American Economic Review 101 (April 2011): 526–555
http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.2.526

The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition

By Uwe Dulleck, Rudolf Kerschbamer, and Matthias Sutter*

Credence goods markets are characterized by asymmetric information between sellers and consumers that may give rise to inefficiencies, such as under- and overtreatment or market breakdown. We study in a large experiment with 936 participants the determinants for efficiency in credence goods markets. While theory predicts that liability or verifiability yield efficiency, we find that liability has a crucial, but verifiability at best a minor, effect. Allowing sellers to build up reputation has little influence, as predicted. Seller competition drives down prices and yields maximal trade, but does not lead to higher efficiency as long as liability is violated. (*JEL* D12, D82)

Repair services, medical treatments, the provision of software programs, or a taxi ride in an unknown city are prime examples of what is known as a credence good in the economics literature. Generally speaking, credence goods have the characteristic that though consumers can observe the utility they derive from the good ex post, they cannot judge whether the type or quality of the good they have received is the ex ante needed one. Moreover, consumers may even ex post be unable to observe which type or quality they actually received. An expert seller, however, is able to identify the type or quality that fits a consumer's needs by performing a diagnosis. He can then provide the right quality and charge for it, or he can exploit the information asymmetry by defrauding the consumer.

Michael R. Darby and Edi Karni (1973) have introduced the term *credence goods* and added this type of good to Phillip Nelson's (1970) classification of ordinary, search and experience goods.[1] It is noteworthy that Darby and Karni (1973) were, in fact, obstetricians for two strands of literature on credence goods. While one strand takes the characteristic mentioned above—*consumers do not know what*

* Dulleck: School of Economics and Finance, Faculty of Business, Queensland University of Technology, GPO Box 2434, Brisbane, Qld 4001 Australia (email: uwe.dulleck@qut.edu.au); Kerschbamer: Department of Economics, University of Innsbruck, Universitaetsstrasse 15, A-6020 Innsbruck, Austria (email: rudolf.kerschbamer@uibk.ac.at); Sutter: Department of Public Finance, University of Innsbruck, Universitaetsstrasse 15, A-6020 Innsbruck, Austria, and Department of Economics, University of Gothenburg, Vasagatan 1, 40530 Gothenburg, Sweden, and IZA Bonn, Schaumburg Lippe Strasse 5-9, D-53113 Bonn, Germany (email: matthias.sutter@uibk.ac.at). We received helpful comments from Adrian Beck, Dennis Dittrich, Winand Emons, Stephan Kroll, Wolfgang Luhan, Larry Samuelson and participants at the Econometric Society Meeting in Wellington, three referees at the ENABLE-Meeting in Mannheim, the third Australian Workshop on Experimental Economics in Melbourne, and seminar participants in Berlin, Bonn, Brisbane, Hannover, Jena, Stavanger, and Vienna. Financial support from the Max Planck Society, the German Science Foundation (through the Gottfried Wilhelm Leibniz Price of the DFG, awarded to Axel Ockenfels) and the Austrian Science Foundation (FWF-grant P20796) is gratefully acknowledged.

[1] Ordinary goods (such as petrol) have well-known characteristics, and subjects know where to get them. Search goods (like clothes) need to be inspected before buying in order to observe their characteristics. Experience goods (like wine) have unknown characteristics, but they are revealed after buying or consuming them.

they need, but they observe the utility from what they get—as the defining feature, the second is based on the definition that credence goods have qualities which are expensive to judge even after purchase. Typical examples mentioned in this second strand of literature are goods vertically differentiated by process attributes (as, e.g., whether food has been produced organically or not, whether tuna has been caught with dolphin-friendly methods or not, or whether electricity has been generated with a low-emissions technology or not), and a typical assumption made is that *consumers know what they want or need, but observe neither what they get nor the utility derived from what they get.*[2] In this article we adhere to the first definition, which is also the definition used by Dulleck and Kerschbamer (2006) in their survey of the credence goods literature.

The information asymmetries prevalent in credence goods markets lead to several types of inefficiencies which capture wide public attention. To define them systematically, consider a car owner bringing his vehicle to a garage for repair. The mechanic—as an expert seller—might have an incentive to cheat on the consumer in two dimensions: First, the repair might be inefficient. The mechanic might replace more parts than are actually necessary to bring the car back on the road (and charge for the additional time and material). This case is referred to as *overtreatment* because the additional benefits to the consumer are smaller than the additional costs. The mechanic's repair might also be insufficient, thus leaving the consumer with a bill, but with a car that is still not working properly. This latter case is referred to as *undertreatment* since any material and time spent on the repair is pure waste. Second, the repair might be appropriate, but the mechanic might charge the consumer for more than he has actually done (e.g., by claiming to have changed a filter without having done so). This kind of problem is known as *overcharging* and it can also lead to inefficiencies in the long run if the fear of getting overcharged deters consumers from trading on credence goods markets in the future, thereby creating an Akerlof-type of market breakdown (George Akerlof 1970).

In this paper we investigate how informational and institutional restrictions and market conditions affect behavior on credence goods markets and how the inefficiencies arising from the asymmetric information on credence goods markets can be contained. Despite the importance of credence goods markets for many day-to-day decisions, this is the first large-scale experiment on credence goods. Our experiment involves 936 participants and is based on a full $2 \times 2 \times 2 \times 2$ factorial design, varying the following factors:

- *Liability*, i.e., the necessity for the seller to provide a good of sufficient quality to meet a consumer's needs (as opposed to an environment where the seller can undertreat).
- *Verifiability* of a seller's action, i.e., the necessity for the seller to charge for the quality provided (as opposed to a setup where the seller can overcharge).
- *Reputation* building, i.e., giving consumers the possibility to identify their trading partners (as opposed to an anonymous market).

[2] Contributions to the second strand of literature include Timothy J. Feddersen and Thomas W. Gilligan (2001) and Soham Baksi and Pinaki Bose (2007). We thank an anonymous referee for asking us to clarify the differences between the two strands of literature.

- *Competition*, i.e., giving consumers an option to choose from several sellers (as opposed to bilateral matching between sellers and consumers).

