is similar. □

Observation 2. *For any β_a and any γ, $[(l,l|l) \succ (l,h|l) \Leftrightarrow (h,l|h) \succ (h,h|h)]$ and $[(l,h|l) \succ (l,l|l) \Leftrightarrow (h,h|h) \succ (h,l|h)]$. Furthermore, for any β_a, $[(l,l|l) \sim (l,h|l) \Leftrightarrow (h,l|h) \sim (h,h|h)]$, and $(l,l|l) \sim (l,h|l)$ for at most a single value of γ.*

Proof. The first part of the statement follows from the fact that $E_{\beta_a}(\pi_c)$ does not depend on the realizations of η and θ and that – for the case where the consumer's problem is solved – her actual payoff does not depend on the realization of those variables either. Thus, when the expert decides to solve the consumer's problem then charging P_l instead of P_h involves exactly the same costs and benefit in both states; in both states the cost is material and amounts to $P_h - P_l$, and in both states the benefit consists in reducing the psychological cost by $\gamma \max\{E_{\beta_a}(\pi_c) - (V - P_h), 0\} \leqslant \gamma(P_h - P_l)$. This latter consideration also implies that for each β_a, indifference is consistent with at most a single value of γ: for $E_{\beta_a}(\pi_c) > V - P_h$ it is consistent with $\gamma = (P_h - P_l)/[E_{\beta_a}(\pi_c) - V + P_h]$; and for $E_{\beta_a}(\pi_c) \leqslant V - P_h$ it is inconsistent with any $\gamma \geqslant 0$. □

Observation 3 (part a). *For any β_a and any γ, $(l,l|h) \succcurlyeq (l,h|h) \Rightarrow (h,h|h) \succ (l,l|h)$.*

Proof. First note that $(l,l|h) \succcurlyeq (l,h|h)$ is equivalent to $P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_l, 0\} \geqslant P_h - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_h\}$. Now $E_{\beta_a}(\pi_c) \geqslant q_l V - P_h \ \forall \beta_a$; furthermore, $q_l V - P_h + P_l \geqslant 0$, by assumption (3). Thus, $E_{\beta_a}(\pi_c) + P_h \geqslant E_{\beta_a}(\pi_c) + P_l \geqslant 0$, implying that the above inequality reduces to $\gamma \geqslant 1$. Now suppose $(l,l|h) \succcurlyeq (h,h|h)$, in violation of the claim. Then $P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_l, 0\} \geqslant P_h - C_h - \gamma \max\{E_{\beta_a}(\pi_c) - V + P_h, 0\} \geqslant P_h - C_h - \gamma[E_{\beta_a}(\pi_c) - V + P_h]$. Using again the fact that $E_{\beta_a}(\pi_c) + P_l \geqslant 0$ we get $C_h - C_l \geqslant P_h - P_l + \gamma[V - P_h + P_l]$ or $[(C_h - C_l) - (P_h - P_l)]/[V - (P_h - P_l)] \geqslant \gamma$. But, since $V > C_h - C_l$ by assumption (1), the LHS of this inequality is strictly smaller than 1, contradicting $\gamma \geqslant 1$. □

Observation 3 (part b). *For any β_a and any γ, $(h,l|h) \succcurlyeq (h,h|h) \Rightarrow (h,l|h) \succ (l,l|h) \succcurlyeq (l,h|h)$.*

