DISCUSSION PAPER SERIES

IZA DP No. 4827

# Guilt from Promise-Breaking and Trust in Markets for Expert Services: Theory and Experiment

Adrian Beck
Rudolf Kerschbamer
Jianying Qiu
Matthias Sutter

March 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Guilt from Promise-Breaking and Trust in Markets for Expert Services: Theory and Experiment

**Adrian Beck**
*University of Innsbruck*

**Rudolf Kerschbamer**
*University of Innsbruck*

**Jianying Qiu**
*University of Innsbruck*

**Matthias Sutter**
*University of Innsbruck,*
*University of Gothenburg and IZA*

Discussion Paper No. 4827
March 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4827
March 2010

# ABSTRACT

## Guilt from Promise-Breaking and Trust in Markets for Expert Services: Theory and Experiment*

We examine the influence of guilt and trust on the performance of credence goods markets. An expert can make a promise to a consumer first, whereupon the consumer can express her trust by paying an interaction price before the expert's provision and charging decisions. We argue that the expert's promise induces a commitment that triggers guilt if the promise is broken, and guilt is exacerbated by higher interaction prices. An experiment qualitatively confirms our predictions: (1) most experts make the predicted promise; (2) proper promises induce consumer-friendly behavior; and (3) higher interaction prices increase the commitment value of proper promises.

JEL Classification: C72, C91, D82

Keywords: promises, guilt, trust, credence goods, experts, reciprocity

Corresponding author:

Matthias Sutter
Department of Economics
University of Innsbruck
Universitätsstrasse 15
A-6020 Innsbruck
Austria
E-mail: matthias.sutter@uibk.ac.at

* We acknowledge financial support from the Austrian Science Fund (FWF) through grant number P20796 and from the Austrian National Bank (OeNB Jubiläumsfonds) through grant number 13602.

*"I will prescribe regimens for the good of my patients according to my ability and my judgment and never do harm to anyone."*

Excerpt from the Hippocratic Oath

# 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 frequently consumed. Examples include car repair services, where the mechanic knows more about the type of service the vehicle needs than the owner; taxicab 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. Despite the informational asymmetries prevalent in markets for credence goods, the turnover in such markets is huge.[1]

From the viewpoint of standard economic theory (relying on rational, risk-neutral and own-money-maximizing agents) 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).

[1] For example, the online site researchandmarkets.com reports that the U.S. auto repair industry includes 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 is still rising (OECD World Health Statistics 2009).

In this paper, we examine in a theoretical model and an experiment the influence of trust (on the consumer's side) and guilt from promise breaking (on the expert's side) on the efficient provision of credence goods. While institutional safeguards against fraud (like liability and verifiability) and market forces (like competition) have been shown recently to increase efficiency on credence goods markets[2], so far the impact of "soft" factors such as making promises and expressing one's trust has been ignored as possibly important for limiting undertreatment, overcharging, and overtreatment, and thus contributing to the efficient provision of credence goods. The starting point of our paper is the assertion that – due to the informational asymmetries – consumers might be reluctant to enter a credence goods market. Anticipating consumers' worries about the quality of the credence good provided and charged for, experts may try to alleviate consumers' concerns by promising good service (like, e.g., the slogan "Direct Route from A to B" by a British taxi company). Though being cheap talk for own-money-maximizing experts, promises might increase efficiency on credence goods markets if experts feel committed to keep their word (see, e.g., Ellingsen and Johannesson, 2004; Gneezy, 2005; Charness and Dufwenberg, 2006; or Vanberg, 2008).[3] If consumers anticipate the commitment value of promises, proper promises might induce them to enter the market. Entering a credence good market already requires some trust on the consumer's side. We are interested in situations where consumers have in addition an opportunity to communicate the intensity of their trust. Our main hypothesis is that the commitment value of promises increases in the trust expressed by the consumer. We model this hypothesis theoretically and test it experimentally.

[2]See Dulleck et al. (2009) for a comprehensive experimental study on the effects of verifiability, liability, competition, and reputation on markets for credence goods, and Huck et al. (2006, 2007, 2010) on the role of competition and reputation and information exchange on markets for experience goods. For a distinction of the different types of goods see Darby and Karni (1973).

[3]Several explanations for promise keeping have been put forward: In Charness and Dufwenberg's (2006) theory of guilt aversion promises are kept because they influence the payoff expectation of others and because people have a disposition to feel guilty when letting down others' payoff expectations. In other theories, people have an expectations-unrelated preference for keeping their word (see Ellingsen and Johannesson, 2004) or have a cost of lying (Gneezy, 2005). Vanberg (2008) presents an experiment designed to discriminate between different theories.

Our model builds on a simplified version of Dulleck and Kerschbamer's (2006) model of credence goods with one expert and one consumer. We modify this model in two ways: (i) At the beginning of the game we introduce a pre-play stage where the expert can make a non-binding promise to the consumer. The promises available to the expert can be interpreted as different versions of the oaths taken by practitioners in the fields of medicine and law, or simply as some advertisement by the expert through which the consumer is reassured to receive an appropriate quality and/or to pay the correct price. (ii) After the expert has made his promise, the consumer can voluntarily pay a non-negative price in order to interact with the expert. A positive interaction price can be interpreted as an additional cost which may arise for the consumer by trusting a particular expert. For example, the consumer may not visit the doctor closest to her but a doctor recommended by a friend, and by doing so she incurs additional costs of transportation and time.

These two modifications provide an opportunity to contribute to the previous literature in the following ways. The first modification is related to earlier studies which have investigated the effects of non-binding promises (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008). Our novelty here is that, due to the nature of credence goods, promises in our model can take three natural forms, depending on the different dimensions in which fraud can happen on credence goods markets. Each of these three promises implies a particular restraint on the expert's behavior. In that respect, credence goods provide a richer framework (than, e.g., in Charness and Dufwenberg, 2006) in which to examine the effects of different types of promises. In particular, our model allows investigating the endogenous selection among promises that differ in more than one dimension, and their effectiveness on trade in credence goods markets.

The second modification is related to earlier studies investigating the impact of trust on the efficiency of trade on markets where contracts are necessarily incomplete. Our novelty here is that the interaction price paid by a consumer in our setup is not a transfer to the other party (as it is the case in standard gift exchange and trust games; see Fehr et al., 1993, and Berg et al., 1995), but an upfront cost. The attractive feature of modeling the interaction price as an upfront cost and not as a transfer to the expert is that this way we get a measure of *pure trust* defined as a consumer's readiness to be vulnerable to the actions of the expert. Since our measure abstracts from any kind of gift-giving it allows us to study the role of trust for economic outcomes, independent of any motives of material reciprocity (Fehr and Gächter, 2000). Closest to this feature of our model is the treatment "(5,5) A Messages" in Charness and Dufwenberg (2006), where first-movers in a trust game can send arbitrary messages to the second-mover which – at least in principle – allows them to communicate their trust. Charness and Dufwenberg found no effects of the possibility to send such messages on the rate of cooperative behavior, which may be due to the fact that "words are cheap" in their setup. By contrast, in our setting consumers can express their trust in the expert in a costly way. That is, consumers can put their money where their mouth is.[4]

Both modifications taken together allow investigating whether there is an interaction between the two factors guilt from promise-breaking and trust. If a consumer's trust in an expert increases, the costs from breaking a (non-binding) promise might rise for the expert. As a consequence, this might induce the expert not to exploit his informational advantage and, thus, increase efficiency on credence goods markets. So far, we are not aware of any attempt to model and measure the impact of a larger amount of trust on the behavioral effects of a promise. However, such an interaction may turn out to be important for the efficiency in markets with informational asymmetries.

Based on our theoretical model we run an experiment with a total of 272 participants. The experimental results qualitatively confirm our hypotheses. We find that (i) most experts make the kind of promise predicted by our model;

[4]The fact that our measure is based on a *behavioral definition of trust* also distinguishes our study from the literature working with a purely belief-based definition. See Fehr (2009) for a discussion of this literature and the consequences of using different definitions.

(ii) proper promises increase consumer-friendly behavior; and (iii) the higher the interaction price paid by the consumer under proper promises, the nicer experts behave, which clearly indicates an interaction effect of trust and guilt.

The rest of the paper is organized as follows. In Section 2 we introduce the sequential credence goods game, motivate the utility functions of both players, and derive theoretical results for a model in which there are two types of experts, selfish ones who care exclusively for their own monetary payoff, and honest ones who feel guilty if they break their promises when consumers trust them. Section 3 explains the experimental design, and Section 4 reports the experimental results. Finally, Section 5 concludes. All proofs are in the Appendix.

# 2 The Model

## 2.1 Basic Structure

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 $e$ (he) and a consumer $c$ (she). The consumer has a problem $\theta$ that is with an ex ante probability $p$ major ($\theta = h$) and with probability $1 - p$ minor ($\theta = l$). If the consumer decides to visit the expert, the expert finds out the severity of the problem by performing a diagnosis.[5] He then provides a service or good of either high ($h$) or 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 ($\tau = h$) solves both types of problems, while the low quality ($\tau = l$) is only sufficient for the minor problem. Different qualities have different costs for the expert, where $C_\tau$ denotes the costs of quality $\tau$, and where $C_l < C_h$. For the quality he claims to have provided the expert can charge one of two exogenously given prices, either $P_l$ or $P_h$. In what follows, we

[5] For simplicity it is assumed that the diagnosis itself involves no cost and reveals the consumer's problem with certainty (no diagnosis errors).

denote the index of the quality charged for (the expert's "charging decision") by $\iota \in \{l, h\}$. The price the expert charges does not need to correspond to the quality he has actually provided. That is, $\tau$ can be different from $\iota$. If the quality is sufficient, i.e. $\tau = h$ when $\theta = h$ or $\tau \in \{l, h\}$ when $\theta = l$, the consumer receives a payoff of $V_\theta$, where $V_h \geq V_l$; otherwise she receives a payoff of zero. In order to make trade attractive, the exogenously given prices satisfy the following conditions: $P_l > C_l$, $P_h > C_h$, and $V_h > P_h > P_l$.

