Verifiability in Markets for Credence Goods

Uwe Dulleck*, Rudolf Kerschbamer† and Matthias Sutter‡

Very Preliminary Version (February 2009)

Abstract

Theory predicts that efficiency prevails on credence goods markets if customers are able to verify which quality they receive from an expert seller. In a series of experiments with endogenous prices we observe that verifiability fails to result in efficient provision behavior and leads to very similar results as a setting without verifiability. Some sellers always provide appropriate treatment even if own money maximization calls for over- or undertreatment. Overall our endogenous-price-results suggests that both inequality aversion and a taste for efficiency play an important role for experts' provision behavior. We contrast the implications of those two motivations theoretically and discriminate between them empirically using a fixed-price design. We then classify experimental experts according to their provision behavior.

JEL Classifications L15, D82, D40

Keywords Credence Goods, Experts, Verifiability, Inequality Aversion, Preferences for Efficiency

*corresponding author, School of Economics and Finance, Queensland University of Technology, E-Mail uwe.dulleck@qut.edu.au.

†Department of Economics, University of Innsbruck and CEPR, E-Mail rudolf.kerschbamer@uibk.ac.at, Tel. +43-512-507 7400, Fax +43-512-507 2980

‡Department of Public Economics, University of Innsbruck, E-Mail matthias.sutter@uibk.ac.at.

1 Introduction

Credence goods markets suffer from informational asymmetries between expert sellers and customers, because customers are unable to observe the quality they need, whereas expert sellers are able to do so. Depending on informational conditions and prices for different qualities, expert sellers may have monetary incentives to provide unnecessary high quality or insufficient low quality or to charge for a higher quality than provided. Dulleck and Kerschbamer (2006) show that either liability, i.e. experts being forced to provide sufficient treatment by institutional rules, or verifiability, i.e. customers being able to verify which quality they receive, induces experts to choose prices such that efficiency prevails on the market for credence goods.

We conducted experiments with endogenous prices and find that verifiability alone is of little help. In fact, aggregate behavior in markets where verifiability holds but liability is violated is very similar to the behavior in markets where both verifiability and liability are violated. Additionally, the overall performance in both conditions - with or without verifiability - is better than the standard prediction for a market without verifiability, but worse than the standard prediction for a market with verifiability. Also, in both types of markets some experts always provide appropriate treatment even if own-money-maximization calls for under- (or over-) treatment. In many cases prices are often such that the gains of trade are split equally between expert and consumer if the former provides the appropriate quality and charges honestly. Finally, equal mark-up prices – the theoretical solution to the problems of over- and undertreatment – often lead to overtreatment.

Overall our findings from the experiments with endogenous prices suggest that two motives play an important role for experts' behavior: inequality aversion (see Fehr and Schmidt 1999, and Bolton and Ockenfels 2000) and preferences for efficiency (see Charness and Rabin 2002 and Engelmann and Strobel 2004). The inequality aversion predicts that experts' behavior depends not only on the own monetary payoff but also on that of the consumer thereby –for some price vectors– justifying appropriate treatment to reduce inequality when money maximization calls for over- or undertreatment. For other price vectors it calls for under- or overtreatment when own money maximization is consistent with proving the appropriate treatment. Preferences for efficiency predict appropriate treatment unless the benefit from over- or undertreatment is such that it compensates for the utility cost of the associated efficiency loss.

We explore the implications of both motives for experts' provision behavior theoretically and discriminate between them empirically using treatments with exogenously given price vectors. We then classify experimental experts according to their provision behavior. We find that the behavior of a majority of subjects is consistent with one of the two motivations. Specifically, a taste for efficiency seems to play an important role for provision behavior for more than 40% of the experimental subjects, while inequality aversion seems to be a major motivation for about 30% of experts.

Credence goods markets are important markets in modern economies, health and many repair services but also financial consulting services and sales services of technical equipment are examples. Empirical and anecdotal evidence clearly documented the problems with credence goods. Wolinsky (1993, 1995) refers to a survey conducted by the Department of Transportation estimating that more than half of car repairs are unnecessary. Hubbard (1998) shows that car mechanics conduct vehicle inspections differently depending on whether the vehicles are on warranty or not. In a medical context, Emons (1997) cites a Swiss study reporting that the average person's probability of receiving one of seven major surgical interventions is one third above that of a physician or a member of a physician's family, indicating that customers' (presumed) education and information level affects the quality of treatment they receive. He also mentions a study by the Federal Trade Commission that documents the tendency of optometrists to prescribe unnecessary treatment. Iizuka (2007) investigates the Japanese drug prescription market where doctors often both prescribe and dispense drugs and finds that their prescriptions respond to markup differences. Gruber et al. (1999) show that the relative frequency of Cesarean deliveries compared to normal child births reacts to the fee differentials of health insurance programs for both types of treatments. Hughes and Yule (1992) find that the number of cervical cytology treatments is positively correlated with the fee for this treatment. Fuchs (1978) and Jürges (2007) show that a large part of patients' demand for health care services is supply-driven, because physician density, for instance, has a significant positive effect on the number of doctor visits or operations.

Though empirical studies on credence goods markets provide compelling evidence for inefficiencies, they generally suffer from the lack of a controlled variation of factors that influence the efficient provision of credence goods. Either these papers provide evidence that overtreatment is happening, without systematically exploring the conditions leading to it (see, e.g., the case studies mentioned in Wolinsky 1993 and 1995, or Emons 1997). Or they vary only one particular aspect that influences the provision of credence goods – for example the price differential between Cesarean section deliveries and normal child births (Gruber et al. 1999) – without controlling for and varying other important factors. Experiments allow for a much broader variation of important factors under ceteris paribus conditions. The main focus of the present paper is on the effect of the presence or absence of a specific condition, the verifiability of the provided treatment quality.

Verifiability of the provided treatment quality is likely to hold in important credence goods markets, including dental services, automobile and equipment repair and pest control. For more sophisticated repairs, where the customer is usually not physically present during the treatment, verifiability is often secured indirectly through the provision of *ex post* evidence. In the automobile repair market, for instance, it is quite common that broken parts are handed over to the customer to substantiate the claim that replacement, and not only repair, has been performed. Similarly, in the historic car restoration market the type of treatment is usually documented step by step in pictures.

There exists some theoretical literature assuming verifiability (see Dulleck and Kerschbamer 2006 for a survey). Emons (1987 and 1993) studies the incentives of experts to under- or overtreat consumers and finds that whether the market mechanism induces non-fraudulent behavior depends on the amount of information consumers have at hand to infer the experts' incentives to be honest. Dulleck and Kerschbamer (2009) study competition between experts and discounters and show that under some conditions experts will not survive competition by discounters. Alger and Salanié (2006) study a homogeneous consumer model in which the degree of verifiability is a continuous variable. They identify an equilibrium in which experts defraud consumers in order to keep them uninformed, as this deters them from seeking a better price elsewhere.