While under standard assumptions, theoretical analysis predicts that liability or verifiability yield efficiency, our experimental results show that liability has a crucial, but verifiability at best a minor, effect. Allowing sellers to build up reputation has little influence on market efficiency, as predicted. Seller competition drives down prices and yields maximal trade, as expected. However, it does not lead to higher efficiency as long as liability is violated, because sellers frequently provide insufficient quality to consumers in this case. Hence, our results suggest that legal liability clauses are most suitable to cure many of the inefficiencies associated with the provision of credence goods.

The remainder of the paper is organized as follows. In Section I we discuss briefly related literature and how our experiment can add insights and thus complement other empirical evidence. In Section II we describe the basic model and introduce the various conditions under which consumers and sellers might engage in trade. Section III presents the experimental design and the theoretical predictions. Section IV presents the experimental results. Section V summarizes our main findings and discusses some implications for policy as well as for further theoretical analysis of credence goods markets.

I. Related Literature and Rationale for an Experimental Study

Darby and Karni's (1973) seminal paper on credence goods studies how market conditions (the presence or absence of idle capacities, regulation, etc.) and reputation concerns affect the equilibrium amount of fraud, with the latter meaning under- and overtreatment, and overcharging. Our paper focuses precisely on these problems in credence goods markets. Other theoretical contributions have focused, for example, on whether a competitive sampling of opinions makes it attractive for experts to provide costly but unobservable diagnostic effort (Wolfgang Pesendorfer and Asher Wolinsky 2003), or under which conditions sellers have an incentive to lie about the true diagnosis in order to prevent informed customers from asking for the right quality of the good from a competing seller (Ingela Alger and Francois Salanié 2006) or in order to substitute for price discrimination of customers (Yuk-Fai Fong 2005). These theoretical questions will not be dealt with in this paper, as we will abstain from search costs, seller's diagnosis efforts, or consumer heterogeneity in order to keep the model and the experiment as succinct as possible.

The empirical evidence on the problems with credence goods originates mainly from the markets for car repairs and for health care services. Wolinsky (1993, 1995) refers to a survey conducted by the Department of Transportation estimating that more than half of car repairs are unnecessary, which is an indication of overtreatment. Thomas N. Hubbard (1998) shows that car mechanics conduct vehicle inspections differently depending on whether the vehicles are on warranty or not. Referring to the health care sector, David Hughes and Brian Yule (1992) find that the number of cervical cytology treatments is positively correlated with the fee for this treatment. Likewise, Jonathan Gruber and Maria Owings (1996) and Gruber, John Kim, and Dina Mayzlin (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. Toshiaki Iizuka (2007) investigates the Japanese drug prescription market where doctors often not only prescribe but also dispense drugs. Controlling for patients' health status, he finds that doctors' prescriptions respond to markup differences, i.e., to monetary incentives that are unrelated to warranted medication.[3] These studies provide evidence that monetary incentives matter for the provision of credence goods. Winand 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 a consumer's (presumed) education or information level can also affect the quality of treatment and the likelihood of overtreatment.

Though empirical studies on credence goods markets document the existence of inefficiencies, they generally lack a controlled variation of factors that might influence the level of efficiency. For instance, some papers show that overtreatment is happening, without systematically exploring the conditions leading to it (see, e.g., the case studies in Wolinsky 1993, 1995; or Emons 1997). Other studies 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, Kim, and Mayzlin1999)—without controlling for and varying other important factors (like liability or verifiability or reputation building of sellers). Of course, field data are naturally limited in the number of conditions that can be varied. By running a controlled laboratory experiment we are able to systematically vary several factors that may affect the provision of credence goods, and identify the effects of these factors on sellers' and consumers' behavior. Hence, our experiment complements the empirical literature by allowing for a much broader variation of important factors under ceteris-paribus conditions.

Turning to the experimental literature, closest to our paper are two papers by Steffen Huck, Gabriele K. Lünser, and Jean-Robert Tyran (2007, 2008) who use a binary version of the well-known trust game (by Joyce Berg, John W. Dickhaut, and Kevin A. McCabe 1995) and interpret it as modeling a market for experience goods.[4] Huck, Lünser, and Tyran (2008) show that experience goods are more efficiently provided when sellers can build up reputation than if this is not the case. Yet, it does not make a difference whether buyers can only observe how a particular seller has served them in the past or whether they know all past quality choices of all sellers in the market. Introducing competition (as compared to a bilateral matching of sellers and consumers), Huck, Lünser, and Tyran (2007) find even higher efficiency levels than with reputation, because competition lets sellers provide high quality in the present to attract consumers also in the future. But taking a trust game as an example for an experience goods market limits the analysis of inefficiencies to

[3] Christopher C. Afendulis and Daniel Kessler (2007) show that the integration of diagnosis and treatment has raised the treatment costs of coronary artery disease. However, they also find that under some conditions the integration of diagnosis and treatment can lead to better health outcomes.

[4] Experience goods differ from credence goods in several important dimensions. For example, (i) while the valuation of a consumer is strictly increasing in quality with experience goods, it is constant whenever the quality is sufficient with credence goods; (ii) for given prices a consumer can tell exactly which quality he prefers in the case of experience goods, but he does not know it with credence goods; (iii) whereas the quality of the good is unobservable ex ante but perfectly observable ex post with experience goods, it may be observable either ex ante, or ex post, or neither ex ante nor ex post with credence goods.

undertreatment (low quality for a given price) and no market interaction. The framework of credence goods is a much richer one as it adds opportunities for *overtreatment* and *overcharging,* both of which constitute persistent problems on credence goods markets. It is noteworthy, however, that the setup of Huck, Lünser, and Tyran (2007, 2008) is a special case of our more general model. In the final section of this paper we will explain in more detail how the experience goods model of Huck, Lünser, and Tyran (2007, 2008) is embedded in our model and how our results compare to theirs.

II. A Simple Model of a Credence Goods Market

A. *The Basic Setup*

We build our model on the framework presented in Dulleck and Kerschbamer (2006).[5] Consumers are ex ante identical and know that they need a high quality product (q^h) with probability h and a low quality product (q^l) with probability $1 - h$. Each consumer (he) is randomly matched with one seller (she) who sets prices p^h and p^l for the high, respectively low, quality (with $p^h \geq p^l$). The seller has costs c^h for the high quality and c^l for the low quality (with $c^h > c^l$).