The sequence of events is as follows. First, both the consumer and the expert observe $P_l$ and $P_h$. Then the consumer decides whether to visit the expert or not. If the consumer decides against a trade, both parties receive an outside payment, where the outside payment for party $i \in \{c, e\}$ is denoted by $o_i$ and where $p(V_l - P_h) + (1 - p)(0 - P_h) < o_c$.[6] If the consumer interacts with the expert, a random move of nature determines the severity of her problem $\theta$, and then the expert decides which quality $C_\tau$ ($\tau \in \{l, h\}$) to provide and which price $P_\iota$ ($\iota \in \{l, h\}$) to charge.[7] We define *overcharging* as charging for the high quality while providing the low one ($\tau = l$, $\iota = h$), and *undertreatment* as providing the low quality when the consumer has the major problem ($\theta = h$, $\tau = l$).

If trade takes place the material payoff of the expert, $\pi_e(\tau, \iota)$, and that of the consumer, $\pi_c(\theta, \tau, \iota)$, are as follows:

$$\pi_e(\tau, \iota) = P_\iota - C_\tau, \quad \iota \quad \text{and} \quad \tau \in \{l, h\}. \tag{1}$$

$$\pi_c(\theta, \tau, \iota) = \begin{cases} -P_\iota & \text{if} \quad \theta = h \text{ and } \tau = l, \\ V_\theta - P_\iota & \text{otherwise.} \end{cases} \tag{2}$$

Notice that in this game neither is the expert obliged to provide a sufficient quality that solves the consumer's problem (no ***liability***), nor can the expert's

[6]This assumption is made to create an interesting tension: Under pure monetary interests of both players the consumer should always opt out, causing a breakdown of the market.

[7]For simplicity we denote both the quality provided and the cost of the quality provided by $C_\tau$. No confusion should result.

action be (perfectly) verified (no ***verifiability***).[8] We denote this basic game by $\Gamma$.

We now modify the basic game $\Gamma$ to the extended game $\Gamma(B, M)$ by adding two stages. First, at the beginning of the game the expert is allowed to send a non-binding message $m \in M = \{NP, LB, VF, HO\}$ to the consumer, where $NP, LB, VF$, and $HO$ are defined as follows:

***NP*:** an irrelevant message (no promise)

***LB*:** a promise to provide a sufficient quality (liability)

***VF*:** a promise to charge the price of the provided quality (verifiability)

***HO*:** a promise to provide the appropriate quality (which matches the problem) and to charge accordingly (honesty)

Secondly, after the expert has made a promise, the consumer, instead of deciding whether to visit the expert or not, needs to state a price $b \in B \cup \{\emptyset\}$ for playing the game, where $B=[0, \bar{B}]$ is a non-empty set of non-negative interaction prices, while the "interaction price" $\emptyset$ stands for not participating and choosing the outside option. In case of participation ($b \in B$) the interaction price $b$ is deducted from the consumer's final monetary payoff (but is not transferred to the expert), and the expert is informed about the magnitude of this price before deciding on the quality provided and the price charged. Accordingly, the consumer's material payoff in the case of participation changes to

$$\pi_c(b, \theta, \tau, \iota) = \begin{cases} -P_\iota - b & \text{if} \quad \theta = h \text{ and } \tau = l, \\ V_\theta - P_\iota - b & \text{otherwise.} \end{cases} \tag{3}$$

[Insert Figure 1 about here.]

[8] In the case of $V_h = V_l$, only when the consumer has a major problem $h$ and the expert provides the low quality, the consumer can indirectly infer her problem type and the quality the expert provided since then she receives no positive value.

The expert's material payoff function stays the same. A game tree considering only material payoffs is shown in Figure 1. If both players care only for their own material payoff, $\Gamma(B, M)$ can be easily solved via backward induction. Let $b(m)$ denote a pure strategy of the consumer and $(m, \tau(m, b, \theta), \iota(m, b, \theta))$ a strategy of the expert.[9] Then, in games involving credence goods without verifiability or liability, the predictions are typically such that in any equilibrium

- $m$ is arbitrary,
- $b(m) = \emptyset \; \forall \; m$, and
- $\tau(m, b, \theta) = l$ and $\iota(m, b, \theta) = h \; \forall (m, b, \theta)$.

That is, whatever the values of $m$, $b$, and $\theta$, the expert always provides the low quality and charges the price of the high quality. Knowing this, the consumer decides against the visit of the expert which leads to a market breakdown. The expert is indifferent in making any promise.

When players have non-standard preferences, however, the introduction of promises and interaction prices might change the behavior of both parties substantially. We argue that the expert feels guilty if the consumer trusts him and he does not keep his word. Furthermore, the more trust the consumer shows by paying a higher interaction price $b$, the more guilt the expert suffers if he breaks his promise.[10] In the next subsection we model these hypotheses theoretically.

## 2.2 Guilt and Trust

We proceed by defining an expert's guilt from promise breaking and the trust of a consumer in the extended game $\Gamma(B, M)$. We first consider the expert. At the

[9] Both strategies might also depend on the exogenously imposed prices, of course. In order not to burden the notation further we omit those variables whenever there is little risk of confusion.

[10] The expert feels guilty if he does not keep his word – this can be due to letting down others' expectations or due to breaking a "moral obligation". See Footnote 3 for references.

beginning of the game, the expert has the opportunity to make a promise. Let $\pi_c^{prom}(m,\theta)$ denote the payoff the consumer receives in state $\theta$ if, lexicographically, the expert first maximizes his material payoff, subject to the constraint that his promise $m$ is kept, and secondly (in case of a tie in own material payoffs) maximizes the payoff of the consumer (over the options that yield him the same monetary payoff).

**Definition 1.** *Given the expert's promise $m$, the consumer's problem $\theta$, and the expert's provision and charging decisions $\tau$ and $\iota$, define*

$$\gamma(m,\theta,\tau,\iota) = \max\{0, \pi_c^{prom}(m,\theta) - \pi_c(\theta,\tau,\iota)\}$$

*as the expert's amount of basic guilt from breaking promise $m$.*

Notice that via this specification the expert feels guilty if he delivers less than what he has promised; but he has no negative guilt if he delivers more than promised.

In line with the trust literature (see e.g. Coleman, 1990, or Rousseau et al., 1998) we define the consumer's trust as the "normalized" amount of money she leaves at risk by paying an interaction price of $b \in B$. Let $\pi_c^{min}(\theta)$ denote the consumer's ex post material payoff if she has a problem of type $\theta$ and the expert behaves selfishly, that is

$$\pi_c^{min}(\theta) = \begin{cases} -P_h & \text{if} \quad \theta = h \\ V_l - P_h & \text{if} \quad \theta = l \end{cases} \tag{4}$$

and define $\bar{\pi}_c^{min} = p\pi_c^{min}(l) + (1-p)\pi_c^{min}(h) = pV_l - P_h$.[11] Similarly, let $\pi_c^{max}(\theta)$ denote the consumer's ex post material payoff in state $\theta$ if the expert behaves honestly, i.e.

[11]Note that $\pi_c^{min}(\theta) = \pi_c^{prom}(NP,\theta)$.

$$\pi_c^{max}(\theta) = \begin{cases} V_h - P_h & \text{if} \quad \theta = h \\ V_l - P_l & \text{if} \quad \theta = l \end{cases} \tag{5}$$

and define $\bar{\pi}_c^{max} = p\pi_c^{max}(l) + (1-p)\pi_c^{max}(h) = p(V_l - P_l) + (1-p)(V_h - P_h)$.[12]

We are now in the position to define the consumer's trust as follows:

**Definition 2.** *Given the consumer's interaction price $b$, define*

$$\psi(b) = \frac{o_c + b - \bar{\pi}_c^{min}}{\bar{\pi}_c^{max} - \bar{\pi}_c^{min}}$$

*as the consumer's amount of trust.*

Our main hypothesis is that the expert's total guilt from promise breaking is influenced by the amount $b \in B$ paid by the consumer. More specifically, the more trust the consumer shows, the more the expert suffers if he delivers less than what he promised.

As before, let $\pi_e(\tau, \iota)$ denote the expert's material payoff if he chooses $(\tau, \iota) \in \{l, h\} \times \{l, h\}$. Then the expert's ex post utility is assumed to be given by

$$U_e(m, b, \theta, \tau, \iota) = \pi_e(\tau, \iota) - \psi(b) \cdot \gamma(m, \theta, \tau, \iota). \tag{6}$$

For future reference we define the following:

**Definition 3.** *Given the expert's promise $m$, the consumer's interaction price $b$, the consumer's problem $\theta$, and the expert's provision and charging decisions $\tau$ and $\iota$, define*

$$\psi(b) \cdot \gamma(m, \theta, \tau, \iota)$$

*as the expert's total amount of guilt.*

As before, let $\pi_c(\theta, \tau, \iota)$ denote the consumer's material payoff if the expert

[12] Note that $\pi_c^{max}(\theta) = \pi_c^{prom}(HO, \theta)$.

chooses $(\tau, \iota)$ in state $\theta$. The consumer's utility function is then simply

$$U_c(b, \theta, \tau, \iota) = \pi_c(\theta, \tau, \iota) - b. \tag{7}$$

## 2.3 Benchmark Solution

We use subgame perfection as the solution concept. Firstly, we solve the subgames which start after the expert's choice of $m$, the consumer's choice of $b$, and nature's choice of $\theta$. After deriving the expert's provision policy $\tau(m, b, \theta)$ and charging policy $\iota(m, b, \theta)$ in those subgames, we proceed by finding the consumer's equilibrium strategy $b(m)$. Finally, we derive the expert's optimal choice of $m$.