The remainder of the paper is organized as follows. Section 2 introduces our experimental design with endogenous prices and presents the main evidence we find. In Section 3 we introduce the two competing models, one based on preferences for efficiency and one based on inequality aversion. This section results in predictions that are then used in the Section 4 to empirical discriminate between the explanations using treatments with exogenous prices. In this section we also classify experts according to their provision behavior. Section 5 concludes.

2 Credence Goods: Standard Theory and Experimental Evidence - The Role of Verifiability

We start this section with a description of the implemented game and standard theory's predictions.

2.1 Basic Model and Standard Predictions

Customers (he) are ex ante identical and know that they need a major treatment (t^h) with probability h, and a minor treatment (t^l) with probability $1-h$. Each customer (he) is randomly matched with one seller (she) who sets prices p^h and p^l for the major, respectively minor, treatment (with $p^h \geq p^l$). The seller has costs c^h for the major treatment, and c^l for the minor one (with $c^h > c^l$).

The customer only knows the prices for the different treatments, but not the type of treatment that he needs, when he makes his decision whether or not to interact with the seller. In case of interaction, the seller gets to know which type of treatment the customer needs. Then she provides one of the two treatments and charges one of the two prices. Customers in need of the minor treatment t^l are sufficiently treated in any case (both if the seller chooses t^l and if she chooses t^h). However, if the customer needs the major treatment t^h, then only t^h is sufficient. A sufficient treatment yields a value $v > 0$ for the customer, an insufficient treatment yields a value of zero. If the customer decides against interaction then both the customer and the seller receive an outside option of $o \geq 0$. In case of an interaction, the monetary payoff for the consumer is the value from being treated minus the price to be paid, whereas the seller receives as a monetary payoff the price charged minus the costs of the provided treatment (c^l if t^l has been provided, otherwise c^h). Figure 1 presents this game.

Standard theory predicts that experts, once the customer chooses to interact, always deliver low quality and charge for high quality. This implies a market break down if $hv - c^l < 2o$ because consumers will choose not to interact unless the seller sets the high price below $hv - o$. Furthermore, even if $hv - c^l > 2o$ customers are predicted to be consistently overcharged and more importantly undertreated.

One solution to the dilemma is to introduce verifiability. Verifiability

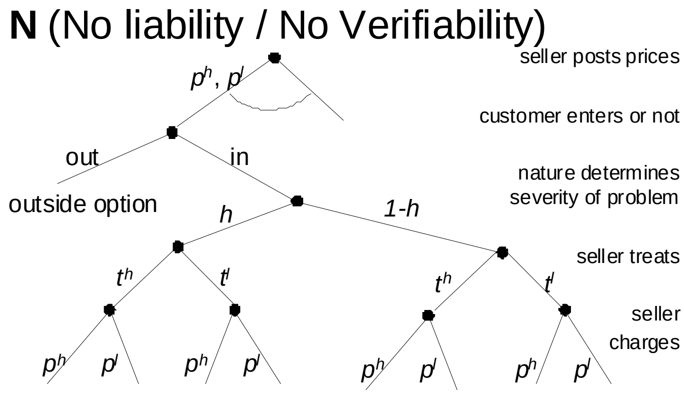

Figure 1: The Credence Goods Game

means that consumers are able to observe and verify ex post the treatment that has been provided by the seller (without knowing, however, whether this treatment is the appropriate one). As a consequence, verifiability prevents overcharging but it does not preclude under- and/or overtreatment.

How does the presence of verifiability change the standard theory's prediction? With verifiability the expert can choose equal mark-up prices that keep her indifferent between providing different treatments. Either by assumption, i.e. that if indifferent the expert will provide in the best interest of the customer, or by referring to the limit of a mixed strategy equilibrium, one can show that such prices will indeed be chosen in equilibrium by experts with standard preferences, leading to efficient provision behavior. For the two quality case, Figure 2 shows in the space of price vectors (p^h, p^l) the set of vectors that induce efficient service. Below this line the expert is induced to always provide low quality (undertreatment in case the consumer needs t^h) and above the line the expert is induced to always provide high quality (overtreatment in case the consumer needs t^l). The line has a slope of one and intersects the vertical axis at $c^h - c^l$ connecting all points where $p^h - c^h = p^l - c^l$.

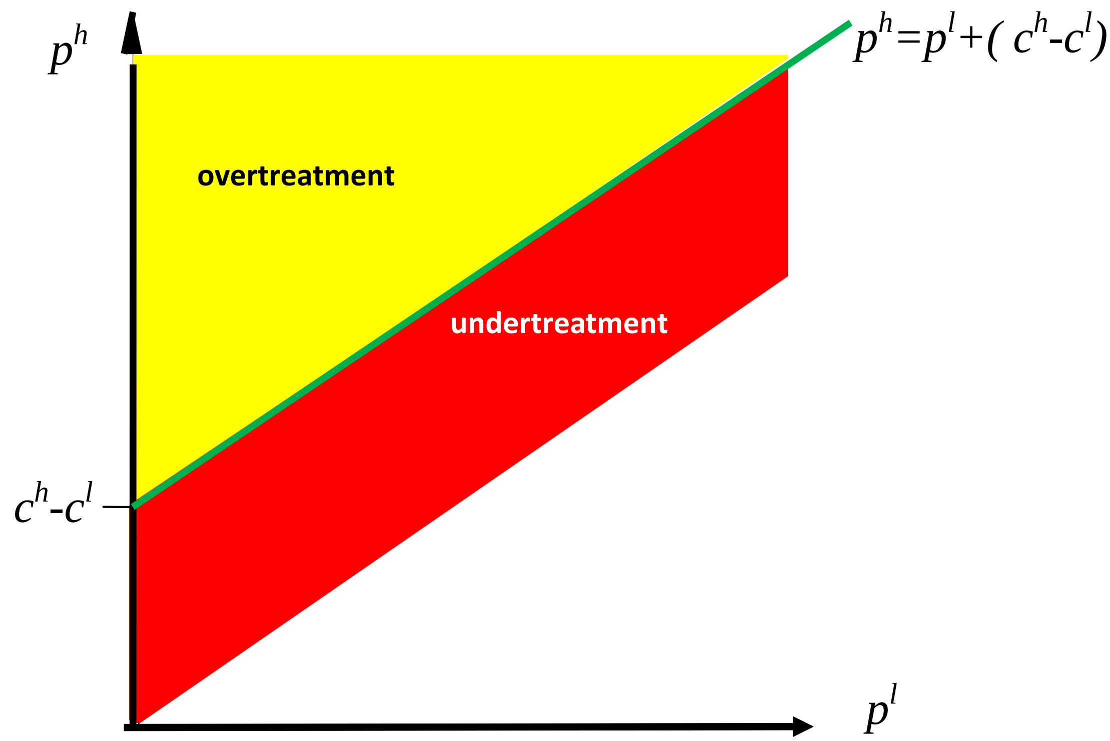

Figure 2: Price vectors that induce efficient provision of credence goods under the assumption of standard preferences.