The consumer only knows the prices for the different qualities, but not the type of quality that generates the highest surplus (we refer to this as the quality he needs), when he makes his decision whether or not to trade with the seller. If the consumer decides against engaging in trade with the seller then both the consumer and the seller receive an outside option of $o \geq 0$. If trade takes place, the seller gets to know which quality the consumer needs. Then she provides one of the two qualities and charges one of the two prices. Consumers in need of the low quality q^l are sufficiently treated in any case (receiving either q^l or q^h). However, if the consumer needs the high quality q^h, then only q^h is sufficient. A sufficient quality yields a value $v > 0$ for the consumer, an insufficient quality yields a value of zero. In case of an interaction, a consumer earns the value from being served with a particular quality (which is either v or zero) minus the price to be paid, while a seller receives the price charged minus the cost of the provided quality (c^l if q^l has been provided, c^h otherwise).

In the following, we extend this basic setup by considering, first, *liability* and *verifiability* as institutional, respectively informational, restrictions on the seller's action space and, second, *reputation* building and seller *competition* as two important features of market conditions.

[5] In order to keep the exposition as succinct as possible, we present the basic model as it will be implemented in the experiment. Of course, more general specifications would be possible, such as varying the value of the outside option with the type of player (seller versus consumer) or the value from receiving a sufficient quality with the quality needed by the consumer (high versus low quality), or considering strictly positive diagnosis costs (which we set at zero here). Dulleck and Kerschbamer (2006) discuss various generalizations, showing that many of them do not affect the theoretical results qualitatively. Hence, we use the simpler framework, which also prevents us from making the experiment too complicated for participants.

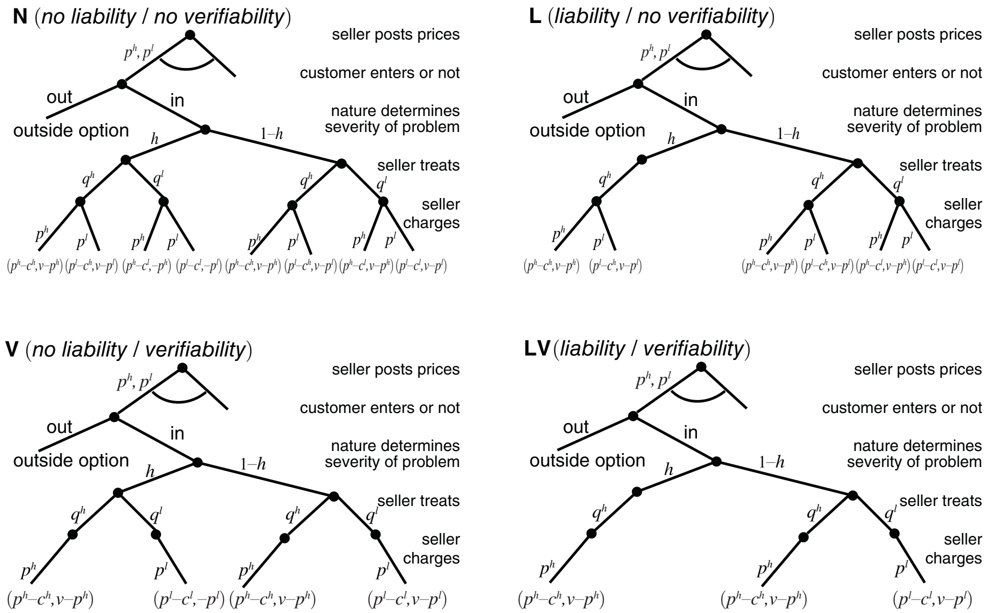

FIGURE 1. THE FOUR INSTITUTIONAL CONDITIONS

B. *Adding Institutional Restrictions: Liability and Verifiability*

Liability in credence goods markets implies the requirement that sellers provide a quality that is sufficient to solve the consumer's problem. Thus, liability prevents undertreatment, but it does not preclude overtreatment and/or overcharging. *Verifiability* means that consumers can observe and verify ex post the quality that has been provided by the seller (without knowing, however, whether this quality was needed). As a consequence, verifiability prevents overcharging, but it does not preclude under- and/or overtreatment.

The factorial combination of liability and verifiability creates four different institutional conditions that imply different sets of available actions for the seller (see Figure 1 for the sequence of actions and the resulting monetary payoffs).

(i) In condition **N** (*no liability*/*no verifiability*) the seller is completely free in her choice of product quality and in which of the two posted prices she charges.

(ii) In condition **L** (*liability*/*no verifiability*) the seller must provide a sufficient quality. However, she is allowed to charge any of her posted prices.

(iii) In condition **V** (*no liability*/*verifiability*) the seller is not restricted in her choice of quality, but she must charge the price of the quality actually provided.

(iv) In condition **LV** (*liability*/*verifiability*) the seller must provide sufficient quality and charge the price of the quality actually provided.

C. *Adding Market Conditions: Reputation and Competition*

The basic setup places two particular restrictions on consumers. First, consumers cannot identify their (past) trading partners. Second, consumers do not have a choice between different potential trading partners, because they are bilaterally matched with one seller only. The first restriction is lifted by considering a *reputation* condition in which sellers are identifiable, such that a consumer can keep track of his past experience with a particular seller (but without knowing how this seller has treated other consumers).[6] The second restriction is removed by considering a *competition* condition where consumers can choose among several sellers, knowing the prices posted by them. In the competition condition the matching becomes endogenous. We assume that consumers face zero costs from comparing the different sellers' prices. Also, there are no capacity constraints, meaning that sellers may serve up to four consumers, while some sellers may serve none. A factorial combination of allowing for reputation building and seller competition yields the following four conditions of market structure:

(i) In condition **B** (for "baseline"; *no competition/no reputation*) the matching of consumers and sellers is random and one-to-one, and consumers cannot identify their trading partners.

(ii) In condition **R** (*no competition/reputation*) consumers can identify their (past) trading partners, but they cannot choose among trading partners because the matching of consumers and sellers is random and one-to-one, as in condition **B**.

(iii) In condition **C** (*competition/no reputation*) consumers can choose among several sellers, but they cannot identify their trading partners.

(iv) In condition **CR** (*competition/reputation*) consumers can choose among different sellers, and they can identify them.