For future reference we define three kinds of price vectors ($PV$):

***EM*:** an equal markup price vector, where $P_h - C_h = P_l - C_l$

***UT*:** an undertreatment price vector, where $P_h - C_h < P_l - C_l$

***OT*:** an overtreatment price vector, where $P_h - C_h > P_l - C_l$

Since the three categories are mutually exclusive and comprehensive, we have $PV \in \{EM, UT, OT\}$.

Notice that the consumer has never an incentive to pay an interaction price $b$ greater than $\bar{b} = \bar{\pi}_c^{max} - o_c$, since then her expected payoff is smaller than the value of her outside option. In what follows we therefore restrict our attention to scenarios where $b \leq \bar{b} = \bar{\pi}_c^{max} - o_c$. Let $b^* = \frac{C_h - C_l}{V_h}(\bar{\pi}_c^{max} - \bar{\pi}_c^{min}) + \bar{\pi}_c^{min} - o_c$.

**Result 1.** *Given prices $P_l$ and $P_h$, expert's promise $m$, consumer's interaction price $b$, and nature's choice $\theta$,*

- *the expert's provision policy $\tau(PV, m, b, \theta)$ is given by*
  - $\tau = \theta$ *if* $b \geq b^*$, $\forall\ (PV, m) \in \{EM, UT, OT\} \times \{VF, LB, HO\} \setminus (UT, VF)$,

– *and* $\tau = l$ *otherwise; and*

- *the expert's charging policy* $\iota(PV, m, b, \theta)$ *is given by*

  – $\iota = h \; \forall \; (PV, m, b, \theta)$.

**Result 2.** *Given prices* $P_l$ *and* $P_h$*, and expert's promise* $m$*, the consumer pays* $b = b^*$ *if* $\frac{V_h}{C_h - C_l} \geq 1 + \frac{p(P_h - P_l)}{(1-p)V_h}$ *and*

- $m \in \{LB, HO\} \; \forall \; PV$*, or*

- $m = VF$ *and* $PV \neq UT$*,*

*and the consumer decides against visiting the expert otherwise.*

**Result 3.** *Given prices* $P_l$ *and* $P_h$ *such that* $\frac{V_h}{C_h - C_l} \geq 1 + \frac{p(P_h - P_l)}{(1-p)V_h}$ *the expert makes the promise*

- $m = LB$ *if* $PV \in \{EM, UT\}$*;*

- $m \in \{LB, VF\}$ *if* $PV = OT$*.*

The intuition for the above results is as follows. The promise not to undertreat is potentially attractive because breaking it causes the expert a considerable amount of basic guilt while yielding only a moderate monetary gain. Thus, given an appropriate interaction price, the expert will keep this promise. By contrast, the promise not to overcharge will never be kept. This is because breaking it causes a one-to-one increase in both the monetary component and the basic guilt component in the expert's utility function and because the latter has less weight in this function than the former under any reasonable interaction price. These two observations together imply that messages $LB$ and $HO$ (which both implicitly contain the promise to solve the problem) always imply exactly the same provision and charging behavior, but $HO$ yields more guilt than $LB$ (because $HO$ also contains the promise not to overcharge). Thus, message $HO$ is dominated by message $LB$, implying that the former will never be sent in equilibrium. The payoff promise to the consumer implied by message $VF$ is more complicated since it depends on the price vector imposed. Under $UT$

vectors message $VF$ does not imply a promise to solve the problem, it rather yields exactly the same provision and charging behavior as message $NP$, but more guilt. Thus, under $UT$ vectors message $VF$ is clearly unattractive. Under $EM$ vectors message $VF$ implies the same payoff promise as $HO$, implying that it is unattractive too. Only under $OT$ vectors the $VF$ message is attractive since here it is equivalent to $LB$.

### 2.4 Honest and Selfish Experts: An Extension

Of course, by assuming that all experts have the same preferences, our basic model is simplistic. A more elaborate model would have a population of experts differing in their preferences for keeping their word. An easy way to allow for such a heterogeneity is to add a parameter $\lambda$ in front of the guilt term as specified in Definition 3 and to assume that experts differ in the value of $\lambda$. In this section we analyze the simplest version of such a model, one in which (i) there are two types of experts, selfish ones ($S$-types) with $\lambda = 0$ and honest ones ($H$-types) with $\lambda = 1$; (ii) each expert knows his own $\lambda$; and (iii) consumers only know the distribution of types.

Let $q$ denote the commonly known probability that the expert is honest, and let $\mu(m)$ denote the probability consumers assign to the event that promise $m$ comes from an honest expert. Solving this extended model using Perfect Bayesian Equilibrium (PBE) as solution concept yields two types of pooling PBEs[13] (the details of the derivation can be found in Appendix A). Here we characterize them for the case of undertreatment price vectors. We discuss the other two cases below. In the following propositions we refer to the relative frequency with which consumers trade with the expert (i.e., choose $b \in B$) as the acceptance rate.

**Proposition 1.** *Suppose $PV = UT$. If $q < q^* = \frac{C_h - C_l}{V_h}[1 + \frac{p(P_h - P_l)}{(1-p)V_h}]$, then there is no PBE with positive acceptance rate. If $q \geq q^*$, then there exist two*

[13] We use PBEs as abbreviation for Perfect Bayesian Equilibria.

*types of pooling PBEs (and no other types of PBEs) with positive acceptance rate:*

- *Pooling PBEs in which*
  - *both types of experts make promise $LB$;*
  - *consumers pay $b^*$ with probability 1 when observing $LB$, pay $b^*$ or $\emptyset$ (or mix between $b^*$ and $\emptyset$) when observing $HO$, and do not interact ($b = \emptyset$) when observing $NP$ or $VF$;*
  - *beliefs are $\mu(LB) = q$, while $\mu(NP)$, $\mu(VF)$, and $\mu(HO)$ are arbitrary.*
- *Pooling PBEs in which*
  - *both types of experts make promise $HO$;*
  - *consumers pay $b^*$ with probability 1 when observing $HO$, and do not interact ($b = \emptyset$) when observing any $m \neq HO$;*
  - *beliefs are $\mu(HO) = q$ and $\mu(LB) < q^*$, while $\mu(NP)$ and $\mu(VF)$ are arbitrary.*

*In both kinds of pooling PBEs $H$-types provide the appropriate quality (i.e., $\tau = \theta$) and charge $P_h$ in any state along the equilibrium path, while $S$-types always provide the low quality ($\tau = l$) and charge for the high quality ($\iota = h$).*

**Remark 1.** PBEs with positive interaction rates in which both types of experts make promise $HO$ require beliefs such that $\mu(LB) < q^* \leq q$. But is this a plausible, or "reasonable", belief? After all, an $S$-type could never be made better off by promising $LB$ instead of $HO$, *regardless* of what consumers believe. However, an $H$-type expert is better off when choosing $LB$ if this promise is accepted. Hence, any belief that has $\mu(LB) < q$ seems unreasonable.

We therefore propose a refinement (denoted R1) which is in the spirit of the Intuitive Criterion (Cho and Kreps, 1987). To introduce the refinement, let $u_e^*(t, m)$ be the equilibrium payoff of type $t \in \{S, H\}$ in a PBE in which both types of experts make promise $m$. We say that promise $m'$ is *(weakly) equilibrium dominated* for type $t$ in a PBE in which both types of experts make promise $m$ if $u_e^*(t, m)$ is (weakly) higher than the maximal payoff type $t$ could get by deviating to promise $m'$, *regardless* of the beliefs consumers have after this choice. Then we require the following:

**Refinement R1:** *Consider a pooling PBE in which both types of experts make promise $m$ and the out-of-equilibrium information set corresponding to promise $m' \neq m$.*

*(a) If $m'$ is equilibrium dominated for type $t \in \{S, H\}$, then – if possible – consumers' beliefs at $m'$ should place zero probability on type $t$.*

*(b) If (a) has no bite, then replace "equilibrium dominated" by "weakly equilibrium dominated".*

Note that part (a) of refinement R1 is almost equivalent to Cho and Kreps's (1987) Intuitive Criterion . Part (b) strengthens the Intuitive Criterion further and yields the following result:

**Proposition 2.** *Suppose $PV = UT$. The only PBEs with positive acceptance rate that survive refinement R1 are those where (i) both types of experts make promise $LB$; (ii) consumers pay $b^*$ with probability 1 when observing $LB$ and abstain from paying an interaction price under any other promise; and (iii) beliefs are such that $\mu(LB) = q$, $\mu(HO) = 0$, and $\mu(m)$ is arbitrary for $m \in \{NP, VF\}$.*

**Remark 2.** Note that the refinement R1 does not only eliminate all PBEs in which both types of experts make promise $HO$ (part (b) of R1 is responsible for that), it also restricts beliefs in those pooling PBEs that survive the refinement (which is a result of part (a) of R1): After a deviation to $HO$ consumers *must* believe that this promise comes from an $S$-type for sure. This seems reasonable since promise $HO$ can only make $H$-types worse off compared to what they get in equilibrium.