Proof. From the proof of Observation 2 we know that $(h,l|h) \succcurlyeq (h,h|h)$ requires $E_{\beta_a}(\pi_c) > V - P_h$ and $\gamma \geqslant (P_h - P_l)/[E_{\beta_a}(\pi_c) - V + P_h] \geqslant 1$, where the last inequality follows from $E_{\beta_a}(\pi_c) \leqslant V - P_l \ \forall \beta_a$. First suppose $(l,l|h) \succcurlyeq (h,l|h)$, in violation of the claim. Then $P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_l, 0\} = P_l - C_l - \gamma[E_{\beta_a}(\pi_c) + P_l] \geqslant P_l - C_h - \gamma \max\{E_{\beta_a}(\pi_c) - V + P_l, 0\} = P_l - C_h$, where the former equality follows from $E_{\beta_a}(\pi_c) > V - P_h > 0$ and the latter (again) from $E_{\beta_a}(\pi_c) \leqslant V - P_l \ \forall \beta_a$. But then $C_h - C_l \geqslant \gamma[E_{\beta_a}(\pi_c) + P_l] = \gamma[E_{\beta_a}(\pi_c) - V + P_h] + \gamma(V - P_h + P_l) \geqslant P_h - P_l + \gamma(V - P_h + P_l)$, where the last inequality follows from the lower bound on γ mentioned at the start of the proof. But $0 \geqslant P_h - C_h + C_l + \gamma(V - P_h) + (\gamma - 1)P_l$ is inconsistent with $P_h > C_h$, $C_l \geqslant 0$, $V \geqslant P_h$, $\gamma \geqslant 1$ and $P_l \geqslant 0$. Next suppose $(l,h|h) \succ (l,l|h)$, in violation of the claim. Then $P_h - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_h, 0\} > P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_l, 0\} \Leftrightarrow P_h - P_l > \gamma(P_h - P_l)$, where the $\Leftrightarrow$ follows from the lower bound on $E(\pi_c)$ given at the start of the proof (which implies that the max operator selects the first term in {.} on both sides of the inequality). But $\gamma < 1$ violates the lower bound on γ given at the start of the proof. □

Observations 1, 2, 3a and 3b together imply the following result:

Lemma 1. *Generically, the expert's behavior in the subgame starting after the consumer has accepted is of one of the following three types*:

- ***type*-1 *behavior*:** *the expert provides the low quality and charges the price for the high quality in both states*;
- ***type*-2 *behavior*:** *the expert provides the low quality if the consumer has the minor problem and the high quality if the consumer has the major problem, and he charges the price for the high quality in both states*;
- ***type*-3 *behavior*:** *the expert provides the low quality if the consumer has the minor problem and the high quality if the consumer has the major problem, and he charges the price for the low quality in both states.*

Given Lemma 1, the requirement that in equilibrium beliefs of all orders have to be consistent with actual behavior, and the facts that (i) type-1 behavior implies $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = q_l V - P_h$, (ii) type-2 behavior implies $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = V - P_h$, and (iii) type-3 behavior implies $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = V - P_l$, it is straightforward to prove the following results[26]:

Proposition 1a. *Generically the baseline game Γ_B admits three types of PBE*:

- *in* ***type*-1** *equilibria the consumer rejects and the expert exhibits type*-1 *behavior in case of acceptance*;
- *in* ***type*-2** *equilibria the consumer accepts and the expert exhibits type*-2 *behavior in case of acceptance*;
- *in* ***type*-3** *equilibria the consumer accepts and the expert exhibits type*-3 *behavior in case of acceptance.*

The bounds for existence of each of the three types of PBE are shown in Fig. 2.

Proof. We derive the bounds for existence of each of the three types of equilibria in turn.

Bounds for type-1 equilibrium: In a type-1 equilibrium $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = q_l V - P_h$. We have to show that for this expectation type-1 behavior (as described in Lemma 1) is optimal for the expert for any $\gamma \leqslant (C_h - C_l)/q_l V$, while it is suboptimal outside this range. **For** $\theta = l$ Observation 1 implies that we only need to show that for the above expectation and for guilt-sensitivity parameters in the given range we have $(l, h, |l) \succcurlyeq (l, l|l)$. Suppose not. Then $P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_l), 0\} = P_l - C_l > P_h - C_l - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_h), 0\} = P_h - C_l$, which violates the assumption that $P_h \geqslant P_l$. **For** $\theta = h$ the fact that $(l, h|l) \succcurlyeq (l, l|l)$ (together with Observation 2) implies that we need to check that $(l, h|h) \succcurlyeq (h, h|h)$ and $(l, h|h) \succcurlyeq (l, l|h)$ hold for parameters within the stated range and that at least one of these restrictions is violated for parameters outside the range. First note that $(l, h|h) \succcurlyeq (h, h|h)$ is equivalent to $P_h - C_l - \gamma \max\{q_l V - P_h - (-P_h), 0\} = P_h - C_l - \gamma q_l V \geqslant P_h - C_h - \gamma \max\{q_l V - P_h - (V - P_h), 0\} = P_h - C_h$. Thus, this inequality is equivalent to $C_h - C_l \geqslant \gamma q_l V$ or $\gamma \leqslant (C_h - C_l)/q_l V$. Next note that $(l, h|h) \succcurlyeq (l, l|h) \Leftrightarrow [P_h - C_l - \gamma \max\{q_l V - P_h - (-P_h), 0\} \geqslant P_l - C_l - \gamma \max\{q_l V - P_h - (-P_l), 0\}] \Leftrightarrow [P_h - C_l - \gamma q_l V \geqslant P_l - C_l - \gamma(q_l V - P_h + P_l)] \Leftrightarrow [P_h - P_l \geqslant \gamma(P_h - P_l)] \Leftrightarrow \gamma \leqslant 1$, which is satisfied because $\gamma \leqslant (C_h - C_l)/q_l V$ and $V > (C_h - C_l)/q_l$, by assumption (3).