Combining institutional and market conditions yields 16 different conditions for the interaction between consumers and sellers on credence goods markets. Table 1 summarizes the different conditions and how they are characterized with respect to the presence or absence of *liability*, *verifiability*, *reputation*, and *competition*. The abbreviations in Table 1 for the 16 conditions are of the form **X/Y**, where **X** $\in$ {**B, R, C, CR**}, and **Y** $\in$ {**N, L, V, LV**}. For example, **C/LV** denotes the condition with seller competition, where sellers can not build up reputation, but where both liability and verifiability apply. In the following we will often refer to

[6] Huck, Lünser, and Tyran (2008) do not observe differences between conditions where consumers get to know only their own history with a particular seller and where they are informed about a particular seller's past history with all of her consumers. In reality, consumers can always be sure about their own experience, but only rarely observe how sellers have treated other consumers. Hence, we have opted for the information condition where consumers do not know how others were treated. Sellers can never identify consumers in our model, because we are primarily interested in the effect of seller reputation.

TABLE 1—THE 16 CONDITIONS OF INTERACTION

	Market condition			
Institutional condition	**B** (*no competition/ no reputation*)	**R** (*no competition/ reputation*)	**C** (*competition/ no reputation*)	**CR** (*competition/ reputation*)
N (*no liability/no verifiability*)	**B/N**	**R/N**	**C/N**	**CR/N**
L (*liability/no verifiability*)	**B/L**	**R/L**	**C/L**	**CR/L**
V (*no liability/verifiability*)	**B/V**	**R/V**	**C/V**	**CR/V**
LV (*liability/verifiability*)	**B/LV**	**R/LV**	**C/LV**	**CR/LV**

a set of conditions by using a single element of the tuples **X/Y** defined above. For instance, a reference to the set **R** includes **R/N**, **R/L**, **R/V**, and **R/LV**.

III. Experimental Design and Predictions

In this section we present, first, the experimental treatments and parameters. Then we proceed with more details on the experimental procedure and the matching protocol used, and conclude with the theoretical predictions. Our choice of the experimental parameters is motivated by our research interest in the effects of liability, verifiability, reputation and competition. The theoretical literature shows that both verifiability and liability are effective means to guarantee efficient trade in credence goods markets (see Dulleck and Kerschbamer 2006). To test the equivalence of both factors we have determined the experimental parameters such that the maximal trade volume and market efficiency is predicted by standard theory when either liability (in **B/L**) or verifiability (in **B/V**) holds. To see whether the prevalence of liability or verifiability makes a difference on trade in credence goods markets, the parameters imply market breakdown when neither liability nor verifiability applies (in condition **B/N**). This yields the sharpest contrast for predictions. The parameters are also chosen in order to check whether reputation has an effect on trade and market efficiency even when it should not. Finally, in the absence of liability and verifiability, the parameters yield almost maximal trade—but not full market efficiency, when competition applies (in **C/N**). Adding reputation to competition (in **CR/N**) keeps the trade volume almost at the maximum, but also admits equilibria with almost full efficiency. This allows for a crisp test of whether competition combined with reputation can be considered a substitute for verifiability or liability.

A. *Experimental Treatments and Parameters*

The 16 conditions of interaction between consumers and sellers in Table 1 constitute a 2×2×2×2 factorial design with 16 different experimental treatments. In each treatment we let the consumer's probability of needing the high quality be $h = 0.5$, and the value of receiving a sufficient quality be $v = 10$. The costs of providing the low (high) quality is $c^l = 2$ ($c^h = 6$). 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 is set to $o = 1.6$, both for the seller and the consumer.

The stage game is repeated for 16 periods, which makes the matching protocol of subjects important. In all treatments we use matching groups of eight subjects each, which is common knowledge. Four subjects in each matching group are in the role of consumers, 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.

In the treatments *without reputation* (i.e., in sets **B** and **C**) it must not be possible for sellers to build up reputation in the course of the repeated interaction. This precludes the use of a partner matching (in which a seller would be matched with the same consumer in all 16 periods). Therefore, we use a stranger matching in which consumers and sellers are randomly rematched after each period.[7]

In the treatments *with competition* (i.e., in sets **C** and **CR**) the four sellers have to post prices first, and each of the four consumers is informed about the prices of all four sellers. Only then consumers have to choose with which seller, if any, to trade. Note that in set **C** consumers can not identify their (potential) trading partners. In order to make that more transparent, we stress in the experimental instructions (see online supplement S3) that the order of presenting the four sellers' prices on the screen will be randomly determined in each period.

In the treatments *with reputation* (i.e., in sets **R** and **CR**) consumers can keep track of their past experience with a particular seller through fixing the sellers' IDs. The feedback consists of information on which seller they have traded with in a particular period, which prices have been posted in that period, and what has been the consumer's profit from trade.

In all treatments consumers are informed after each period about their payoff. From the latter, consumers can infer whether they have received a good of sufficient quality or not and which price has been charged. Consumers are not informed about which quality was actually needed, although they can infer this information in case of undertreatment. In treatments *without verifiability*, the information feedback is also silent on which quality was provided.

B. *Experimental Procedure*

All experimental sessions were run computerized (using zTree; Urs Fischbacher 2007) and recruiting was done with ORSEE (Ben Greiner 2004). A total of 936 undergraduate students participated in the experiment, with none of them participating in more than one session. In all treatments we ran three sessions, except for treatment **B/N** with four sessions. The number of participants per session was always 16 in treatments of sets **R**, **C**, and **CR**, while the number was 16, 24, or 32 in treatments of set **B**. All sessions started with an extensive description of the game. All parameters as well as the matching procedure were made common knowledge to all participants by reading them aloud. Before the start of an experimental session, participants had to answer a set of control questions correctly to ensure that

[7] Of course, the probability of meeting a particular seller again is one-quarter, meaning that implicit reputation formation that encompasses the whole matching group is feasible. However, the stranger matching precludes reputation formation of *individual* sellers, which is our main concern and distinctive feature in comparison to the treatments with opportunities for reputation building (i.e., to sets **R** and **CR**).

they had fully understood the instructions. For every session we invited four subjects more than needed in order to make sure that we got enough subjects answering all questions correctly. Once the number of subjects required to start a session had answered all questions correctly, the four remaining subjects were paid 4 euros and dismissed. The average session length, including instructions and control questions, was 1.5 hours. Participants received 6 points as starting endowment, with an exchange rate of 4 points for 1 euro. At the end of the experiment they were paid privately and earned on average 14 euros.