2.2 Experimental Design

To understand the role of verifiabilty, we compare two experimental conditions.[1] In one we impose no verifiability - condition N - and in the other we impose verifiability - condition V. In these conditions, we follow the standard credence goods game as depicted in Figure 1 with the exception that the last stage is degenerate in condition V. In condition V the expert has to charge the price posted for the provided treatment. In section 4 of this article we investigate an additional condition - condition VFix - where verifiability applies and prices are exogenously set, i.e. experts and customers decide under exogenously given prices.

In all experimental conditions we let the customer's probability of needing the major treatment be $h = 0.5$, and the value of a sufficient treatment be $v = 10$. The costs of providing the minor, respectively major, treatment are $c^l = 2$, and $c^h = 6$. The prices posted by the sellers, p^l and p^h (with $p^l \leq p^h$), have to be chosen in integer numbers from the interval $\{1, 11\}$. The outside option if no trade takes place between the seller and the customer is set to $o = 1.6$.

We always use matching groups of eight subjects each, which is common knowledge in all conditions. Four subjects in each matching group are in the role of customers, and four in the role of sellers. The assignment to roles is randomly determined at the beginning of the experiment, and roles are kept fixed throughout the entire experiment.

There are 16 periods of interaction between sellers and consumers in both conditions. Due to the repetition of the stage game, the matching of subjects is important. Since our main focus is on the role of verifiability on the one hand, and the impact of non-standard preferences on the provision behavior of experts on the other hand, we do not want to give sellers an opportunity to build up a reputation in the course of the repeated interaction. Therefore, we use a stranger matching in which customers and sellers are randomly rematched after each period.

All experimental sessions were run computerized (using zTree; Fischbacher, 2007) and recruiting was done via ORSEE (Greiner, 2004). A total of 248 subjects participated in the experiment, most of them studying economics, business administration, and social and life sciences. All sessions started with an extensive description of the game. All parameters as well as the match-

[1]Since in our story sellers proved a "treatment" to consumers we will refer to experimental treatments as conditions throughout the article.

ing procedure were made common knowledge to all participants by reading them aloud. Before the experiment started, participants had to answer a set of control questions correctly to ensure that they had fully understood the instructions. For every session we invited four subjects more than needed in order make sure that we get enough subjects who could answer all questions correctly. Once a number of subjects required to start a session had answered all questions correctly, the four remaining subjects were paid 4 Euro and dismissed. The average session length was 1.5 hours, and subjects earned on average 15 Euro.

2.3 Standard Theory and the Role of Verifiability

We can solve the game by backward induction. Without verifiablity, the expert will always charge the higher price p^h and always provide the cheaper treatment c^l. Anticipating this consumers will only accept if $p^h \leq hv - o = 3.4$. But with such a price the expert earns less than the outside option $(p^h - c^l < o)$. Thus, without verifiability the market breaks down. With verifiability the expert cannot charge for a treatment other than the provided one and the provided treatment depends on the mark-up $p^i - c^i$, $i \in \{l, h\}$. To characterize the provision behavior of experts it is useful to distinguish the following three types of price vectors:

- an equal mark-up price-vector is defined as one that satisfies $p^h - p^l = c^h - c^l = 4$.
- an undertreatment price-vector satisfies $p^h - p^l < c^h - c^l = 4$.
- an overtreatment price-vector is characterized by $p^h - p^l > c^h - c^l = 4$.

Under equal mark-up price vectors sellers provide the appropriate treatment, while they provide always the minor (major) treatment under the undertreatment (overtreatment) price vectors. Anticipating this, consumers will accept an equal mark-up vector iff $p^h \leq 10$, an undertreatment vector iff $p^l \leq 3$, and an overtreatment vector iff $p^h \leq 8$. Thus, to maximize profits, experts will choose $p^h = 10$ and $p^l = 6$ which will be accepted by a risk neutral consumer. This leads to the following hypothesis.

Prediction 1 *Assume that subjects have standard preferences. Then in condition N no interaction will take place on the experimental credence goods market while in condition V all interactions will be carried out and full efficiency prevails.*

2.4 Aggregate Behavior

Observation 1 (Aggregate Behavior) *Verifiability has no significant impact on the frequency of interaction and the degree of efficiency on the experimental credence goods market. Indeed, the frequency of interaction, the undertreatment rate, the overtreatment rate and overall efficiency are not significantly different in conditions V and N. The overall performance in both conditions is better than the standard prediction for condition N, but worse than the standard prediction for condition V.*

Table 1 below presents some aggregate data. Verifiability does, in no way, help to solve the credence goods problem. Interaction, efficiency, under- and overtreatment rates are not significantly different in condition V compared to condition N. Thus, Prediction 1 can clearly be rejected.

Table 1: Summary statistics for conditions N and V.

Averages per period	Condition N	Condition V
Interaction[a]	0.45	0.50
Efficiency[b]	0.18	0.16
Undertreatment[c]	0.53	0.60
Overtreatment[d]	0.06	0.05
Overcharging[e]	0.86	-
p^l with Interaction	**4.67**	**5.84**
p^l without Interaction	**5.17**	**6.21**
p^h with Interaction	**7.28**	**7.70**
p^h without Interaction	7.91	7.82
Actually charged price	**7.08**	**6.44**
Profits Sellers[f]	2.69	2.58
Profits Customers[f]	1.00	1.06

bold numbers are significantly different 5% Mann-Whitney U-tests for pairwise differences between treatments (with matching groups of 8 subjects as one independent observation)

[a] relative frequency,

[b] calculated as (actual average profit – outside option) / (maximum possible average profit – outside option),

[c] customer needs t^h, but seller provides t^l,[d] customer needs t^l, but seller provides t^h,

[e] seller provides t^l, but charges t^h (with $p^h > p^l$and customer needs t^l),

[f] in experimental currency units.

Before proceeding it is worthwhile to note that standard theory provides not only the equilibrium predictions of no interaction at all in condition N, it also predicts that if interaction takes place in N, the undertreatment rate should be (close to) one! This is obviously not what we observe: The undertreatment is with 53% high, but far from the predicted 100%. Additionally, it is surely not significantly higher in N than in V.