Bounds for type-2 equilibrium: In a type-2 equilibrium $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = V - P_h$. We have to show that for this expectation type-2 behavior (as described in Lemma 1) is optimal for the expert for any $\gamma \geqslant (C_h - C_l)/V$, while it is suboptimal outside this range. **For** $\theta = l$ the check is exactly the same as for type-1 equilibrium. **For** $\theta = h$ we need to check that $(h, h|h) \succcurlyeq (l, h|h)$ and $(h, h|h) \succcurlyeq (l, l|h)$ hold for parameters within the stated range and that at least one of these restrictions is violated for $\gamma < (C_h - C_l)/V$. First note that $(h, h|h) \succcurlyeq (l, h|h) \Leftrightarrow [P_h - C_h - \gamma \max\{V - P_h - (V - P_h), 0\} \geqslant P_h - C_l - \gamma \max\{V - P_h - (-P_h), 0\}] \Leftrightarrow [\gamma V \geqslant C_h - C_l] \Leftrightarrow \gamma \geqslant (C_h - C_l)/V$. Next note that $(h, h|h) \succcurlyeq (l, l|h) \Leftrightarrow [P_h - C_h - \gamma \max\{V - P_h - (V - P_h), 0\} \geqslant P_l - C_l - \gamma \max\{V - P_h - (-P_l), 0\}] \Leftrightarrow \gamma[(V - P_h + P_l) \geqslant (C_h - C_l) - (P_h - P_l)] \Leftrightarrow \gamma \geqslant [(C_h - C_l) - (P_h - P_l)]/[V - (P_h - P_l)]$, which is fine because $(C_h - C_l)/V \geqslant [(C_h - C_l) - (P_h - P_l)]/[V - (P_h - P_l)]$, by assumption (1).

Bounds for type-3 equilibrium: In a type-3 equilibrium $E_{\alpha_a}(\pi_c) = E_{\beta_a}(\pi_c) = V - P_l$. We have to show that for this expectation type-3 behavior (as described in Lemma 1) is optimal for the expert for any $\gamma \geqslant 1$, while it is suboptimal outside this range. **For** $\theta = l$ we need only to check that for this expectation and those values of γ we have $(l, l|l) \succcurlyeq (l, h|l) \Leftrightarrow [P_l - C_l - \gamma \max\{V - P_l - (V - P_l), 0\} \geqslant P_h - C_l - \gamma \max\{V - P_l - (V - P_h), 0\}] \Leftrightarrow \gamma \geqslant 1$. **For** $\theta = h$ Observation 2 and Observation 3 together imply that we need only to check $(h, l|h) \succcurlyeq (h, h|h) \Leftrightarrow [P_l - C_h - \gamma \max\{V - P_l - (V - P_l), 0\} \geqslant P_h - C_h - \gamma \max\{V - P_l - (V - P_h), 0\}] \Leftrightarrow \gamma \geqslant 1$. □