C. *Predictions*

In the following analysis we assume that both sellers and consumers are rational, risk neutral and only interested in their own monetary payoff, and that this fact is common knowledge. The equilibrium concept we apply is Perfect Bayesian Equilibrium and our main focus will be on symmetric equilibria. Note that in the treatments *without reputation* (i.e., in sets **B** and **C**) the predictions for the finitely repeated game are the same as the predictions for the underlying stage game. This is not the case in the treatments *with reputation* (i.e., in sets **R** and **CR**) where consumers can keep track of their past experience with a particular seller. The online supplement S1 contains the detailed proofs of our predictions. Here we present only the equilibrium predictions—and a short rationale for it—in order to keep the exposition succinct. Table 2 summarizes all predictions by indicating whether trade takes place, whether undertreatment or overtreatment occurs, and which price vectors are posted and which prices are charged by the sellers in equilibrium.

***PREDICTION B*: *The Effects of Liability and Verifiability* (*Set B*).—**

(*i*) *In* **B**/**N** *the market breaks down. The reason is that in* **B**/**N** *sellers provide the low, but charge for the high quality under each price vector. Anticipating this, consumers would enter the market only if* $p^h \leq 3$. *But with such a low* p^h *even cheating sellers earn less than the value of their outside option* (*because* $p^h - c^l \leq 1 < o = 1.6$).

(*ii*) *In* **B**/**L** *the market generates the maximal trade volume and full efficiency. The argument is that in* **B**/**L** *undertreatment is impossible and overtreatment is dominated by overcharging. Thus, sellers always provide the appropriate quality*[8] *and charge for the high quality under each price vector. Anticipating this, consumers are willing to trade as long as* $p^h \leq 8$. *With* $p^h = 8$ *and expected costs for providing the appropriate quality of* 4 $(= (1-h)\,c^l + h\,c^h)$ *a seller's expected profit is 4, which is larger than the value of her outside option. Thus, sellers post a price vector with* $p^h = 8$, *while* p^l *is indeterminate* (*because it will never be charged when* $p^l < p^h$).

(*iii*) *In* **B**/**V** *the market generates the maximal trade volume and full efficiency. Here the argument is that in* **B**/**V** *sellers provide the appropriate quality under equal markup vectors, but always the low* (*high*) *quality under*

[8] Here and throughout the paper *appropriate quality* refers to providing q^h if the consumer needs the high quality and providing q^l if the consumer needs the low quality; *undertreatment* refers to providing q^l when the consumer needs q^h; *overtreatment* refers to providing q^h when the consumer needs q^l.

TABLE 2—PREDICTIONS ON TRADE AND PRICING-, PROVISION- AND CHARGING-POLICY
(*Symmetric Perfect Bayesian Equilibria*)

Institutional condition		Market conditions: B (*no competition/ no reputation*)	R (*no competition/ reputation*)	C (*competition/ no reputation*)	CR (*competition/ reputation*)
N		**B/N**	**R/N**	**C/N**	**CR/N**
(*no liability / no verifiability*)	[1]	no trade	no trade	mixed eq.: high prob. on {n.d, 3}	mixed as in C/N; or {n.d.,5} in earlier rounds
	[2]	no trade	no trade	undertreatment	undertreatment or efficient provision
	[3]	no trade	no trade	overcharging	overcharging or honest charging
L		**B/L**	**R/L**	**C/L**	**CR/L**
(*liability / no verifiability*)	[1]	{n.d., 8}	{n.d., 8}	mixed eq.: high prob. on {n.d, 5}, low prob. on {n.d., 6}	mixed as in C/L; or {n.d., 5}
	[2]	efficient provision	efficient provision	efficient provision	efficient provision
	[3]	overcharging	overcharging	overcharging	overcharging or honest charging
V		**B/V**	**R/V**	**C/V**	**CR/V**
(*no liability / verifiability*)	[1]	{6, 10}	{6, 10}	mixed eq.: high prob. on {3, 7}, low on {4, 8}	mixed as in C/V; or {3, 7}
	[2]	efficient provision	efficient provision	efficient provision	efficient provision
	[3]	honest charging	honest charging	honest charging	honest charging
LV		**B/LV**	**R/LV**	**C/LV**	**CR/LV**
(*liability / verifiability*)	[1]	{6, 10},{7, 9}, or {8, 8}	{6, 10},{7, 9}, or {8, 8}	mixed eq.: high prob. on {5, 6} or {4, 7}	mixed as in C/LV; or {4, 5} or {3, 6}
	[2]	efficient provision	efficient provision	efficient provision	efficient provision
	[3]	honest charging	honest charging	honest charging	honest charging

Legend

[1]	prediction on interaction and (if applicable) on posted prices
{x, y}	seller posts price x for q^l and price y for q^h.
n.d.	not determined in theoretical solution, but p^l has to satisfy $p^l \leq p^h$.
mixed eq.	equilibrium in mixed strategies
prob.	probability
[2]	prediction on provision policy
undertr.	undertreatment
effic. p.	efficient provision, i.e., providing the quality that is needed
[3]	prediction on charging policy

undertreatment (overtreatment) price vectors.[9] *Anticipating this, consumers are willing to trade under an equal markup vector if and only if* $p^h \leq 10$, *under an undertreatment vector if and only if* $p^l \leq 3$, *and under an overtreatment vector if and only if* $p^h \leq 8$. *Thus, sellers post the equal markup vector {6, 10} which yields the highest profits for them.*

[9] Note the following definition of price vectors: an *equal markup 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 $p^h - p^l > c^h - c^l = 4$.

(*iv*) *In* **B/LV** *the market generates the maximal trade volume and full efficiency. The argument is similar to the one in the previous paragraph, except that equal markup vectors and undertreatment vectors yield the same provision policy and therefore—for given expected prices—the same volume of trade. Thus, sellers post the equal markup vector {6, 10} or an undertreatment vector that yields the same profit in expectation.*

Summarizing Prediction **B** we observe that if both liability and verifiability are violated, then the market breaks down. As soon as either **L** or **V** (or both) apply, however, sellers have an incentive to post prices which induce them to provide the appropriate quality, making it profitable for consumers to enter the market in **B/L**, **B/V**, and **B/LV**. Verifiability and liability are equally effective in inducing full market efficiency, even though the price vectors posted by sellers in equilibrium differ.