What went wrong in condition V? According to the theoretical prediction sellers should choose equal mark-up prices. However, such prices are very rare in condition V - they are chosen in less than 5% of all transactions. Similarly rare are overtreatment price vectors, i.e. price vectors that provide incentives to overtreat the customer since $p^h - c^h > p^l - c^l$. Most posted price vectors are from the undertreatment variety, ie. $p^l - c^l > p^h - c^h$. Table 2 below reports the frequencies of the five most popular price vectors in conditions N and V.

Table 2: The five most popular price vectors posted in conditions N and V.

Condition N			Condition V		
(p^l, p^h)	absolute #	frequency	(p^l, p^h)	absolute #	frequency
(6,8)	176	22.92%	(6,8)	265	37.64%
(4,8)	84	10.94%	(7,8)	89	12.64%
(5,7)	50	6.51%	(5,8)	46	6.53%
(5,8)	44	5.73%	(4,8)	17	2.41%
(4,7)	39	5.08%	(8,8)	15	2.13%
	393 (of 768)	51.17%		432 (of 704)	61.36%

Notice, in both conditions the price vector (6,8) is by far the most frequent price vector. This price vector splits the gains from trade equally between consumers and sellers - if sellers always provide the appropriate treatment and charge for the provided treatment. The prominence of this price vector suggests that a concern for relative payoffs plays a role for aggregate behavior in the experiment.

There is other evidence that points in the same direction and that also helps to explain the poor performance in condition V: In those rare cases where equal mark-up vectors have been chosen they frequently lead to overtreatment - overall the overtreatment rate under equal mark-up prices is about 1/3! Table 3 contains the five most frequently posted price vectors and shows how the aggregate under- and overtreatment rates change in the price difference $p^h - p^l$.

Table 3: Under- and overtreatment rates in condition V

(p^l, p^h)	Overtreatment rate	Undertreatment rate
(4,8)	37.5%	0%
(5,8)	14.3%	33%
(6,8)	1.25%	53%
(7,8)	0%	65%
(8,8)	0%	100%

The next subsection looks at the individual behavior.

2.5 Individual Behavior

Turning to individual behavior it is interesting to note that in both conditions there exist two types of players that exhibit the same provision policy throughout the game. Let's first look at condition N. Standard theory predicts that all sellers consistently provide c^l independently of the price vector under which they operate and independently of the type of treatment the customer needs. Looking at the data we observe that 14 of the 48 sellers, that is 29% of all sellers, behave exactly in this manner. The number of sellers displaying this behavior increases to 25 (50%) if we only look at the final 10 periods of the experiment. More surprising is, that 12 of the 48 sellers (25%) always choose to provide the appropriate treatment.

The picture in condition V is similar: One group of sellers, consisting of 17 of the 44 sellers, or 39% of sellers in this experiment, consistently posts undertreatment price vectors and always provides c^l independent of the customer's needed treatment. Again, this number increases to almost 50% if we concentrate on the last 10 periods. The second group of experts - consisting of 7 experts (or 16%) always provides the appropriate treatment even though most the time those experts posted undertreatment vectors, too.

Table 4: Number and frequency of subjects with consistent behavior over all 16 periods (in brackets values for the last ten rounds)

	Condition N		Condition V	
Behavior	# of Subjects	Frequency	# of Subjects	Frequency
Own profit maximizing	14 (25)	29(52)%	17 (21)	39(48)%
appropriate treatment provision	12	25%	7	16%

In both conditions it is interesting to note that those experts posting a (6,8) price vector are more likely to provide the appropriate treatment: In condition N, more than 50% of the (6,8) price vectors are posted by experts always providing the appropriate treatment while they make up only 25% of the population. In condition V more than 20% of the (6,8) price vectors are posted by experts always providing the appropriate treatment, while they make up only 15% of the population.

Observation 2 (Individual Behavior) *In both conditions, N and V, there exist two types of players that exhibit the same behavior throughout the 16 periods of the game. The first group consists of 'underproviders'. In condition N its members always provide low quality and always charge for high quality. In condition V this group consists of 40% of the subjects and its members always choose undertreatment price vectors and always provide low quality. The second group consists of 'appropriate providers'. Members of this group always provide the appropriate quality in both conditions. Members of this group have a tendency to choose the price vector* $(6,8)$.

3 Non-Standard Preferences

The results of our experimental investigation suggest that the behavior of a subset of subjects is not only motivated by their own monetary payoff. On the one hand the prominence of price vector (6,8) suggests that inequality aversion is relevant for behavior, on the other hand, the result on undertreatment behavior is consistent with the hypothesis that at least some subjects have a taste for efficiency. In this section we explore the implications of such motives of experts for their provision behavior. To keep the exposition simple, our main focus will be on condition V.

3.1 Preferences for Efficiency

Suppose that (some) sellers derive an extra utility from providing the appropriate treatment or that they feel a moral cost if they under- or overtreat a customer. A preference for efficiency has been postulated, for example, in Charness and Rabin's (2002) influential paper on social preferences, where they assume that a subject's utility does not only depend on its own payoff, but also on the minimum of payoffs in a peer group and the sum of payoffs in this group. The latter argument in a subject's utility function represents a desire for efficiency. Engelmann and Strobel (2004) report that Charness and Rabin's (2002) model is very suitable for describing behavior in simple distributional games. A desire for efficiency will render undertreatment and overtreatment less attractive, as both actions bear a cost for the seller.

We formulate the taste for efficiency of sellers as follows: Let $\theta \in \{l, h\}$ be the index of a customer's type of problem and $\mu \in \{l, h\}$ the index of the treatment provided. Then the utility of a seller of type (α, β) is assumed to be given by

$$U_{\alpha,\beta}(p^l, p^h, \theta, \mu) = p^\kappa - c^\mu - \alpha I_{\theta>\mu} - \beta I_{\theta<\mu}. \tag{1}$$

where $\alpha \geq 0$ is the disutility from undertreating and $\beta \geq 0$ is the disutility from overtreating. I denotes an indicator variable which takes the value of one if the condition in the subscript is met and the value of zero otherwise.[2] Following the literature we assume that buyers and sellers are heterogeneous with respect to their inequality aversion, ie. an agent's type is characterized by values of α and β. Types are independently drawn from the same cumulative distribution $F(\alpha, \beta)$ with strictly positive density on $[0, \alpha^{\max}] \times [0, \beta^{\max})$. Given that we are only interested in sellers' provision behavior, it is only important that α, β are seller specific and do not change in short periods of time.