Proposition 1b. *For any $\gamma > (C_h - C_l)/P_h$ only type*-2 *and type*-3 *equilibria are consistent with forward-induction reasoning in Γ_B.*

Proof. Suppose the consumer accepts. If she is rational, she must believe that the expert behaves in a way that implies a material payoff of at least 0 for her. Thus, if the expert believes that the consumer is rational, then he infers $E_{\alpha_a}(\pi_c) \geqslant 0$ from her acceptance. Thus, $E_{\beta_a}(\pi_c) \geqslant 0$. The proof that $E_{\beta_a}(\pi_c) \geqslant 0$ together with $\gamma > (C_h - C_l)/P_h$ implies $(h, h|h) \succ (l, h|h)$ is by contradiction. Suppose $(l, h|h) \succcurlyeq (h, h|h)$. Then $P_h - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_h, 0\} = P_h - C_l - \gamma[E_{\beta_a}(\pi_c) + P_h] \geqslant P_h - C_h - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_h), 0\}$, which is equivalent to $C_h - C_l \geqslant \gamma[E_{\beta_a}(\pi_c) + P_h] - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_h), 0\}$. Now the RHS of this inequality is minimal when $E_{\beta_a}(\pi_c) > V - P_h$. For this case the inequality reduces to $\gamma \leqslant (C_h - C_l)/V$, which is inconsistent with $V > P_h$ (by assumption (1)) and $\gamma > (C_h - C_l)/P_h$. The proof that the two inequalities

[26] For completeness we mention the existence of two non-generic pure-strategy PBE, one in which the expert always provides the appropriate quality and charges accordingly and one in which he always provides the appropriate quality and always charges the "wrong" price. However, since each of those equilibria exists only at a single point [the former at $\gamma = 1/q_l$, the latter at $\gamma = 1/(1 - q_l)$] we ignore them in the sequel.

at the start of the proof also imply $(h,l|h) \succ (l,l|h)$ is again by contradiction. Suppose $(l,l|h) \succcurlyeq (h,l|h)$. Then $P_l - C_l - \gamma \max\{E_{\beta_a}(\pi_c) + P_l, 0\} = P_l - C_l - \gamma[E_{\beta_a}(\pi_c) + P_l] \geqslant P_l - C_h - \gamma \max\{E_{\beta_a}(\pi_c) - (V - P_l), 0\} = P_l - C_h$, where the first equation is implied by $E_{\beta_a}(\pi_c) \geqslant 0$ and the second by $E_{\beta_a}(\pi_c) \leqslant V - P_l$. Thus, we get $(C_h - C_l)/[E_{\beta_a}(\pi_c) + P_l] \geqslant \gamma$, which is inconsistent with $E_{\beta_a}(\pi_c) \geqslant 0$, $P_h > P_l$, and $\gamma > (C_h - C_l)/P_h$. □

A.2. Formal results for the game with promise option, Γ_P

Proposition 2a. *Consider the game Γ_P. Suppose promises act as equilibrium selection devices. Then for any $\gamma \in [(C_h - C_l)/V, (C_h - C_l)/q_l V]$ the expert makes promise SQ and exhibits type-2 behavior after acceptance. For parameter values outside this range the promise is arbitrary and has no effect on expert's behavior.*

Proof. Under the assumptions that promises are believed (and believed to be believed) if and only if believing in the promise is consistent with equilibrium, promises *NO* and *SQ&AP* never have any effect, the former because it has no content and the latter because generically there is no PBE in which the expert provides sufficient quality and charges the price for the quality provided in both states.[27] By contrast, the impact of promises *SQ* and *AP* on the expert's behavior in the subgame starting after the consumer has accepted depends on the realization of γ. For $\gamma < (C_h - C_l)/V$ and for $\gamma \in ((C_h - C_l)/q_l V, 1)$ those promises have no effect because there is a unique prediction in those ranges anyway – see Fig. 2. For $\gamma \geqslant 1$ they have no effect either; for *SQ* this follows from the fact that the expert provides sufficient quality in both equilibria and for *AP* from the observation that the expert charges correctly in none of them. Promises *AP* and *SQ* have an effect in the range $\gamma \in [(C_h - C_l)/V, (C_h - C_l)/q_l V]$. In this range *SQ* selects the type-2 equilibrium, while *AP* selects the type-1 equilibrium; since type-1 equilibrium involves rejection while type-2 equilibrium entails profitable trade, the model predicts that the expert makes promise *SQ* and that he exhibits type-2 behavior after acceptance.[28] For other values of γ promises *SQ* and *AP* have no effect and our model makes no prediction regarding which promise is actually made. □