***PREDICTION R*: *The Effects of Reputation* (*Set R*)**.—*Reputation itself does not affect the predicted behavior of sellers and consumers, since the stage game has a unique equilibrium and the repeated game a fixed, commonly known end date. Hence, prediction* **B** *also applies to set* **R**.

***PREDICTION C*: *The Effects of Competition* (*Set C*)**[10].—

(*i*) *In* **C/N** *the volume of trade is almost 100 percent although sellers still provide low quality and charge for high quality under each price vector. The possibility of trade even when liability and verifiability are violated in* **C/N** *is a major difference to predictions* **B** *and* **R**, *where market breakdown was predicted under condition* **N**. *The intuition for (almost) maximal trade in* **C/N** *runs as follows. Although sellers can still not be induced to provide the high quality, each seller can now serve more than one consumer. The latter fact implies that there is now room for prices that are profitable for both parties of the interaction. In equilibrium, each seller posts {n.d, 3}*[11] *with probability $x = 0.844$ and a price vector which is unattractive for consumers (due to $p^h > 3$) with probability $1 - x$.*[12] *If at least one seller posts {n.d, 3} then all consumers are (under)treated, otherwise (with probability $(1 - x)^4 = 0.0006$) there is no trade. Note that the increase in the volume of trade (compared to* **B/N**) *translates only into a minor increase in efficiency (less than 1/7 of the potential gains from trade are realized), as consumers are always undertreated in equilibrium.*

(*ii*) *In* **C/L** *the provision and charging policy and the volume of trade are the same as in* **B/L**, *but prices are lower. In equilibrium each seller posts {n.d., 5} with probability $x = 0.839$ and {n.d., 6} with probability $1 - x$. Consumers trade with the seller with the lowest p^h.*

[10] As mentioned earlier our focus is on symmetric equilibria. In the price-posting stage of set **C** there are also asymmetric equilibria. Because there is no obvious way for sellers to coordinate on a specific asymmetric equilibrium we regard such equilibria as less plausible, and thus mention them only here in a footnote. The full proof of Prediction **C** (in online supplement S1) includes the asymmetric equilibria in conditions **C/L**, **C/V**, and **C/LV**.

[11] The abbreviation "n.d." means theoretically not determined. Yet, p^l has to satisfy $p^l \leq p^h$, of course.

[12] We round the probabilities to three decimals here.

(*iii*) *In* **C**/**V** *the provision and charging policy and the volume of trade are the same as in* **B**/**V**, *but prices are lower. In equilibrium each seller posts the equal markup vector* {*3, 7*} *with probability* $x = 0.839$ and {*4, 8*} *with probability* $1 - x$. *A consumer's trading decision does not only depend on both prices,* p^l *and* p^h, *but also on the kind of price vector posted.*

(*iv*) *In* **C**/**LV** *the provision and charging policy and the volume of trade are the same as in* **B**/**LV**, *but prices are lower. In equilibrium each seller posts either* {*4, 5*} *or* {*3, 6*} *with probability* $x = 0.132$, *either* {*5, 5*} *or* {*4, 6*} *or* {*3, 7*} *with probability* $y = 0.280$, *and either* {*5, 6*} *or* {*4, 7*} *with probability* $1 - x - y$. *Within the three sets of price vectors, both sellers and consumers are indifferent to which price vector is accepted if trade takes place.*

In all treatments of set **C** it is important to note that the gains from trade shift completely from the seller's side (as in treatments without competition) to the consumer's side.

***PREDICTION CR*: *The Combined Effects of Reputation and Competition* (*Set CR*).**—*Prediction* **C** *remains an equilibrium prediction also in set* **CR**. *However, in* **CR**/**N** *there are additional equilibria where* (*some*) *sellers post* {*n.d, 5*} (*instead of* {*n.d., 3*}) *in the first nine periods and in which consumers enter the market, because they anticipate* (*correctly*) t*hat they will get the appropriate quality with sufficiently high probability. In these equilibria, adding reputation to competition increases the efficiency of trade* (*in comparison to* **C**/**N**) *because consumers can now costlessly reward a seller who has treated them appropriately in the past, simply by buying from this* (*and not from another*) *seller again, even in the last periods of the experiment where sellers are expected to act opportunistically in any case.*

IV. Experimental Results

In line with the presentation of the model and the predictions, Section IVA deals with the impact of liability and verifiability in set **B**, and Section IVB examines the effects of reputation and competition in sets **R**, **C**, and **CR**. Section IVC presents an econometric estimation of the effects of liability, verifiability, reputation, and competition.

A. *Aggregate Behavior in Set B* (*Baseline*)

***RESULT 1* (*On the Role of Liability and Verifiability*)**: *The market does not break down in* **B**/**N**. *Liability has a significantly positive impact on the frequency of trade and on the degree of efficiency, as theory predicts. However, verifiability has no significant impact on those variables, contrary to the theoretical prediction. In fact, aggregate behavior with respect to trade volume and market efficiency does not differ between* **B**/**N** *and* **B**/**V**, *showing that verifiability is an ineffective means to make credence goods markets more efficient.*

Table 3 presents aggregate data for set **B**. The overall pattern emerging from Table 3 reveals the following results that add further details to our Result 1.

(i) *Trade volume and efficiency*. The first row in Table 3 shows that the average relative frequency of trade between consumers and sellers is only around 50 percent in treatments *without liability* (**B**/**N**, **B**/**V**), but significantly higher and above 80 percent in treatments *with liability* (**B**/**L**, **B**/**LV**).[13] Looking at efficiency in the second row yields a similar picture.[14] Efficiency is below 20 percent without liability, but significantly higher and above 80 percent with liability. Hence, Prediction **B** is basically correct as far as the importance of liability for trade volume and efficiency of credence goods markets is concerned. However, verifiability fails in the experiment.