This utility function has straightforward behavioral implications. First, consider condition N. If $c^h - c^l$, i.e. the additional profit a seller receives when referring to undertreatment, is small compared to the expert's disutility α from undertreating a customer then the expert will use a policy to provide the appropriate treatment. Overtreatment is never optimal for an expert in N. Next consider condition V. The mark-up difference, i.e. $(p^h - c^h) - (p^l - c^l)$, becomes important. Figure 3 shows the areas of undertreatment and overtreatment when verifiability applies. Note, that without verifiability either a seller always undertreats or always chooses the appropriate treatment under each price vector. When verifiability holds, the expert chooses either the appropriate treatment or always provides the treatment with the higher mark-up, depending on the prices posted. Important for our analysis is that a taste for efficiency predicts appropriate treatment in a corridor along the equal mark-up line.

[2] We use the convention that $l \leq h$, but not vice versa.

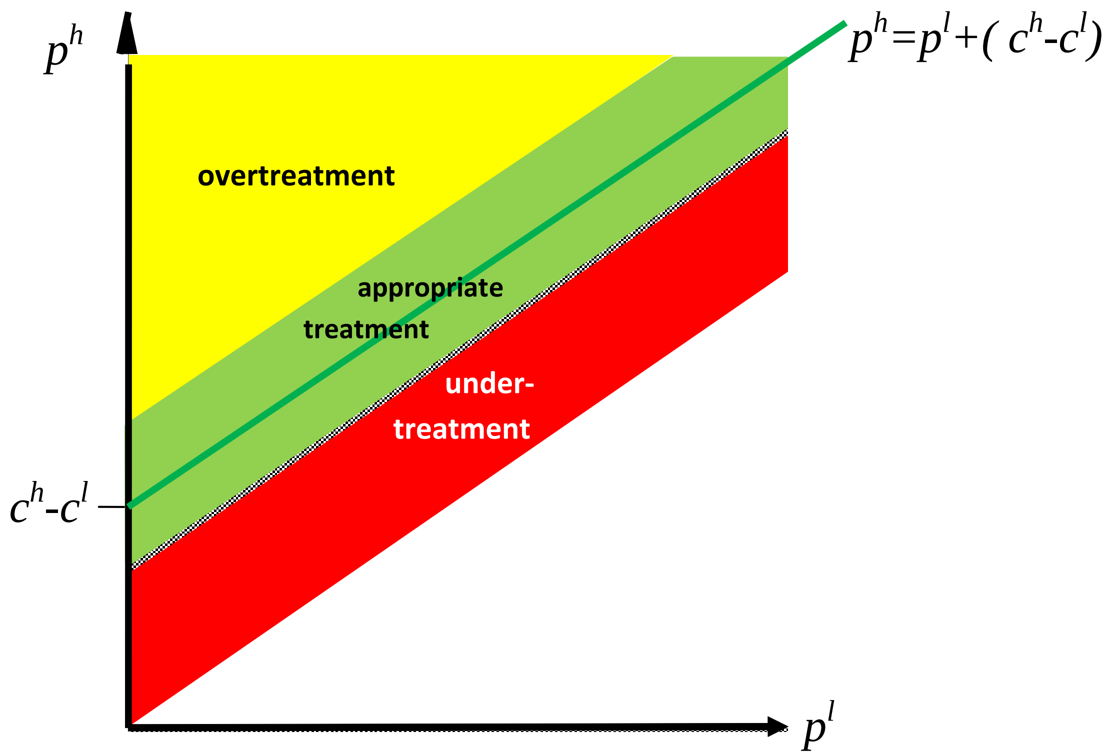

Figure 3: Price vectors that induce efficient provision of credence goods under the assumption of efficiency loving preferences.

3.2 Inequality Averse Preferences

Suppose that (some) sellers derive a disutility from an unequal distribution of the gains from an interaction. Since Fehr and Schmidt's (1999) and Bolton and Ockenfels' (2000) seminal contributions, it has often been argued that subject behavior in lab experiments is heavily affected by inequality aversion. In the following we analyze the consequences of Fehr and Schmidt preferences for the provision behavior on credence goods markets. More specifically, we assume that the utility of a seller of type (a, b) does not only depend depend on her own monetary payoff y_s but also on the payoff of the consumer, y_c.

$$U_{a,b}(y_s, y_c) = y_s - a(\max\{y_c - y_s, 0\}) - b(\max\{y_s - y_c, 0\}). \tag{2}$$

The assumptions from the literature apply: $a \geq b$, i.e. a person suffers more from disadvantageous inequality than from inequality in the person's favor; and $b < 1$, i.e. the seller refrains wasting money to reduce advantageous inequality, because the direct effect on y_s is stronger than the reduced disutility due to a more equal outcome. This is the two-person model of Fehr and Schmidt (1999). Following the observation of Fehr and Schmidt (1999, 2003) we assume that buyers and sellers are heterogeneous with respect to their inequality aversion, ie. an agent's type is characterized by values of a and b. Types are independently drawn from the same cumulative distribution $G(a, b)$ with strictly positive density on $[0, a^{\max}] \times [b^{\min}, 1)$ for all $a \geq b$ and $G(a, b) = 0$ for $a < b$. Given that we are only interested in sellers' provision behavior, it is only important that a, b are seller specific and do not change in short periods of time.

Similar to the previous model, we restrict attention to the case where verifiability holds. Given that inequality aversion is stronger or weaker depending on whether inequality is to the seller's advantage or disadvantage, one first has to determine where these structural breaks take place. For simplicity of presentation we concentrate in what follows on prices with $p^l > \frac{c^l}{2}$. This restriction seems quite natural and it ensures that undertreatment (i.e. providing t^l when t^h is needed) implies an outcome where the monetary payoff of the expert exceeds that of the consumer. To characterize the provision behavior of an expert with inequality averse preferences it is useful to subdivide the space of price vectors in 4 areas depending on the sign of the difference in monetary payoffs of the two trading partners in case of appropriate treatment. In area A, the expert's monetary payoff exceeds that of the customer in both cases, when the customer needs t^h and appropriately receives t^h and when he needs t^l and appropriately receives t^l. This area is defined by $v - p^h < p^h - c^h$ and $v - p^l < p^l - c^l$. In area B the expert is better off if she appropriately provides high quality but worse off if she appropriately provides low quality and low quality is sufficient. This area is defined by $v - p^h < p^h - c^h$ and $v - p^l > p^l - c^l$. In area C the expert is always worse off if she provides the appropriate quality. This area is defined by $v - p^h > p^h - c^h$ and $v - p^l > p^l - c^l$. And finally in area D where the expert is worse off if she provides appropriately high quality but better off if she provides appropriately low quality. This area is defined by $v - p^h > p^h - c^h$ and $v - p^l < p^l - c^l$. Figure 4 displays the four areas, beginning with area A in the upper right quadrant continuing counter clockwise with areas B, C and D, respectively.