Proposition 2b. *Suppose promises are believed and believed to be believed in game Γ_P. Then the expert either makes promise SQ or promise AP and each of those promises is accepted by the consumer. The expert's behavior after acceptance of promise $p \in \{NO, AP, SQ, SQ\&AP\}$ is as shown in Fig. 3.*

Proof. Let α_p denote the consumer's (initial) belief about the expert's behavior when he has made promise $p \in \{NO, SQ, AP, SQ\&AP\}$ and the consumer has accepted. Also let β_p denote the expert's expectation of α_p conditional on the consumer's acceptance after the promise p has been made. Then

$$E_{\alpha_{NO}}(\pi_c) = E_{\beta_{NO}}(\pi_c) = q_l V - P_h,$$
$$E_{\alpha_{SQ}}(\pi_c) = E_{\beta_{SQ}}(\pi_c) = V - P_h,$$
$$E_{\alpha_{SQ\&AP}}(\pi_c) = E_{\beta_{SQ\&AP}}(\pi_c) = V - q_l P_l - (1 - q_l) P_h,$$

while the expectation under the promise *AP* depends on the price-vector. It is the same as that for *SQ* for "overtreatment" price-vectors (defined by $P_h - C_h > P_l - C_l$), it is the same as that for *SQ&AP* for "equal mark-up" price-vectors (defined by $P_h - C_h = P_l - C_l$), and it is

$$E_{\alpha_{AP_{UT}}}(\pi_c) = E_{\beta_{AP_{UT}}}(\pi_c) = q_l V - P_l$$

for "undertreatment" price-vectors (defined by $P_h - C_h < P_l - C_l$). Given those expectations it is straightforward to show that the expert's behavior in the subgame starting when the consumer has accepted after promise p has been made is as shown in Fig. 3.[29] First notice that the promise *NO* yields rejection if it is believed. Thus, the expert will never use this promise if he expects that promises are believed. Next notice that the promise *SQ* strictly dominates the promise *SQ&AP* for any $\gamma > 0$. To see this, first note that both promises yield acceptance by the consumer. Furthermore, for $\gamma > 1/q_l$ the promise *SQ&AP* implies a lower price for the consumer than *SQ* at a material cost for the expert, but exactly the same amount of guilt; and for $\gamma \leqslant 1/q_l$ both promises yield exactly the same behavior (and thereby also the same material payoff for the expert), but *SQ&AP* yields a higher amount of guilt than *SQ*. Thus, promise *SQ&AP* will never be made if it is expected to be believed. So we are left with the promises in the set $\{SQ, AP\}$. Here, the comparison is more complicated; it depends on the kind of price-vector under which transactions (potentially) take place, on the magnitude of V and potentially also on γ.[30] □

[27] For completeness we mention that there exists such an equilibrium for a single realization of γ, namely for $\gamma = 1/q_l$. However, since this equilibrium exists only at a single point we ignore it in the sequel.

[28] Our discussion assumes that there is some uncertainty about the expectation and behavior of the other side of the market when there is equilibrium multiplicity.

[29] In the figure AP_{UT} stands for the promise *AP* under an undertreatment price-vector. The behavior after the promise *AP* under an overtreatment vector is the same as after promise *SQ*, and the behavior under an equal mark-up vector is the same as after *SQ&AP*.