(ii) *Undertreatment*. The third row of Table 3 reveals that, apart from the low trade volume, the high undertreatment rates (of 53 percent in **B**/**N**, respectively 60 percent in **B**/**V**) are responsible for the low efficiency in **B**/**N** and **B**/**V**. Undertreatment does not differ significantly between **B**/**N** and **B**/**V**. In equilibrium, undertreatment should not occur in **B**/**V** if sellers post equal markup prices. Such prices are posted only very seldom, however, as noted under item (v) below. Rather, the large majority of sellers post undertreatment price vectors in **B**/**V** which provide an incentive for undertreatment. Looking at individual behavior over all 16 periods of the experiment we observe that a considerable fraction of sellers behave persistently in an egoistic way. In treatment **B**/**N** we observe 13 out of 48 sellers (29 percent) who *always* provide the low quality (and charge the high price) across the 16 periods of the experiment. Likewise, 17 out of 48 sellers (39 percent) always provide the low quality in treatment **B**/**V**. Nevertheless, it is interesting and important to note that even in treatments **B**/**N** and **B**/**V** a nonnegligible fraction of sellers (27 percent in **B**/**N**, respectively 16 percent in **B**/**V**) provide the appropriate quality across *all* 16 periods. Hence, a substantial fraction of sellers are honest, thus running counter to the standard prediction of undertreating consumers whenever possible.

(iii) *Overtreatment*. This form of market inefficiency occurs in between 3 percent and 6 percent of observations in set **B** (see fourth row of Table 3) and is therefore no substantial problem in any of the treatments. Note that overtreatment need not be a mistake of the seller in treatments with verifiability, since under overtreatment price vectors providing the high quality (instead of the low quality) increases the seller's profit (by definition). The latter type of price vector is very rare, though, as indicated in item (v) below. Looking at how overtreatment in treatment **B**/**V** changes in the price difference $p^h - p^l$, we observe an inverse relationship (as expected under the assumption that behavior responds to monetary incentives). For instance, in **B**/**V** the overtreatment rate is 0 percent for vectors $\{7, 8\}$ and $\{8, 8\}$, but it increases monotonically with a decrease in the low price, reaching 38 percent for vector $\{4, 8\}$. Note that the price vector $\{4, 8\}$ is an equal markup vector, where standard theory predicts appropriate treatment. Looking only at equal markup vectors (which

[13] In Table 3 (and Table 4 below) we check for significant differences between two treatments each by using two-sided nonparametric Mann-Whitney U-tests (with a matching group of eight subjects as one independent observation).

[14] Efficiency is measured as the ratio of the average actual profit per subject to the average maximally possible profit per subject, where the outside option is disregarded for both measures.

Table 3—Overview of Results in BASELINE (*set B*)

Averages per period	**B/N** (*no liability/ no verifiability*)	**B/L** (*liability/ no verifiability*)	**B/V** (*no liability/ verifiability*)	**B/LV** (*liability/ verifiability*)
Trade[1]	$0.45^{a,c}$	$0.82^{a,d}$	$0.50^{d,f}$	$0.82^{c,f}$
Efficiency[2]	$0.18^{a,c}$	$0.84^{a,d}$	$0.16^{d,f}$	$0.81^{c,f}$
Undertreatment[1, 3]	0.53	–	0.60	–
Overtreatment[1, 4]	0.06	0.02	0.05	0.03
Overcharging[1, 5]	0.88	0.75	–	–
p^l with trade	$4.67^{a,b,c}$	$5.94^{a,e}$	$5.84^{b,f}$	$6.88^{c,e,f}$
p^l without trade	$5.17^{a,b,c}$	$6.12^{a,e}$	$6.21^{b,f}$	$7.46^{c,e,f}$
p^h with trade	$7.28^{a,b,c}$	$8.00^{a,d}$	$7.70^{b,d,f}$	$7.96^{c,f}$
p^h without trade	$7.91^{a,c}$	$8.77^{a,d}$	$7.82^{d,f}$	$8.83^{c,f}$
Actually charged price	$7.08^{a,b,c}$	$7.77^{a,d,e}$	$6.44^{b,d,f}$	$7.46^{c,e,f}$
Profits sellers[6]	$2.69^{a,c}$	$3.44^{a,d,e}$	$2.58^{d,f}$	$3.10^{c,e,f}$
Profits consumers[6]	$1.00^{a,c}$	$2.12^{a,d,e}$	$1.06^{d,f}$	$2.37^{c,e,f}$
Most prominent price vectors	{6,8} 23%	{6,8} 24%	{6,8} 38%	{7,8} 35%
	{4,8} 11%	{7,8} 23%	{7,8} 13%	{8,8} 27%
	{5,7} 7%	{8,8} 13%	{6,7} 11%	{6,8} 17%
Number of subjects	96	96	88	80

[1] relative frequency
[2] calculated as (actual average profit − outside option) / (maximum possible average profit − outside option)
[3] consumer needs q^h, but seller provides q^l.
[4] consumer needs q^l, but seller provides q^h.
[5] seller provides q^l, but charges p^h (with $p^h > p^l$ and consumer needing q^l)
[6] in experimental currency units
Mann-Whitney U-tests for pairwise differences between treatments (with matching groups of 8 subjects as one independent observation)
[a] B/N versus B/L ($p < 0.05$)
[b] B/N versus B/V ($p < 0.05$)
[c] B/N versus B/LV ($p < 0.05$)
[d] B/L versus B/V ($p < 0.05$)
[e] B/L versus B/LV ($p < 0.05$)
[f] B/V versus B/LV ($p < 0.05$)

are posted only in about 4 percent of the cases; see point (v) below) we observe that they all induce a considerable amount of overtreatment in **B/V** (the overtreatment rate under equal markup vectors is 40 percent while the average overtreatment rate in **B/V** is only 5 percent).[15] We will discuss this evidence in the final section of this paper.

(iv) *Overcharging*. Overcharging occurs whenever the low quality is provided, but the high quality is charged for, given that $p^h > p^l$ is satisfied. Overcharging is predicted to occur in 100 percent of cases when sellers post price vectors with $p^l < p^h$ in treatments without verifiability (see prediction **B**). In **B/N**, the overcharging rate is, with 88 percent, actually rather close to the predicted 100 percent, and it is only insignificantly smaller in **B/L**. In both treatments we find that overcharging is increasing in the price difference $p^h - p^l$ which is consistent with the assumption that experts respond to monetary incentives.[16]

[15] For further details on how overtreatment (or undertreatment) depends on the endogenously chosen price vectors see the (extended) working paper version of this paper (Dulleck, Kerschbamer, and Sutter 2009).

[16] Holding the high price constant at $p^h = 8$ in **B/N** (**B/L**) the relative frequency of overcharging is 75 percent (65 percent) when the low price is $p^l = 7$, but increases monotonically to 100 percent (100 percent) with a decrease in the low price down to $p^l = 3$.