Within each area it is straightforward to solve for the provision behavior of the expert depending on what the consumer needs (details available from the authors on request). Figure 4 shows the combinations of prices that induce an expert to provide the appropriate treatment, to always provide high quality (overtreatment if the customer needs t^l), to always provide low quality (undertreatment if the customer need t^h) and to always provide exactly the wrong quality, i.e. provide t^h if t^l is needed and t^l if t^h is needed. For our arguments below it is important to note that the provision behavior of experts with Fehr and Schmidt is qualitatively as depicted in the figure for all values of a and b. The only exception is that for high values of a the *always wrong treatment* region enters area B (implying that the always *appropriate treatment* region disappears from area C).[3]

3.3 Contrasting the Two Models of Non-Standard Preferences

First note that all three models of expert's behavior - the standard model, the taste for efficiency model, and the inequality aversion model - share a monotonicity in the predicted behavior: Consider the standard model first. Holding p^h constant and increasing p^l from some point below $p^h - (c^h - c^l)$ to a price above, the expert is predicted to first overtreat the customer, then, exactly at $p^l = p^h - (c^h - c^l)$, the experts behavior is undetermined (and

[3] Note that the *always wrong* region is neceaasrily to the low left of the always *appropriate treatment* region and that those to region necessarily intersect in exactly one point.

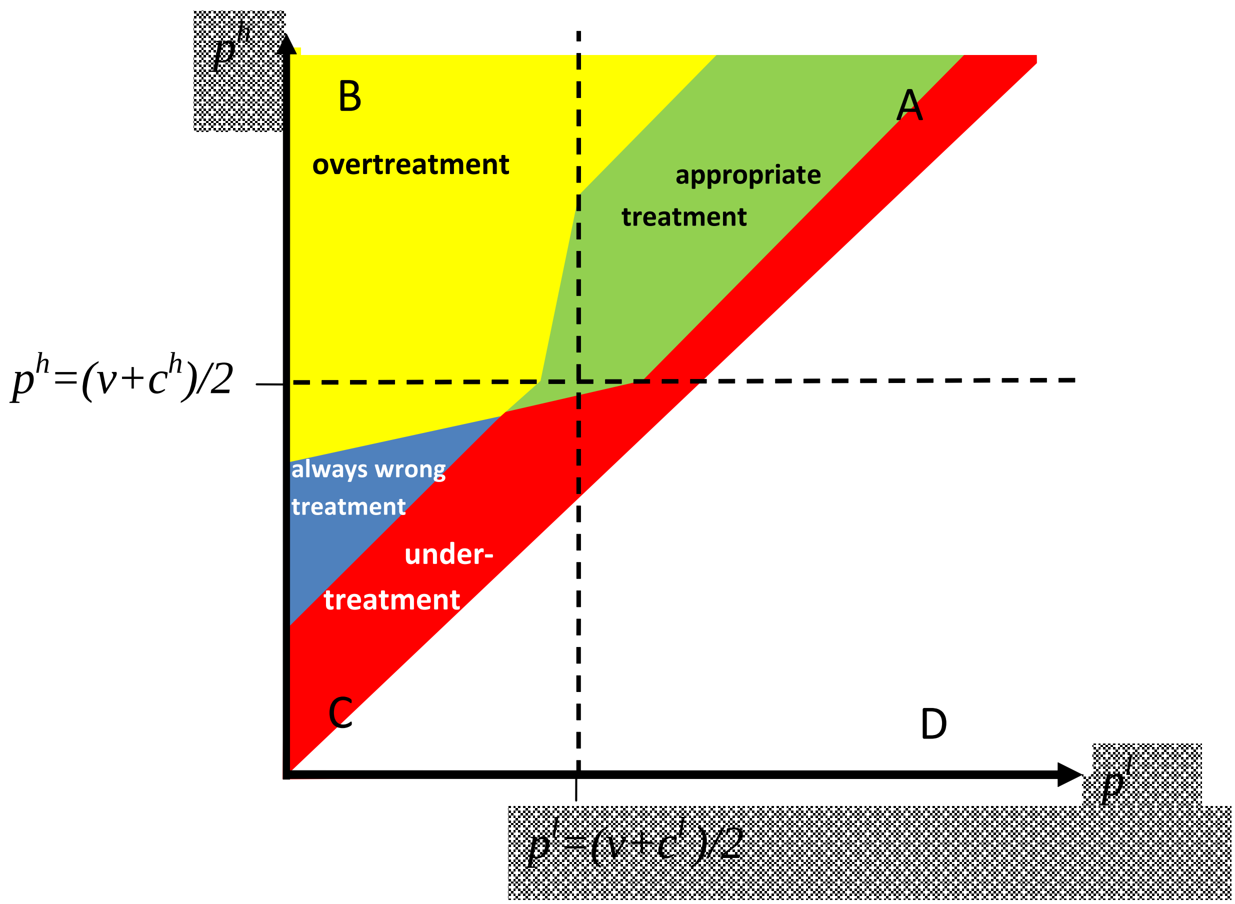

Figure 4: Price vectors such that the expert provides always the appropriate treatment, always t^h (overtreatment), always t^l (undertreatment) and always the wrong treatment ($a = 0.5, b = 0.25$).

usually it is assumed she provides the appropriate treatment), and for higher p^l's ($> p^h - (c^h - c^l)$) the expert is predicted to undertreat the customer. A taste for efficiency implies an even stricter monotonicity, the expert first overtreats, then she provides the appropriate treatment (as a unique prediction), then she undertreats. Additionally, the measure of the region around the equal mark-up prices $p^l = p^h - (c^h - c^l)$ where the expert provides the appropriate treatment is positive. Inequality averse preferences are consistent with two different behavior patterns, first moving from over- to appropriate to undertreatment and second moving from overtreatment to providing always the wrong treatment to undertreatment. See the figures above for an illustration.

Prediction 2 *Behavior of sellers can be explained by one of the three models if, holding p^h constant and varying p^l, leads to one of the following three patterns: a change of behavior from overtreatment to appropriate treatment to undertreatment.*

a) overtreatment - appropriate treatment - undertreatment;

b) overtreatment - always wrong treatment - undertreatment;

c) overtreatment - arbitrary treatment (exactly at $p^l = p^h - (c^h - c^l)$) - undertreatment.

The two alternative models encompass the standard model for $\alpha = \beta = 0$ or $a = b = 0$ respectively. To discriminate, we therefore assume that all parameters are strictly positive. One difference between the two non-standard models is, that equal mark-up contracts always induce efficient provision by subjects with a taste for efficient in the expert's role, but that subjects with inequality averse preferences may refer to overtreatment or always wrong treatment under equal mark-up prices.