[30] Under equal mark-up price-vectors *SQ* dominates *AP* for all V and all γ (because *AP* yields the same payoff expectations as *SQ&AP* for this case); under overtreatment price-vectors the two promises are equivalent to each other (because they yield the same payoff expectations for the consumer); and under

A.3. Formal results for the game with money-burning option, Γ_M

Proposition 3. *Consider the game Γ_M. For any $\gamma \in ((C_h - C_l)/(P_h + x\epsilon), 1)$ the path predicted by (iterated) forward-induction reasoning is unique and it involves full efficiency without money burning.*

Proof. We start by showing that – for guilt-sensitivity parameters in the range $((C_h - C_l)/(P_h + \epsilon x), 1)$ – for any $m \in M = \{0, \epsilon, 2\epsilon, \ldots, x\epsilon\}$ only type-2 behavior is consistent with the assumption of common knowledge of rationality. To see this, first suppose the consumer burns the maximal feasible amount $x\epsilon$. If the expert is rational and believes in the rationality of the consumer when he has to decide in the subgame starting after the amount $x\epsilon$ has been burned, then he infers that $E_{\alpha_{x\epsilon}}(\pi_c) \geqslant x\epsilon$. Therefore, $E_{\beta_{x\epsilon}}(\pi_c) \geqslant x\epsilon$.[31] We now show that an expert with such an expectation necessarily provides C_h in state h for any $\gamma > (C_h - C_l)/(P_h + x\epsilon)$. By the arguments given earlier it is sufficient to show that $(h, h|h) \succ (l, h|h)$ under those conditions. The proof is by contradiction. Suppose $(l, h|h) \succcurlyeq (h, h|h)$. Then $P_h - C_l - \gamma \max\{E_\beta(\pi_c) - (-P_h), 0\} \geqslant P_h - C_h - \gamma \max\{E_\beta(\pi_c) - (V - P_h), 0\}$. Now, since $E_\beta(\pi_c) \geqslant x\epsilon$ this is equivalent to $C_h - C_l \geqslant \gamma[E_\beta(\pi_c) + P_h] - \gamma \max\{E(\pi_c) - (V - P_h), 0\} \geqslant \gamma[E(\pi_c) + P_h] \geqslant \gamma[P_h + x\epsilon]$. Thus, $\gamma \leqslant (C_h - C_l)/(P_h + x\epsilon)$, which is inconsistent with the restriction on γ we started with. Taking the same logic one step further, suppose now the consumer burns the amount $(x-1)\epsilon$. Then the expert infers from that that $E_{\alpha_{(x-1)\epsilon}}(\pi_c) - (x-1)\epsilon \geqslant V - P_h - x\epsilon$. This is so because – if both the expert and the consumer accept the above reasoning – they believe that by burning the amount $x\epsilon$ the consumer can guarantee a material payoff of at least $V - P_h - x\epsilon$. Hence, burning $(x-1)\epsilon$ is consistent with common knowledge of rationality only if this does not yield less. Thus, $E_{\alpha_{(x-1)\epsilon}}(\pi_c) \geqslant V - P_h - \epsilon \geqslant x\epsilon$, where the last inequality is implied by our assumption on x.[32] Now, since $E_{\beta_{(x-1)\epsilon}}(\pi_c) \geqslant x\epsilon$, we can apply the same arguments as above to show that the expert treats the consumer appropriately, even when she burns only $m = (x-1)\epsilon$. Furthermore, for $m = (x-2)\epsilon$ we get – by the same argument – the inequality $E(\pi_c) - (x-2)\epsilon \geqslant V - P_h - (x-1)\epsilon \Leftrightarrow E(\pi_c) \geqslant V - P_h - \epsilon \geqslant x\epsilon$ and thus again appropriate treatment. Proceeding this way further we get to the prediction that for any $m \in M$ (even for $m = 0$!) the expert treats the consumer appropriately. Anticipating appropriate treatment for any $m \in M$, the consumer decides not to burn any money (that is, to choose $m = 0$). □

Appendix B. Instructions

INSTRUCTIONS FOR THE EXPERIMENT

Thank you for participating in this experiment. Please do not talk to any other participant until the experiment is over.

General remarks

The aim of this experiment is to explore choice behavior. During this experiment you and the other participants will be asked to make decisions. By making these decisions you will earn money. Your earnings will depend on your decisions and the decisions of the other participants. You will receive your payment anonymously and in cash. All participants receive the same information about the rules of the game, including the costs and payoffs for certain actions. Neither you nor anybody else will ever be informed about the identity of the participants you interacted with. No one will be informed about the payments to other participants. Please consider all expressions as gender neutral.