So far, we have exclusively looked at undertreatment price vectors. How does Proposition 2 look like for equal-markup and overtreatment price vectors? For $PV = EM$, Proposition 2 remains true, provided we impose the requirement $\mu(VF) = 0$ (because $VF$ is equivalent to $HO$ in this case). For $PV = OT$, the story is slightly more complicated because $LB$ and $VF$ imply exactly the same payoff promise for the consumer in this case. Thus, there are pooling PBEs in which (a) $LB$ is accepted with probability 1, while $VF$ is rejected; (b) $VF$ is accepted with probability 1, while $LB$ is rejected; and (c) both promises are accepted with probability 1. Note, however, that all those pooling PBEs lead qualitatively to exactly the same behavior along the equilibrium path.

Summing up we conclude that the predictions for equilibrium behavior of honest types in our two-type model correspond to those in the basic model: Honest types should make promise $LB$ and they should keep their promise if consumers reveal sufficient trust (by paying $b = b^*$). The two-type model adds the insights that selfish experts have always an incentive to mimic honest ones and that out-of-equilibrium promise $HO$ should be interpreted by consumers as a bad signal. This latter insight is implied by the fact that for honest types this promise is too difficult to keep, so they steer clear of it to avoid feeling guilty. Selfish types, on the other hand, do not keep any promise that is in conflict with monetary self interest and they do not feel guilty for their misbehavior. So, they go for the promise that is expected to be accepted with the highest probability. Thus, depending on $S$-types' expectations they might be willing to send $HO$ with positive probability.

## 3 Experimental Design

We let the consumer's probability of having the minor problem be $p = 0.5$ and the valuation for receiving a sufficient quality (i.e. a quality which solves the problem) be $V_l = V_h = 100$ experimental currency units (ECU). The cost of providing the low quality is $C_l = 0$ ECU, the cost of providing the high quality is $C_h = 30$ ECU, and the exchange ratio is 80 ECU= 1 Euro. In order to encourage interaction, we set the outside option of both the expert and the consumer equal to zero ($o_e = o_c = 0$). Each of the 6 vectors ($P_l$,$P_h$) in the set $\{(30, 50), (30, 60), (30, 70), (30, 65), (40, 65), (50, 65)\}$ is exogenously imposed with equal frequency. This set of price vectors includes all types of price vectors, undertreatment price vectors $(30, 50)$, $(40, 65)$, and $(50, 65)$), equal markup price vectors $(30, 60)$, and overtreatment price vectors $(30, 70)$ and $(30, 65)$. Each of these 6 price vectors was played 4 rounds, two times with the consumer having a minor problem and two times with the consumer having a major problem. In total this sums up to 24 rounds for each subject. The sequence of (price vector, problem)-pairs was randomized on the individual expert's level. At the beginning of the experiment, subjects were informed about their (fixed) role as either an expert or a consumer[14], and they received an initial endowment of 200 ECU, equivalent to 2.5 Euro.

In the following we describe the sequence of actions in each round of the experimental treatment that corresponds exactly to the extended game $\Gamma(B, M)$. The changes for the other experimental treatments are introduced below.

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 expert was given an opportunity to send one out of four possible messages:[15]

***NP*:** "Hello."

***LB*:** "I promise I will provide a sufficient quality."

***VF*:** "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."

[14]Experts were called "Player A" and consumers were called "Player B" (see instructions in Appendix B).

[15]The exact wording of the experimental instructions refers to the consumer's minor or major problem as "Problem I" or "Problem II" and to the quality of the good provided as "Solution I" (low quality) or "Solution II" (high quality). In the text we rephrase the wording so that it matches our terminology used in the model.

***HO*:** "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."

The consumer was informed about the chosen message, and she could then 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 could voluntarily pay an interaction price $b$ from the discrete set $B = \{0, 5, 10, ..., 30\}$ to the experimenter. We deliberately set the range of possible interaction prices large in order not to limit or misguide the choice of the consumer.[16]

If the consumer decided to interact with the expert, the expert was reminded what message he had sent and was informed which interaction price the consumer had paid. Then he 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. At the end of the experiment, the payoffs of all rounds were added up to yield the final payoff.

We had 16 subjects in each session. To obtain more than one independent observation for each session, we assigned 4 experts and 4 consumers into one matching group. Experts only interacted with consumers in the same matching group.

In order to isolate the effects of promises and interaction prices, we implemented a $2 \times 2$ experimental design by varying the opportunity of experts to make promises and of consumers to pay an interaction price. This yields the following four treatments.

**B:** In this baseline treatment experts cannot make any promise and consumers cannot pay an interaction price.

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

**I:** Here consumers can pay an interaction price, but experts cannot make promises.

**P:** Experts can make a promise, but consumers can only decide whether or not to trade, without paying a price for it.

**PI:** Experts can send promises and consumers can pay interaction prices.

The experiment was run in June 2009. All sessions were run computerized (Fischbacher, 2007) and recruiting was done with ORSEE (Greiner, 2004). Four sessions were conducted for treatments **B**, **P**, and **PI**, and five for treatment **I**. In total we had 272 undergraduate students from the University of Innsbruck participating in the experiment. At the beginning of each session, the instructions were read aloud to make them common knowledge. Subjects were also given about 10 minutes to read through the instruction 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 Experimental Results

In reporting the experimental results we first show the isolated effects of promises, then the isolated effects of interaction prices, and finally the combined effects of promises and interaction prices. Given that the sequence of decisions in the experiment is rather complex when being exposed to it for the first time (in particular in treatments with promises and interaction prices) we concentrate in the following analysis on the final 20 rounds (i.e., rounds 5–24), thus considering rounds 1–4 as an opportunity for subjects to accumulate experience with the game. When we have an ex ante directional hypothesis, we use a one sided test, otherwise a two sided test.

### 4.1 The Effects of Promises

Figure 2 displays the distribution of promises in treatments **P** and **PI**. Consistent with our theory, we see a clear dominance of promise *LB*, and the distribution of promises is significantly different from a random one ($Chi^2$ test: $p < 0.01$). In order to test for the pure effects of promises, Table 1 compares treatments **B** and **P**.

The overcharging ratio (hereafter **OR**) is 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. The undertreatment ratio (**UR**) is defined as the ratio of cases where the consumer actually got undertreated (having the major problem while receiving the low quality) over all cases where consumers agreed to interact and had the major problem. The honesty ratio (**HR**) is defined as the ratio of honest behavior (getting the needed quality and paying accordingly) over all cases with interaction.

[Insert Figure 2 about here.]

Recall that consumers didn't have the possibility to state a positive interaction price in treatment **P**, since the restriction $b \in \{0, \emptyset\}$ applied. Notice, however, that by agreeing to interact with the expert, a consumer already signals a positive amount of trust, because her expected payoff if she interacts with an own-money-maximizing expert is smaller than the outside option.[17] Additionally, experts could form beliefs about the level of consumer's trust. Therefore, even in treatment **P** we should expect a significant effect of promises.

[Insert Table 1 about here.]

[17] This is true for all price vectors except (30,50), which gives the consumer an expected payoff of zero even if the expert behaves totally selfishly.

The aggregate picture emerging from the first two columns of Table 1 suggests that the possibility to make promises in treatment **P** increases the interaction ratio, while experts undertreat less often and overcharge slightly more often. However, none of these differences between treatments **P** and **B** is statistically significant (two sided Wilcoxon rank sum test: $p > 0.10$). This suggests that simply allowing for (non-binding) communication is not enough to induce strong effects.

However, the content of communication is important and has a significant impact. A comparison of behavior in treatment **B** with behavior contingent on a specific message in treatment **P** yields a much clearer picture. The message $NP$ is considered as a strong negative signal by consumers: the interaction ratio decreases significantly in comparison to treatment **B** (two sided Wilcoxon rank sum test: $p = 0.0023$). In fact, the reaction of consumers is justified, since experts undertreat in 100% of cases with message $NP$ (two sided Wilcoxon rank sum test: $p = 0.0178$), and are never honest in their provision and charging decisions (two sided Wilcoxon rank-sum test: $p = 0.0067$). In contrast, after a promise $LB$ experts undertreat (weakly) significantly less often (two sided Wilcoxon rank-sum test: $p = 0.0650$), albeit they are not significantly more honest (two sided Wilcoxon rank-sum test: $p = 0.1304$). Experts' behavior after promises $VF$ or $HO$ is not significantly different from behavior in treatment **B**, though (two sided Wilcoxon rank-sum test: $p > 0.10$). Experts even overcharged significantly more often after promise $HO$ (two sided Wilcoxon rank-sum test: $p = 0.0128$). This is fully consistent with the (out-of-equilibrium) prediction of our two-type model that promise $HO$ is not attractive for honest types.[18]

A direct comparison of behavior across messages in treatment **P** reveals further interesting patterns. In comparison to message $NP$, promises always increase

[18] We are aware that equilibrium refinements for signaling games find little support in experimental tests (see e.g. Brandts and Holt, 1992). In our context refinement R1 nevertheless predicts well, possibly because our model is simpler and more intuitive than those tested in previous experiments.

the interaction ratio significantly (two sided Wilcoxon rank-sum test: $p < 0.05$ for all promises), but imply rather heterogeneous provision and charging behavior. After promise $LB$, undertreatment decreases and honesty increases substantially (two sided Wilcoxon rank-sum test: $p = 0.0178$ for the **UR** and $p = 0.0074$ for the **HR**). Promises $VF$ and $HO$ seem to drive behavior in the direction implied by the promise, but the effects are not significant (two sided Wilcoxon rank-sum test: $p > 0.10$). This could be due to the following reason. Our two-type model suggests that experts who take promises seriously are less likely to make promises *VF* and *HO*. This implies that the very fact of making promise *VF* or *HO* might signal that the promising expert is less likely to keep his promise. This inference is supported by the observation that the overcharging rate does not differ significantly among promises *VF*, *HO*, and *LB* (two sided Wilcoxon rank-sum tests: $p > 0.10$), although promises *VF* and *HO* – if kept – would rule out overcharging, while promise *LB* would not. Thus, it seems that only particular kinds of promises foster trust and cooperation, or, in more bloomy words, "if you sound like you mean it, the chances are greater that you will do it"(Charness and Dufwenberg, 2008, page 20).