The following prediction identifies a price vector that allows to differentiate between the two models.

Prediction 3 *Consider an equal-mark-up price vector with $p^h = (v + c^h)/2$ and $p^l = p^h - (c^h - c^l)$. Under this price vector*

- *appropriate treatment is consistent with preferences for efficiency and standard preferences but inconsistent with inequality averse preferences; furthermore, if an increase in p^l still leads to appropriate treatment then this behavior can only be explained by a taste for efficiency.*

- *overtreatment is consistent with inequality averse preferences and standard preferences but inconsistent with preferences for efficiency; furthermore, if an increase in p^l leads to appropriate provision behavior then this behavior can only be explained by inequality averse preferences.*

- *always providing the wrong treatment is consistent with inequality averse preferences and standard preferences; furthermore, on the one hand if an increase in p^l leaves the behavior unchanged (always provision of wrong treatment) then it is only consistent with inequality averse preferences; on the other hand if an increase in p^l leads to undertreatment then this behavior is consistent with standard preferences and inequality averse preferences.*

- *any other behavior is only consistent with standard preferences; furthermore if an increase in p^l leads to undertreatment, the standard preferences are confirmed.*

4 Testing for the Non-Standard Behavior Predictions

To test the predictions for provision behavior of subjects in the expert role, we conducted an additional set of experiments using a fixed price design with verifiability. The timing of the game is exactly the same as in section 2 except that the first stage of the game was replaced by the experimental software choosing a price vector from the set {(4,8), (5,8), (6,8), (7,8)} with equal probability. This set of vectors has two characteristics, first and foremost, it includes an equal mark-up vector that allows to discriminate between experts with (potential) preferences for efficiency and experts with (potential) inequality averse preferences. The sets other prices increase p^l which allows us to check the change in behavior induced by a change in prices consistent with Prediction 3. Second the set consists of the four most frequently chosen price vectors in condition V. Given that the prices were fixed from the sellers' perspective, we refer to this condition as condition V-Fix.

Similar to the analysis of conditions N and V, we report in the following observations on aggregate and individual behavior for condition V-Fix.

4.1 Aggregate Behavior

To concentrate on the provision behavior, we report the under- and overtreatment rate for the four price vectors in Table 5.

Table 5: Under- and overtreatment rates in condition V-Fix

(p^l, p^h)	Frequency	Undertreatment rate	Overtreatment rate
(4,8)	25%	21.05%	41.38%
(5,8)	25%	68.18%	5.88%
(6,8)	25%	60.61%	2.56%
(7,8)	25%	77.27%	3.85%

Three observations are immediate. First, on the aggregate level there exists evidence that monetary incentives work in the direction standard theory predicts, namely that an increase in the payoff for undertreating customers, i.e. an increase in p^l, increases the undertreatment rate. Second, there is an considerable number of subjects providing the appropriate treatment under *each* price vector. Third, the overtreatment rate is very high under the equal mark-up vector (4,8), indicating that inequality avers preferences play an important role.

In the following, we classify individual consumers into three groups according to Prediction 3.

4.2 Individual Behavior

Prediction 3 provided a classification of types of customers. Table 6 applies this classification to subjects that participated in V-Fix, column 2 uses the data from all 16 periods while column 3 relies only on the data from periods 7 to 16, i.e. the final 10 periods. A subject is classified into one of the groups, if for all rounds under inspection his or her behavior corresponds to the behavior identified in Prediction 3 for one of the 3 types. As already mentioned in prediction 3, some individuals can be classified as either type.

Table 6: Classification of Individual Behavior in V-Fix.[4]

Behavior consistent with	# (all 16 periods)	# (last 10 periods)
Standard Preferences	5	6
Taste for Efficiency	6	7
Inequality Aversion	4	4
Standard and Taste f. Eff.	4	6
Standard and Inequ. Av.	4	5
Total	23 (of 32)	28 (of 32)

The main observation one can get from inspecting Table 6 is that the provision behavior of 28 of the 32 subjects is consistent with one of the three models if one only uses data of the last 10 periods. If all 16 periods are considered, it is still possible to explain the behavior of 23 of the 32 subjects. While this is only superficial evidence that the proposed models of non-standard preferences can explain observed behavior well, it indicates that the basic principles upon which these models are based, carry substantial explanatory value.

Similar to Fehr and Schmidt (1999), this result points to the importance of heterogeneity of subjects in economic experiments. Subjects display consistent behavior over a long period of time (the last ten periods or all periods) and their behavior allows to classify more than 85% into one of the three groups. The behavior is clearly driven by different motives and while studying the aggregate data can give first evidence and provide ideas for further elaboration, analyzing the individual behavior is important to understand the incentives that influence behavior of subjects on experimental credence goods markets.

[4] We classify as: *Standard Preferences* all subjects that provide eratic behavior or undertreat at (4,8) and undertreat for all other price vectors; *Taste for Efficiency* all subjects that provide the appropriate treatment at (4,8) and (5,8) and either appropriately treat or undertreat for theother price vectors; *Inequality Aversion* all subjects that either overtreat or provide the appropriate treatment at (4,8) and provide the appropriate treatment for some other price vector; *Standard and Taste for Efficiency* all subjects that provide appropriate treatment at (4,8) and undertreat for all other price vectors; *Standard and Inequality Aversion* all subjects that overtreat or provide always the wrong treatment at (4,8) and undertreat for any other price vector.

5 Conclusions

This article started by documenting expert's behavior in a credence goods experiment. From this experiment one can learn that while standard economic theory can explain some inefficiencies that prevail on experimental (and real life) credence goods markets, it cannot explain why verifiability - as one market institution that is predicted to solve the credence goods problem - fails to improve efficiency on experimental markets. Furthermore, standard theory cannot explain, why a substantial share of subjects in the expert's role provide appropriate treatments even if it is against their own monetary interest or why experts who are provided with the theoretically optimal monetary incentives refer to providing high quality where low quality is sufficient. Finally standard theory cannot explain the frequency of price vectors that split the gains of trade equally between experts and consumers - if experts provide the appropriate treatment and charge for it. Besides drawing such conclusions from aggregate data based on different experimental conditions, we also looked at individual behavior and found indications that some experts have a taste for efficiency, while other display inequality aversion.

We next analyzed two models of non-standard preferences that allow for a Taste for Efficiency on the one hand and for Inequality Aversion on the other hand. For both models we derived predictions with respect to expert's provision behavior under certain sets of prices. Combining these prediction with the predictions under the assumption of standard preferences allowed us to identify three different types of subjects, namely those with standard preferences, those with Inequality Averse preferences and those with a Taste for Efficiency. This classifications can explain the behavior of more than 90% of the subjects who show consistent behavior over the last 10 periods of the experiment.