2 roles and 24 rounds

This experiment consists of **24 rounds**, each of the 24 rounds consists of the same sequence of decisions. This means that the same situation will be simulated 24 times in a row. There are 2 roles in this experiment: **Player A** and **Player B**. At the beginning of the experiment you will be randomly assigned to one of these two roles. On the first screen of the experiment you will see which role you are assigned to. **This role stays the same throughout the experiment.**

A Player A always interacts with a Player B. In each of the 24 rounds you will be randomly matched to a player of the other type, i.e. the pairs of participants, consisting of a Player A and a Player B, will be determined randomly in each round.

In each round you face the following situation:

Player B has one of two possible problems, either Problem I or Problem II. Both problems are equally likely. The problem of a Player B is only known to the Player A who currently interacts with him – Player B himself does not know his own problem. Player A can now sell a solution to Player B, either Solution I or Solution II. Solution I costs Player A 0 points, Solution II costs him 30 points. For this solution, he can either charge the "Price of Solution I" or the "Price of Solution II"

undertreatment price-vectors the comparison depends upon whether AP is accepted or not (for $P_l \leqslant q_l V$ it is accepted, for $P_l > q_l V$ it is rejected) and in case of acceptance, on whether γ is larger or smaller than $(C_h - C_l)/(q_l V + P_h - P_l)$ (in the former case the two promises yield the same behavior and the same amount of guilt, in the latter the behavior or the amount of guilt differs).

[31] We omit the subscript on the variables α and β when there is little scope for confusion.

[32] If the assumption on x is violated, then the arguments in the main text still work when we replace the restriction $\gamma > (C_h - C_l)/(P_h + x\epsilon)$ by the – then more demanding – restriction $\gamma > (C_h - C_l)/(P_h - \epsilon)$.

from Player B. Those prices are randomly determined in each round and are known to both players. Player A can combine solutions and prices as desired, i.e. if Player A chooses e.g. Solution I, he can either charge the "Price of Solution I" or the "Price of Solution II".

The solution sold by Player A to Player B can either be sufficient or not. Solution II is always sufficient and solves the problem of Player B in any case. Solution I only solves Problem I. The following table shows when a solution is sufficient and when it is not:

Player A chooses ⇓	**Player B has ⇒**	**Problem I**	**Problem II**
Solution I		sufficient	not sufficient
Solution II		sufficient	sufficient

It follows that a solution is not sufficient if Player B has Problem II and Player A chooses Solution I. **Player A's earnings do not depend on whether the solution is sufficient or not.** However, Player B receives 100 points if and only if the solution was sufficient. If the solution is not sufficient Player B receives no points in this round. Still, the price Player A charges for his solution has to be paid by Player B in any case.

Sequence of a round:

Explanations referring to the interaction price only apply in treatments **I** *and* **PI**, *and explanations referring to promises only apply in treatments* **P** *and* **PI**.

1. Player A and Player B get to know the "Price of Solution I" and the "Price of Solution II", which are randomly determined for this round.
2. Player B has then 2 possibilities: He **either quits this round** and does not continue to play. Then both players receive no points for this round and they wait until the next round starts, in which the players will be randomly rematched. **Or he accepts to interact with Player A.**
3. Player A gets to know Player B's problem (either Problem I or Problem II). He now chooses a solution (either Solution I or Solution II) and charges a price for this solution (either the "Price of Solution I" or the "Price of Solution II"). The payoff of Player A is as follows:

+ price, charged from Player B ("Price of Solution I" or "Price of Solution II")
− costs of the solution sold (0 or 30 points)
= payoff for Player A in this round

Player B will neither be informed about the problem he had, nor will he be informed about the solution Player A chose. He will only be told the price he has to pay to Player A and whether the solution chosen by Player A was sufficient or not. The payoff of Player B is as follows:

+ 100 or 0 points (100 if the solution was sufficient, 0 if not)
− price charged by Player A ("Price of Solution I" or "Price of Solution II")
= payoff for Player B in this round