Given the behavioral consequences of promises, it seems interesting to examine whether the distribution of promises is stable over time. Figure 3 shows the smoothed development of the distribution of the 4 kinds of promises over time. Promise *LB* stays rather stable, whereas the frequencies of promises *NP* and *VF* decrease over time, and the frequency of promise *HO* increases over time. Figure 4 takes a closer look at the consequences of promise *HO*. It turns out that the increase of promise *HO* is accompanied by experts taking it less seriously: given promise *HO*, the **UR** increases, the **OR** stays basically constant at a very high level, while the **HR** decreases over time, and accordingly the interaction ratio decreases.

[Insert Figures 3 and 4 about here.]

## 4.2 The Effects of Interaction Prices

We now turn to the pure effects of allowing consumers to pay positive interaction prices to trade with the expert. Recall that experts cannot make promises in treatment **I**. Our model suggests that without the restriction of promises experts may simply behave selfishly. In the spirit of Charness and Dufwenberg (2006), however, experts with let-down aversion might behave nicer when they observe higher interaction prices.

[Insert Table 2 about here.]

Recall that our design allows for 7 different levels of interaction prices. We found that, among those who paid strictly positive interaction prices (24% of subjects with $b \in B$ chose $b > 0$), the majority paid 5 or 10 (5: 37%, 10: 23%, 15: 19%, 20: 16%, 25: 3%, 30: 2%). Since categorizing interaction prices into too many levels would make the number of observations too small for robust inferences, we classify them in the following into two categories, $b = 0$ and $b \geq 5$.

Table 2 reports the **OR**, **UR**, and **HR** for treatments **B** and **I**, considering in the latter also $b = 0$ and $b \geq 5$ separately. While on average the **UR** (**HR**) is smaller (larger) under $b \geq 5$ than under $b = 0$ – which would be compatible with the idea that let-down aversion influences the behavior of experts – statistical tests show that these differences are not significant at any reasonable level. Therefore, it seems that – consistent with our model – "standard" reciprocity or let-down aversion do not play an important role in treatment **I**.

## 4.3 The Combined Effects of Promises and Interaction Prices

We are now ready to discuss the combined effects of promises and interaction prices in treatment **PI**. As we have argued earlier and assumed in our model, the existence of interaction prices gives consumers an opportunity to express their trust. Increased trust might magnify the expert's guilt from promise breaking, and consequently increases the probability of experts keeping their promises. Therefore, experts anticipating positive interaction prices $b$ should make promises that are easier to keep afterwards. Since message *NP* and promise *LB* are easier to be kept than the promises *VF* and *HO*[19], when comparing treatment **P** with **PI** we should observe an increase in the frequencies of messages *NP* and *LB*, and a decrease in the frequencies of promises *VF* and *HO*. one sided two sample tests of proportions for each promise confirm this hypothesis ($p < 0.01$ for each promise; see also Figure 1).[20]

Let us now turn to the key issue of our model: How do different levels of the consumer's interaction price influence the expert's behavior when the latter can make promises? As argued earlier, a higher interaction price might signal more trust by the consumer, and this might magnify the expert's guilt from breaking his promise. If this is true, we should observe consumer-friendlier behavior from experts when the interaction price is higher. On the right-hand side of Table 3 we report the **OR**, **UR**, and **HR** depending on the two categories of interaction prices for treatment **PI**. In order to illustrate the effects of interaction prices as clearly as possible, we also display the results from treatments **B**, **P**, and **I**.

[Insert Table 3 about here.]

Our hypothesis is strongly confirmed. In treatment **PI**, when the interaction price $b$ increases from $b = 0$ to $b \geq 5$, the **UR** decreases significantly (one sided Wilcoxon rank-sum test: $p = 0.0415$), while the **OR** does not change significantly. As a side effect of less frequent undertreatment, the **HR** increases (one sided Wilcoxon rank-sum test: $p = 0.0156$).

[19]Note that as soon as a promise includes a statement about future charging behavior, according to our model $\psi \geq 1$ is needed to outweigh the direct monetary gain from overcharging by the total guilt term $\psi(b) \cdot \gamma(m, \theta, \tau, \iota)$. Since consumers have no incentive to pay such a high interaction price a promise to charge $P_l$ is never kept.

[20]The distribution of promises in treatment **P** is *NP*: 6%, *LB*: 52%, *VF*: 14%, and *HO*: 28%, and in treatment **PI** it is *NP*: 13%, *LB*: 59%, *VF*: 8%, and *HO*: 19%.

One might argue that the above observation is also consistent with experts having distributional preferences (see, e.g., Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000, and Charness and Rabin, 2002). Indeed, since a high interaction price decreases the material payoff of the consumer, inequality averse experts could behave in a way consistent with the above data pattern. This explanation can be rejected, however, by comparing treatments **I** and **PI**. If the observations in treatment **PI** were driven by distributional preferences, one would expect a similar pattern in treatment **I**. But in treatment **I** the interaction price does not have a significant effect on experts' behavior, as shown in subsection 4.2. This suggests that guilt from promise breaking drives our results, rather than distributional preferences.

[Insert Table 4 about here.]

Table 4 reports more detailed results on the effects of the interaction price in treatment **PI**. It breaks down Table 3 into different messages and confirms the pattern observed in Table 3: In general, an increase in the interaction price decreases the **UR** and increases the **HR**, while it decreases the **OR**. A closer look reveals that only after promise *LB* the interaction price has a significant effect on experts' behavior, but not after any other promise. This is again consistent with our two-type model, suggesting that experts with stronger other regarding preferences anticipate the larger guilt from promise breaking and accordingly make promises that are easier to keep afterwards. This means that they more often send promise *LB* and they tend to keep their promise if consumers signal their trust (by choosing a $b > 0$). However, experts who make promises *VF* or *HO* are less influenced by interaction prices, possibly because they are more likely to be selfish, as predicted by our two-type model.

To briefly summarize the experimental results, we found that (1) experts mainly make promise *LB* as predicted by our model; (2) proper promises increase consumer-friendly behavior; and (3) the higher the interaction price under proper promises the more likely consumers receive the appropriate quality. It also turned out that too demanding promises are less likely to be followed by consumer-friendly behavior, most likely because they are made by subjects who are less willing to keep them.

## 5 Conclusion

In this paper we have investigated how trust (combined with reciprocity) and non-binding promises (combined with a preference for promise keeping) influence experts' behavior in credence goods markets. Such markets are characterized by an informational asymmetry between consumers and experts, which can lead to various forms of fraud, such as undertreatment, overtreatment, or overcharging. Using the parsimonious model of Dulleck and Kerschbamer (2006) as our working horse, we have incorporated trust and guilt from promise breaking into a sequential game with credence goods. We have argued that promises made by experts induce a commitment that affects experts' behavior since experts suffer guilt from breaking their promise. This assumption has been motivated by recent experimental evidence indicating that subjects tend to keep promises at a personal cost, even in one-shot relations, where reputational concerns cannot play a role (e.g., Charness and Dufwenberg, 2006). A novel feature in our model is the interaction of trust and promises. In particular, we have assumed that the experts' feelings of guilt from promise breaking are exacerbated if consumers show a larger degree of trust in them. Contrary to earlier literature on behavior of players in gift-exchange and trust games – where trust is measured by the amount of resources transferred from one party of the transaction to the other – we allow consumers to express their trust by paying a non-negative interaction price at the start of their relationship with an expert. A higher interaction price signals more trust without implying a gift to the transaction partner. The increased trust of the consumer increases the amount of guilt an expert feels if he breaks his promise and thereby leads to consumer-friendlier behavior. In sum, promises and interaction prices can serve as communication devices to establish trust and guilt (if a promise is not kept). Both factors together predict a higher frequency of interaction in credence goods markets and a more efficient provision of such goods, while standard theory based on rationality and payoff maximization would predict no interaction and the minimum level of efficiency.

Our experimental results largely confirm our theoretical predictions. Consistent with our model, we have found that positive interaction prices alone have no significant effects on experts' provision and charging behavior. Similarly, communication alone is insufficient to trigger large effects. However, when experts can make meaningful promises to consumers and consumers can express their trust via (positive) interaction prices then experts' promises become a better predictor of their behavior and they treat consumers in a friendlier and more efficient way. In the latter case (treatment PI) experts also show a tendency to make those promises that are easier to be kept afterwards. Finally, promises that are difficult to keep do not lead to nicer behavior. Thus, it seems that only particular kinds of promises foster trust and cooperation.

Distributional preferences alone cannot account for the effects reported here. Rather, it is the combination of trust with a preference for promise keeping. Our findings suggest that "soft" factors like making promises and expressing one's trust have strong effects on behavior on – and efficiency of – credence goods markets which are otherwise prone to several types of inefficiencies and fraud due to informational asymmetries. Because of these asymmetries, consumer protection agencies typically call for institutional safeguards (like liability and verifiability) to protect consumers from being exploited. Dulleck et al. (2009) have shown that liability is, indeed, very important for efficiency on credence goods markets, while ex-post verifiability of the expert's actions is not. However, they have also discussed why liability may not work properly in many cases. The soft factors examined in this paper – making promises and expressing one's trust – might therefore be considered suitable and cost-effective substitutes for hard institutional rules.

## References


Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Brandts, J. and Holt, C. (1992). An experimental test of equilibrium dominance in signaling games. *American Economic Review*, 82(5):1350–1365.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Dufwenberg, M. (2008). Broken promises: An experiment. University of California at Santa Barbara, Economics Working Paper Series 10-08, Department of Economics, UC Santa Barbara.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Cho, I. and Kreps, D. (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102(2):179–221.