Several conclusions result. First and foremost, these experiments show that some people indeed have preferences that are geared towards efficiency. This implies, on the one hand, that problems on credence goods markets are less severe than theory predicts, because some experts do not abuse their power due to moral costs of inefficient behavior. On the other hand, it suggests that institutions should be designed such that individuals with these type of preferences are particularly likely to end up in positions where credence goods problems play a great role. The medical sector seems to be the most important sector in question. An avenue for further research is the use of professionals from the health sector in similar experiments to find out whether these professionals are more or less efficiency loving than the general population.

Second, while standard preferences can qualitatively predict some of the efficiency problems, they are not able to take into account problems that may arise due to prevailing inequality. Such preferences seem to be important in credence goods situations and our predictions show that new types of inefficiencies can result, namely overtreatment and the provision of always the wrong treatment. While in this case our evidence is only preliminary, another avenue for further research would be to test these theories also in field experiments. Schneider (2006) provides a study in this direction using car mechanics.

Third, as already pointed out by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), nonstandard preferences always come with heterogeneity. Thus just investigating aggregate behavior in empirical studies, especially on the effect of incentive systems, may be misleading as individuals react in different ways to the (monetary) incentives provided given their preferences. As indicated above, in a policy context, this indicates that selecting the right group of people for a certain profession may be worthwhile. Apart from this recommendation, it also questions to what extent experimental studies concentrating only on aggregate data can help to inform policy design.

References

[1] Alger, I. and Salanié, F. (2006), A Theory of Fraud and Overtreatment in Experts Markets, *Journal of Economics and Management Strategy* **15**, 853-881.

[2] Akerlof, G. (1970), "The Market for Lemons: Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics* **89**, 488-500.

[3] Bolton, G. E., Ockenfels, A. (2000), ERC – A theory of equity, reciprocity and competition. *American Economic Review* **90**: 166-193.

[4] Charness, G., and Rabin, M. (2002), Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, **117**: 817-869.

[5] Darby, M. and Karni, E. (1973), "Free Competition and the Optimal Amount of Fraud", *Journal of Law and Economics* **16**, 67-88.

[6] Dulleck, U. and Kerschbamer, R. (2003), "Price Discrimination Via The Choice of Distribution Channels", Working Paper 0312, Department of Economics, University of Vienna.

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

[8] Dulleck, U. and Kerschbamer, R. (2009), Experts vs. Discounters: Competition and Market Unraveling and Consumers Do Not Know What They Need, *International Journal of Industrial Organization* **27**, 15-23.

[9] Dulleck, U., Kerschbamer, R. and Sutter, M. (2009), The Economics of Credence Goods: On the Role of Liability, Verifiability, Competition and Reputation, NCER Working Paper, QUT

[10] Emons, W. (1997),"Credence Goods and Fraudulent Experts", *Rand Journal of Economics* **28**, 107-119.

[11] Emons, W. (2001), "Credence Goods Monopolists", *International Journal of Industrial Organization* **19**, 375-389.

[12] Engelmann, D. and Strobel, M. (2004), Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments, *American Economic Review* **94**, 857-869

[13] Fehr, E., Schmidt, K. (1999), A Theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* **114**: 817-868.

[14] Fischbacher, U. (2007), Z-tree: Zürich toolbox for ready-made economic experiments. *Experimental Economics* **10**: 171-178.

[15] Fong, Y. (2005), "When Do Experts Cheat and Whom Do They Target", *Rand Journal of Economics* **28**, 113-130.

[16] Fuchs, V. (1978), "The Supply of Surgeons and the Demand for Operations", *Journal of Human Resources* **13**, 35-56.

[17] Greiner, B. (2004), The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, Working Paper Series in Economics 10.

[18] Gruber, J., Kim, J., and Mayzlin, D. (1999), "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery", *Journal of Health Economics* **18**, 473-490.

[19] Hubbard, T. N. (1998), An empirical examination of moral hazard in the vehicle inspection market. *RAND Journal of Economics* **29**: 406-426.

[20] Hughes, D. and Yule, B. (1992), "The Effect of Per-item Fees on the Behavior of General Practitioners", *Journal of Health Economics* **11**, 413-437.

[21] Iizuka, T. (2007), Experts' agency problems: Evidence from the prescription drug market in Japan. *RAND Journal of Economics* **38**: 844-862.

[22] Jürges, H. (2007), Health insurance status and physician-induced demand for medical services in Germany: new evidence from combined district and individual level data. University of Mannheim. MEA-Discussion Paper 119-07.

[23] Mas-Colell, A., Whinston, M., and Green, J. (1995), "*Microeconomic Theory*", Oxford University Press, New York.

[24] Myerson, R. (1991), "*Game Theory: Analysis of Conflict*", Harvard University Press, Cambridge, Massachusetts.

[25] Nelson, P. (1970), "Information and Consumer Behavior", *Journal of Political Economy* **78,** 311-329.

[26] Park, I.-U. (2005), "Cheap-Talk Referrals of Differentiated Experts in Repeated Relationships", *Rand Journal of Economics* **28**(2)*,* in print.

[27] Pesendorfer, W. and Wolinsky, A. (2003), "Second Options and Price Competition: Inefficiency in the Market for Expert Advice", *Review of Economic Studies* **70**, 417-437.

[28] Pitchik, C. and Schotter, A. (1987), "Honesty in a Model of Strategic Information Transmission", *American Economic Review* **77**, 1032-1036; Errata (1988), *AER* **78**, 1164.

[29] Pitchik, C. and Schotter, A. (1993), "Information Transmission in Regulated Markets", *Canadian Journal of Economics* **26**, 815-829.

[30] Richardson, H. (1999), "The Credence Good Problem and the Organization of Health Care Markets", mimeo, Department of Economics, Texas A&M University.

[31] Schneider, H. (2006), A Field Experiment to Measure Agency Problems in Auto Repair, mimeo, Yale University.

[32] Sülzle, K. and Wambach, A. (2005), "Insurance in a Market for Credence Goods", *The Journal of Risk and Insurance* **72**, 159-176.

[33] Taylor, C. (1995), "The Economics of Breakdowns, Checkups, and Cures", *Journal of Political Economy* **103**, 53-74.

[34] Tirole, J. (1990), *The Theory of Industrial Organization*, MIT Press, Cambridge, Massachusetts.

[35] Wolinsky, A. (1993), "Competition in a Market for Informed Experts' Services", *Rand Journal of Economics* **24**, 380-398.

[36] Wolinsky, A. (1995), "Competition in Markets for Credence Goods", *Journal of Institutional and Theoretical Economics* **151**, 117-131.