At the beginning of the experiment you receive an initial endowment of 200 points. With this endowment you can cover losses that might occur during some rounds. Also, losses can be covered with gains from other rounds. At the end of the experiment your initial endowment and your payoffs from each round will be summed up. This amount will be converted into cash money using the following exchange rate

80 points = 1 Euro

References

Bacharach, M., Guerra, G., Zizzo, D., 2007. The self-fulfilling property of trust: An experimental study. Theory Dec. 63 (4), 349–388.
Battigalli, P., Dufwenberg, M., 2007. Guilt in games. Amer. Econ. Rev. 97 (2), 170–176.
Battigalli, P., Dufwenberg, M., 2009. Dynamic psychological games. J. Econ. Theory 144 (1), 1–35.
Beck, A., Kerschbamer, R., Qiu, J., Sutter, M., 2010. Guilt from promise-breaking and trust in markets for expert services: Theory and experiment. IZA Discussion Paper 4827. Institute for the Study of Labor (IZA).
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.
Charness, G., Dufwenberg, M., 2010. Bare promises: An experiment. Econ. Letters 107 (2), 281–283.
Cox, J., Servatka, M., Vadovic, R., 2010. Saliency of outside options in the lost wallet game. Exper. Econ. 13 (1), 66–74.
Darby, M.R., Karni, E., 1973. Free competition and the optimal amount of fraud. J. Law Econ. 16 (1), 67–88.
Dawes, R.M., Messick, D.M., 2000. Social dilemmas. Int. J. Psych. 35 (2), 111–116.
Dufwenberg, M., Gneezy, U., 2000. Measuring beliefs in an experimental lost wallet game. Games Econ. Behav. 30 (2), 163–182.
Dulleck, U., Kerschbamer, R., 2006. On doctors, mechanics, and computer specialists: The economics of credence goods. J. Econ. Lit. 44 (1), 5–42.

Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. Amer. Econ. Rev. 101 (2), 526–555.
Ellingsen, T., Johannesson, M., 2004. Promises, threats and fairness. Econ. J. 114 (495), 397–420.
Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G., 2010. Testing guilt aversion. Games Econ. Behav. 68 (1), 95–107.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exper. Econ. 10 (2), 171–178.
Geanakoplos, J., Pearce, D., Stacchetti, E., 1989. Psychological games and sequential rationality. Games Econ. Behav. 1 (1), 60–79.
Gneezy, U., 2005. Deception: The role of consequences. Amer. Econ. Rev. 95 (1), 384–394.
Greiner, B., 2004. An online recruiting system for economic experiments. In: Kremer, K., Macho, V. (Eds.), Forschung und wissenschaftliches Rechnen 2003. In: GWDG Bericht, vol. 63. Gesellschaft für wissenschaftliche Datenverarbeitung, Goettingen, pp. 79–93.
Guerra, G., Zizzo, D.J., 2004. Trust responsiveness and beliefs. J. Econ. Behav. Organ. 55 (1), 25–30.
Ho, T.H., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental "p-beauty contests". Amer. Econ. Rev. 88 (4), 947–969.
Huck, S., Lünser, G., Tyran, J.-R., 2007. Pricing and trust. Working paper. University College, London.
Huck, S., Lünser, G., Tyran, J.-R., 2010. Consumer networks and firm reputation: A first experimental investigation. Econ. Letters 108 (1), 242–244.
Huck, S., Lünser, G., Tyran, J.-R., 2012. Competition fosters trust. Games Econ. Behav. 76 (1), 195–209.
Nagel, R., 1995. Unraveling in guessing games: An experimental study. Amer. Econ. Rev. 85 (5), 1313–1326.
Orbell, J.M., Van de Kragt, A., Dawes, R.M., 1988. Explaining discussion-induced cooperation. J. Pers. Soc. Psych. 54 (5), 811–819.
Ostrom, E., 1998. A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. Amer. Polit. Sci. Rev. 92 (1), 1–22.
Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. Econometrica 76 (6), 1467–1480.
Vanberg, C., 2010. A short note on the rationality of the false consensus effect. Technical report, mimeo.