Coleman, J. (1990). *Foundations of social theory.* Belknap Press of Harvard University Press.

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

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

Dulleck, U., Kerschbamer, R., and Sutter, M. (2009). The economics of credence goods: On the role of liability, verifiability, reputation and competition. *American Economic Review*, forthcoming.

Ellingsen, T. and Johannesson, M. (2004). Promises, threats and fairness. *Economic Journal*, 114(495):397–420.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, 7(2-3):235–266.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 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. GWDG Bericht 63, Goettingen, Gesellschaft für wissenschaftliche Datenverarbeitung: 79-93.

Huck, S., Lünser, G., and Tyran, J.-R. (2006). Competition fosters trust. Working paper, University College London.

Huck, S., Lünser, G., and Tyran, J.-R. (2007). Pricing and trust. Working paper, University College London.

Huck, S., Lünser, G., and Tyran, J.-R. (2010). Consumer networks and firm reputation: A first experimental investigation. *Economic Letters*, forthcoming.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23:393–404.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica*, 76(6):1467–1480.

# Figures and Tables

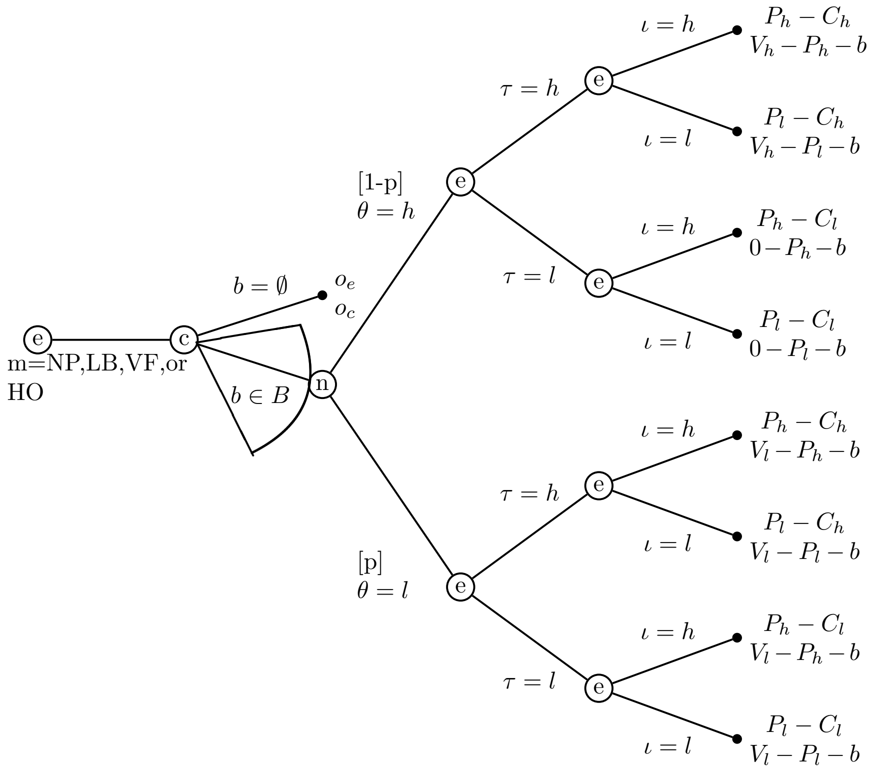


Figure 1: Game Tree of the Extended Game when Only Material Payoffs are Considered ($c$ denotes the consumer, $e$ the expert, and $n$ nature).

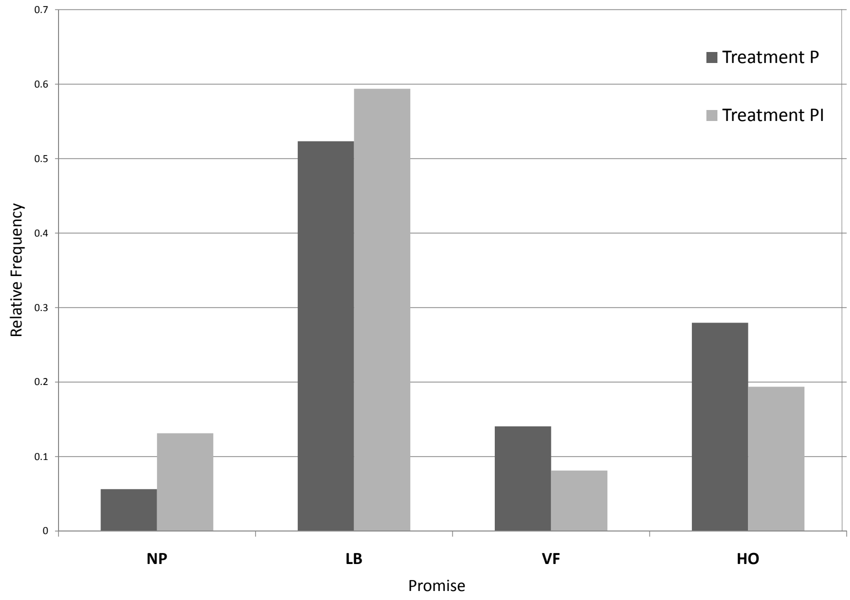


Figure 2: Distribution of Promises in Treatments **P** and **PI**.

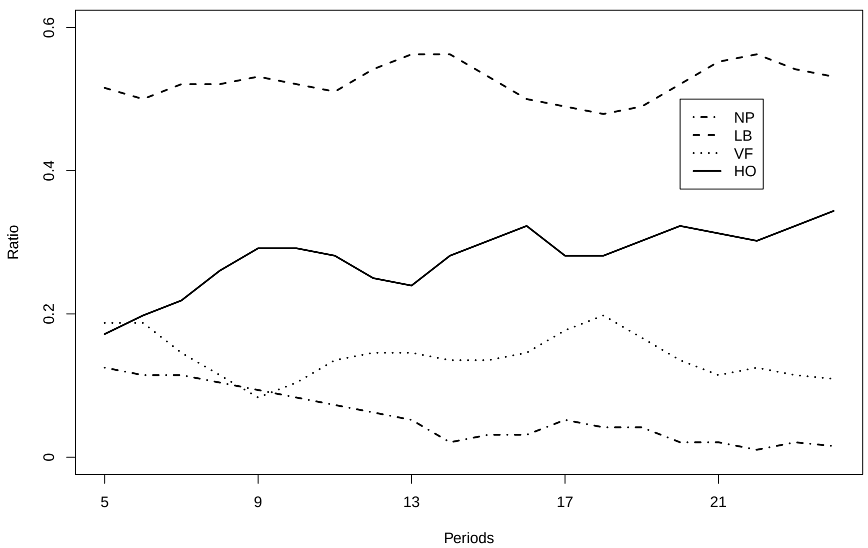


Figure 3: Development of Messages *NP*, *LB*, *VF*, and *HO* over Time in Treatment **P** (smoothed by calculating 3-period moving averages).

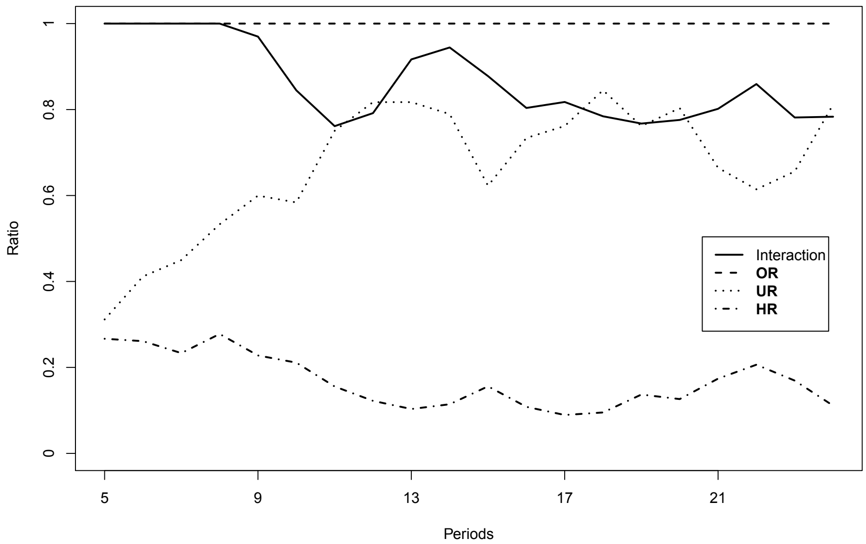


Figure 4: Interaction Ratio, **OR**, **UR**, and **HR** over Time under Promise $HO$ in Treatment **P** (smoothed by calculating 3-period moving averages).

| Treatment | **B** | **P** | | | | |
|---|---|---|---|---|---|---|
| | | Overall | *NP* | *LB* | *VF* | *HO* |
| Interaction | 0.81 | 0.85 | 0.47*** | $0.90_{\diamond\diamond\diamond}$ | $0.82_{\diamond\diamond}$ | $0.85_{\diamond\diamond\diamond}$ |
| **UR** | 0.77 | 0.60 | 1.00*** | $0.49^{*}_{\diamond\diamond\diamond}$ | 0.64 | 0.71 |
| **OR** | 0.96 | 0.99 | 0.94 | 0.99 | 0.97 | 1.00** |
| **HR** | 0.14 | 0.21 | 0.00*** | $0.24_{\diamond\diamond\diamond}$ | 0.20 | 0.16 |

*** / ** / * significantly different from treatment **B** at the 1% / 5% / 10% level
$\diamond\diamond\diamond$ / $\diamond\diamond$ / $\diamond$ significantly different from promise $NP$ at the 1% / 5% / 10% level

Table 1: The Effects of Promises.

| Treatment | **B** | **I** | | |
|---|---|---|---|---|
| Int. Price | NA | all | 0 | $\geq 5$ |
| **UR** | 0.77 | 0.82 | 0.83 | 0.78 |
| **OR** | 0.96 | 0.97 | 0.97 | 0.97 |
| **HR** | 0.14 | 0.10 | 0.09 | 0.12 |

*** / ** / * significantly different at the 1% / 5% / 10% level

Table 2: The Effects of Interaction Prices in Treatment **I**.

| Treatment | **B** | **P** | **I** | | | **PI** | | |
|---|---|---|---|---|---|---|---|---|
| Int. Price | NA | NA | all | $b = 0$ | $b \geq 5$ | all | b=0 | $b \geq 5$ |
| **UR** | 0.77 | 0.60 | 0.82 | 0.83 | 0.78 | 0.57 | 0.64 | 0.44** |
| **OR** | 0.96 | 0.99 | 0.97 | 0.97 | 0.97 | 0.94 | 0.96 | 0.88 |
| **HR** | 0.14 | 0.21 | 0.10 | 0.09 | 0.12 | 0.26 | 0.20 | 0.38** |

*** / ** / * significantly different at the 1% / 5% / 10% level

Table 3: The Effects of Interaction Prices: Comparing $b = 0$ and $b \geq 5$ Within a Given Treatment.

| | Int. Price | **NP** | **LB** | **VF** | **HO** |
|---|---|---|---|---|---|
| Frequency | | 0.13 | 0.59 | 0.08 | 0.19 |
| Interaction | | 0.71 | 0.87 | 0.83 | 0.88 |
| Mean Int. Price | | 4.83 | 2.54 | 4.42 | 3.07 |
| | | | | | |
| **UR** | $b = 0$ | 0.81 | 0.61 | 0.86 | 0.57 |
| | $b \geq 5$ | 0.43 | 0.38* | 0.88 | 0.45 |
| | | | | | |
| **OR** | $b = 0$ | 1.00 | 0.97 | 0.85 | 0.95 |
| | $b \geq 5$ | 0.93 | 0.90 | 0.85 | 0.82 |
| | | | | | |
| **HR** | $b = 0$ | 0.08 | 0.21 | 0.17 | 0.24 |
| | $b \geq 5$ | 0.39 | 0.39* | 0.14 | 0.45 |

*** / ** / * significantly different at the 1% / 5% / 10% level

Table 4: Experts' Behavior Depending on Promises and Interaction Prices in Treatment **PI**: Comparing $b = 0$ and $b \geq 5$ Within a Given Treatment and Promise.

| $m$ | $PV$ | $\pi_c^{prom}(m,h)$ | $\pi_c^{prom}(m,l)$ |
|---|---|---|---|
| $NP$ | all | $-P_h$ | $V_l - P_h$ |
| $LB$ | all | $V_h - P_h$ | $V_l - P_h$ |
| $VF$ | $OT$ | $V_h - P_h$ | $V_l - P_h$ |
| $VF$ | $EM$ | $V_h - P_h$ | $V_l - P_l$ |
| $VF$ | $UT$ | $-P_l$ | $V_l - P_l$ |
| $HO$ | all | $V_h - P_h$ | $V_l - P_l$ |

Table A.1: Consumer's Payoff Implied by Expert's Promise.

| $PV = UT$ | $PV = EM$ | $PV = OT$ |
|---|---|---|
| $NP^* = NP$ | $NP^* = NP$ | $NP^* = NP$ |
| $LB^* = LB$ | $LB^* = LB$ | $LB^* = LB \vee VF$ |
| $VF^* = VF$ | | |
| $HO^* = HO$ | $HO^* = HO \vee VF$ | $HO^* = HO$ |

Table A.2: Redefined Promises.

# Appendix

Appendix can be made available at the journal website.

## Appendix A: Selfish ($S$) and Honest ($H$) Experts

Assume that there are two types $t$ of expert: selfish ones ($S$-types), who are only interested in their own material payoff, and honest ones ($H$-types), who have preferences as specified in the paper. Let $q$ denote the commonly known probability that the expert is honest, i.e., $q = Prob(t = H)$.

The solution concept we use is Perfect Bayesian Equilibrium (PBE; we use PBEs as abbreviation for the plural of PBE).[21] Throughout, our focus will be on symmetric PBEs (in which all experts of a given type behave the same way). Also, we will restrict attention to PBEs in which consumers play undominated strategies (see Remark A below). As will become clear below, both restrictions are without loss of generality. Finally, we use the following tie-breaking rule (which is assumed to be common knowledge): If an $H$-type is indifferent between two or more strategies he decides for the one which is best for the consumer. Since $b$ is a continuous variable this latter assumption is essential for equilibrium existence.[22]

Let $\pi_c^{prom}(m, \theta)$ again denote the payoff the consumer receives if the expert keeps his promise, as defined in Section 2.2. Then:

[Insert Table A.1 about here.]

Given the definition of $\pi_c^{prom}(m, \theta)$ and the definition of guilt (Definition 3) it

[21] Note that in this simple model the notion of PBE is equivalent to the notion of Sequential Equilibrium.

[22] Without this assumption Observation 3 below is only true for $b > b^*$ which results in the usual "nonexistence-of-a-smallest-$\epsilon > 0$" problem. In other words, in any PBE with positive acceptance rate this assumption is satisfied (even if it is not imposed).

is convenient to classify promises as follows:

[Insert Table A.2 about here.]

That is, for undertreatment price vectors the modified definition of promises is as the original one. For equal markup price vectors, $VF$ implies the same payoff promise as $HO$, so both are subsumed under the heading $HO^*$; and for overtreatment price vectors, the verifiability promise is equivalent to the liability promise, so both are subsumed under the heading $LB^*$.

**Observation 1.** *A promise to charge $P_l$ is never kept by any expert in PBE.*

*Proof.* For an $S$-type the claim is obviously true. For an $H$-type it follows from comparing the monetary gain from charging $P_h$ instead of $P_l$, $P_h - P_l$, with the additional guilt from breaking the promise, which is $(P_h - P_l)\psi(b)$. The former exceeds the latter for any $b$ such that $\psi(b) < 1 \Leftrightarrow b < \bar{\pi}_c^{max} - o_c$. But, if this latter inequality is violated the consumer gets a higher payoff by opting out. □

**Observation 2.** *Promise $VF^*$ yields the same behavior as $NP^*$, but the former yields more guilt for $H$-types. Similarily, $HO^*$ yields the same behavior as $LB^*$, but the former yields more guilt for $H$-types.*

*Proof.* Implied by the definition of promises and by Observation 1. □

**Observation 3.** *Promises $LB^*$ and $HO^*$ induce the $H$-type to provide $C_h$ in state $h$ iff $b \geq b^*$, where $b^* = \frac{C_h - C_l}{V_h}[(1-p)V_h + p(P_h - P_l)] + pV_l - P_h - o_c$.*

*Proof.* By undertreating an $H$-type gets an additional monetary payoff of $\Delta_1 = C_h - C_l$. Furthermore, under $LB^*$ and $HO^*$ undertreatment yields additional guilt of $\Delta_2 = \frac{V_h[o_c+b-pV_l+P_h]}{(1-p)V_h+p(P_h-P_l)}$. Thus, an $H$-type refrains from undertreating under $LB^*$ and $HO^*$ iff $\Delta_1 \leq \Delta_2$, which gives the condition above. □

**Remark A.** Observations 1 and 3 and the shape of consumer's payoff function together imply that, for any given promise $m$, any $b(m) \in [0; b^*)$ and any $b(m) > b^*$ is dominated either by $b(m) = b^*$ or by $b(m) = \emptyset$. Thus, for any given promise $m$ and any belief $\mu(m)$ at $m$ (here represented by the probability consumers assign to the event that promise $m$ comes from an $H$-type), consumers' PBE interaction price must be either $b^*$ (if $\mu(m)[V_h - p(V_h - V_l) - P_h - b^*] + [1 - \mu(m)][pV_l - P_h - b^*] \equiv \zeta > o_c$), or $\emptyset$ (if $\zeta < o_c$), or (degenerate or non-degenerate) randomization between $b^*$ and $\emptyset$ (if $\zeta = o_c$). We refer to the probability with which consumers choose $b^*$ at $m$ as the acceptance rate (or the probability of acceptance) of $m$.

**Observation 4.** *There is no PBE in which either $NP^*$ or $VF^*$ is accepted with positive probability.*

*Proof.* The proof follows immediately from the fact that both types of expert always provide $C_l$ and charge $P_h$ under these promises, and that consumers' utility given this behavior is $pV_l - P_h - b$, which is less than $o_c$ for all $b \geq 0$ by assumption. □

**Observation 5.** *In any PBE with positive acceptance rate, $S$-types always make the promise that implies the highest acceptance rate.*

*Proof.* The result follows from the facts that (i) the optimal behavior of an $S$-type is $\tau = l$ and $\iota = h$ $\forall (m, PV, \theta)$; (ii) the payoff of an $S$-type under this behavior is $P_h - C_l$; and (iii) $P_h - C_l > 0 \;\forall\; PV$. Hence, an $S$-type's expected payoff under each price vector increases in the acceptance rate. □

**Observation 6.** *There is no PBE with positive acceptance rate in which $S$-types make either promise $NP^*$ or $VF^*$.*

*Proof.* Follows from Observations 4 and 5. □

**Observation 7.** *There is no (pure- or mixed-strategy, separating or pooling) PBE with positive acceptance rate in which an $S$-type perfectly reveals his type (by making a promise which is never made by an $H$-type).*

*Proof.* The result follows from the fact that a promise that comes from an $S$-type for sure is rejected by consumers with probability 1, and from Observation 5. □

**Observation 8.** *In any PBE in which both $LB^*$ and $HO^*$ are accepted with positive probability, the acceptance rate of $HO^*$ must be strictly higher than that of $LB^*$.*

*Proof.* For the acceptance probability to be strictly positive for both promises, $H$-types must use a strictly mixed strategy. Mixing requires that $H$-types are indifferent between $LB^*$ and $HO^*$. Since both promises lead to the same behavior and to the same per-consumer monetary payoff for the expert, but $HO^*$ yields more guilt (see Observation 2), $HO^*$ must be accepted with higher probability to yield the same expected utility as $LB^*$ for the $H$-types. □

**Observation 9.** *In any PBE with positive acceptance rate we must have* $q \geq q^*$*, where* $q^* = \frac{C_h - C_l}{V_h}[1 + \frac{p(P_h - P_l)}{(1-p)V_h}]$.

*Proof.* Consider a PBE with positive acceptance rate and let $m^*$ be the promise that has the highest acceptance rate. By Observation 4 $m^* \in \{LB^*, HO^*\}$, and by Observation 8 $m^*$ is unique. By Observation 5 all $S$-types choose $m^*$ for sure. Let $\alpha \in [0,1]$ be the probability with which $H$-types make promise $m^*$. For consumers to be willing to pay $b^*$ (which is necessary to induce $H$-types to provide $C_h$ in state $h$, which in turn is necessary for $m^*$ to have a positive acceptance rate) we must have $\alpha q[pV_l + (1-p)V_h - P_h - b^*] + (1-q)(pV_l - P_h - b^*) \geq (\alpha q + 1 - q)o_c$. Since $pV_l + (1-p)V_h - P_h - b^* > pV_l - P_h - b^*$, this condition is easier to satisfy for higher $\alpha$. Setting $\alpha$ equal to 1, plugging in $b^*$, and solving for $q$ yields the condition in Observation 9. □

**Observation 10.** *There is no PBE in which both* $LB^*$ *and* $HO^*$ *are made with positive probability.*

*Proof.* Suppose there is such a PBE. Then, by Observation 8, promise $LB^*$ must be accepted with a probability $\in (0,1)$. In other words, consumers must randomize between $b = b^*$ and $b = \emptyset$ when seeing $LB^*$, implying that they must be indifferent between those actions under this promise. By Observations 5 and 8, $LB^*$ must come from an $H$-type for sure. But randomization of consumers under $LB^*$, where they know that this promise comes from an $H$-type for sure, is inconsistent with paying $b = b^*$ (with strictly positive probability) under promise $HO^*$ (which is necessary for $HO^*$ to be chosen with positive probability), where they know that the fraction of $S$-types making this promise is at least $1 - q$. □

We are now in the position to characterize the PBEs of this game. In Proposition 1 reference is made to the variable $\mu(m)$. This variable denotes the probability consumers assign to the event that promise $m$ comes from an $H$-type.

**Proposition 1.** *If $q < q^* = \frac{C_h - C_l}{V_h}[1 + \frac{p(P_h - P_l)}{(1-p)V_h}]$, then there is no PBE with positive acceptance rate. If $q \geq q^*$, then there exist two types of pooling PBEs (and no other types of PBE) with positive acceptance rate:*

- *Pooling PBEs in which*
  - *both types of expert make promise $LB^*$;*
  - *consumers pay $b^*$ with probability 1 when observing $LB^*$, pay $b^*$ or $\emptyset$ (or mix between $b^*$ and $\emptyset$) when observing $HO^*$, and do not interact ($b = \emptyset$) when observing $NP^*$ or $VF^*$;*
  - *beliefs are $\mu(LB^*) = q$, while $\mu(NP^*)$, $\mu(VF^*)$, and $\mu(HO^*)$ are arbitrary.*
- *Pooling PBEs in which*
  - *both types of expert make promise $HO^*$;*
  - *consumers pay $b^*$ with probability 1 when observing $HO^*$, and do not interact ($b = \emptyset$) when observing any $m \neq HO^*$;*
  - *beliefs are $\mu(HO^*) = q$ and $\mu(LB^*) < q^*$, while $\mu(NP^*)$ and $\mu(VF^*)$ are arbitrary.*

*In both kinds of pooling PBEs $H$-types provide $C_\theta$ in state $\theta$ and charge $P_h$ in any state along the equilibrium path, while $S$-types always provide $C_l$ and charge $P_h$.*

*Proof.* The first sentence follows immediately from Observation 9. Observation 4 implies that only $LB^*$ and $HO^*$ can be accepted with positive probability along an equilibrium path. Observation 10 states that there is no PBE in which both $LB^*$ and $HO^*$ are accepted with positive probability. So there can only be

PBEs in which either $LB^*$ or $HO^*$ is accepted with positive probability. In each case all $S$-types have to make that promise (by Observation 5). Furthermore, all $H$-types have to make the promise since $P_h - C_h > 0$ and $P_l - C_l > 0$ under each price vector and since since $P_h - P_l > (P_h - P_l)\psi(b)$ by Observation 1, so that the payoff of an $H$-type in each state is strictly positive even under $HO^*$. In both types of pooling PBEs $\mu(NP^*)$ and $\mu(VF^*)$ are arbitrary since consumers will not interact (i.e. $b = \emptyset$) under such a promise, no matter what their beliefs are. In the equilibria in which both types of expert make promise $LB^*$, $\mu(HO^*)$ is arbitrary because no type of expert has a strict incentive to deviate to $HO^*$ for any acceptance rate. In the equilibria in which both types of expert make promise $HO^*$, $\mu(LB^*) \leq q^*$, because with $\mu(LB^*) > q^*$ the promise $LB^*$ is accepted with probability 1, which induces $H$-types to deviate to $LB^*$ to avoid guilt (see Observation 1). It remains to be verified that in both configurations strategies are optimal given beliefs, and that beliefs are consistent with strategies along the equilibrium path. This is easily verified. $\square$

As argued in the paper PBEs in which both types of expert make promise $HO^*$ require unreasonable beliefs. To get rid of them we introduce a refinement which is in the spirit of the Intuitive Criterion (Cho and Kreps, 1987). To introduce the refinement, denote the equilibrium payoff of type $t \in \{S, H\}$ in a PBE in which both types of expert make promise $m$ by $u_e^*(t, m)$. We say that promise $m'$ is *(weakly) equilibrium dominated* for type $t$ in the PBE in which both types of expert make promise $m$ if $u_e^*(t, m)$ is (weakly) higher than the maximal payoff type $t$ could get by deviating to promise $m'$, *regardless* of the beliefs consumers have after this choice. Then we require the following:

**Refinement R1:** *Consider a pooling PBE in which both types of expert make promise $m$ and the out-off-equilibrium-path information set corresponding to promise $m' \neq m$.*

*(a) If $m'$ is equilibrium dominated for type $t \in \{S, H\}$, then - if possible - consumers' beliefs at $m'$ should place zero probability on type $t$.*

*(b) If (a) has no bite, then replace "equilibrium dominated" by "weakly equilibrium dominated".*

Note that part (a) of refinement R1 is almost equivalent to Cho and Kreps's Intuitive Criterion. Part (b) strengthens the Intuitive Criterion further and yields the following result:

**Proposition 2.** *The only PBEs with positive acceptance rate that survive refinement R1 are those where (i) both types of expert make promise $LB^*$; (ii) consumers pay $b^*$ with probability 1 when observing $LB^*$ and abstain from paying an interaction price under any other promise; and (iii) beliefs are such that $\mu(LB^*) = q$, $\mu(HO^*) = 0$, and $\mu(m)$ is arbitrary for $m \in \{NP^*, VF^*\}$.*

*Proof.* Obvious and therefore omitted. □

**Returning to our original classification of promises**, what does Proposition 2 imply for equilibrium behavior?

For $PV = UT$, Proposition 2 remains true, even if we replace $m^*$ by $m$.

For $PV = EM$, Proposition 2 remains true if we replace $m^*$ by $m$, provided that we require $\mu(VF) = 0$ (because $VF$ is equivalent to $HO$ in this case).

For $PV = OT$, the story is slightly more complicated because $LB$ and $VF$ imply exactly the same payoff promise for the consumer in this case. Thus, there are pooling PBEs in which $LB$ is accepted with probability 1, while $VF$ is rejected; pooling PBEs in which $VF$ is accepted with probability 1, while $LB$ is rejected; and pooling PBEs in which both promises are accepted with probability 1. Note, however, that all those pooling PBEs lead qualitatively to exactly the same behavior along the equilibrium path.

# 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 a 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" 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 A can send a message to Player B. The content of this message is **not binding** for Player A. He can choose one of the following 4 messages:

   - Message 1: "Hello."
   - Message 2: "I promise I will provide a sufficient solution."
   - Message 3: "I promise I will charge the "Price of Solution I" if I choose Solution I, and I will charge the "Price of Solution II" if I choose Solution II."
   - Message 4: "I promise I will provide Solution I and charge the "Price of Solution I" if you have Problem I, and I provide Solution II and charge the "Price of Solution II" if you have Problem II."

3. Player B sees the non-binding promise of Player A and 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 pays an interaction price for playing the game**. No matter how high this interaction price is, Player B will continue to play the next steps of this round. The interaction price (in steps of 5 between 0 and 30) will be told to the matched Player A, but it will not be paid to him. The interaction price Player B pays will be subtracted from his balance.

4. If Player B decides to pay an interaction price, Player A gets to know the interaction price paid by Player B as well as his 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")**
**− interaction price for this round**
**= 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**