(v) *Pricing*. Across all four treatments, the price vector {6, 8} is the most prominent one, capturing about 25 percent of all observations (see bottom part of Table 3). This price vector would split the gains from trade equally between sellers and consumers—*if and only if* sellers always provided the appropriate quality and always charged for the actually provided quality. However, those two conditions are most often not satisfied. The second most popular price-vector is {7, 8} in about 17 percent of the overall observations. Both price vectors are undertreatment vectors, i.e., price vectors that provide incentives for sellers to undertreat consumers in treatments with verifiability. The equal markup vectors predicted in **B**/**V** occur in only about 4 percent of cases. In these treatments we find about 95 percent of price vectors to be of the undertreatment type, which is inconsistent with prediction **B**. Almost no overtreatment price vectors are chosen—in that respect standard theory can explain the low prevalence of overtreatment observed in the experiment. Looking at the price vectors when trade takes place, it seems noteworthy that more than 80 percent of price vectors satisfy $p^h = 8$ in **B**/**L**, which is the point prediction from prediction **B**.

(vi) *Profits*. From the lower half of Table 3 it becomes clear that the average profits for consumers in treatments without liability are *below* the available outside option. Hence, consumers would have benefited from staying away from the credence goods market in these treatments. Only when liability applies, consumers' profits exceed the value of the outside option. As expected, the bilateral matching allocates market power to sellers (by allowing them to make a take-it-or-leave-it offer), and hence they earn significantly more than consumers in each single treatment.

While Table 3 presents overall averages, Figure 2 illustrates the development of key variables across the 16 periods of the experiment. Panel A shows that the relative frequency of trade is rather stable, and high, if liability holds (in **B**/**L** and **B**/**LV**), while it has a steady downward trend whenever liability is violated (in **B**/**N** and **B**/**V**). This downward trend is due to consumers learning that undertreatment happens and that it has large costs for them. While a consumer's relative frequency of interaction with a seller in period t is 57 percent after having received a sufficient quality in period $t-1$, it is only 32 percent after having been undertreated in period $t-1$ ($p < 0.05$, Wilcoxon signed ranks test using data from **B**/**N** and **B**/**V**). Hence, consumers are rather credulous as long as they receive sufficient quality, but react strongly to the experience of undertreatment. Of course, the learning process is rather slow (see the negative, but relatively small estimated coefficient for "period" in the estimations reported in Section IVC), because consumers who happen to need the low quality cannot experience undertreatment and because a considerable fraction of sellers provide the appropriate quality also if the high quality is needed.

Panels B to D of Figure 2 display the time path of undertreatment, overtreatment, and overcharging, showing that overcharging seems to increase over time, while there is no clear time trend for under- and overtreatment. Panels E and F show the development of (accepted) prices p^l and p^h, indicating that consumers are willing (and have) to pay the highest prices in conditions where liability applies. This means

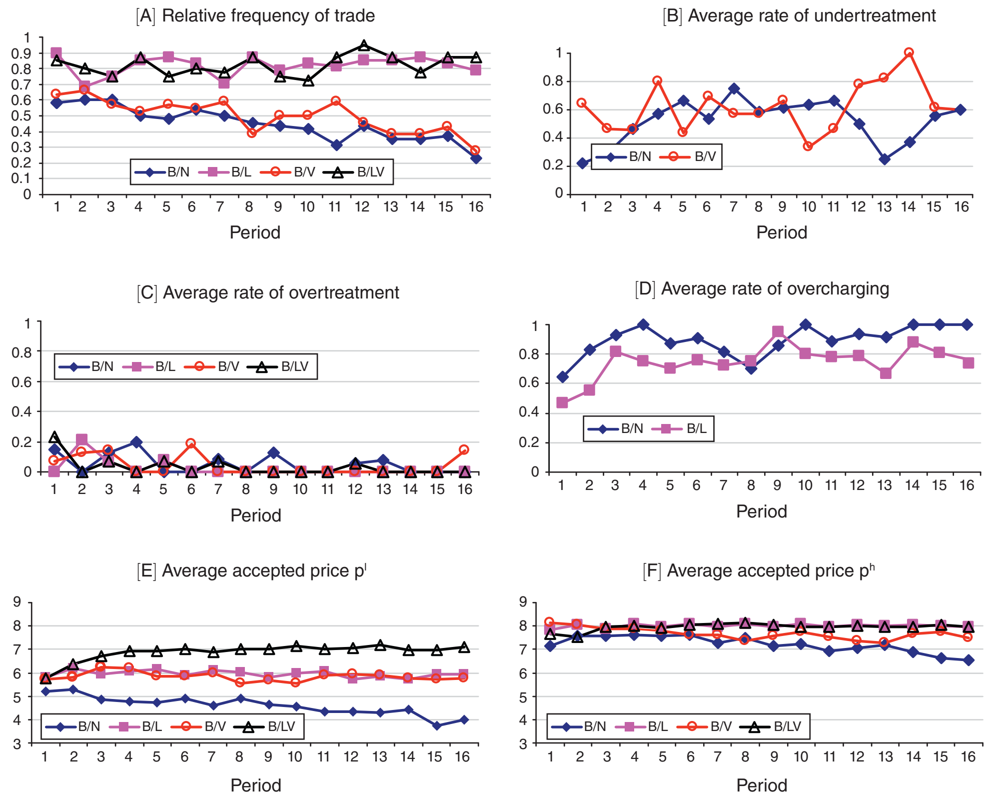

Figure 2. Trade, Provision, Charging, and Pricing Policy in Set B

that an institutional safeguard against cheating in credence goods markets has its price for consumers.

B. *Aggregate Behavior in Sets R, C, and CR*

***RESULT 2* (*On the Role of Reputation and Competition*):**

Set R. *An opportunity for reputation building (without competition) increases the volume of trade and decreases the frequency of overcharging when neither liability nor verifiability applies (in* **R/N**), *while in all other treatments reputation has no significant effect on behavior. The latter finding is consistent with prediction* **R**, *whereas the former is not.*

Set C. *Competition (without an opportunity for reputation building) decreases prices and increases the volume of trade significantly, independently of whether liability and/or verifiability applies or not. The impact of competition on trade volume in* **C/N** *(compared to* **B/N**) *is consistent with prediction* **C**. *Overall efficiency is hardly affected by the introduction of seller competition, which is also in line with prediction* **C**.

Set CR. *Adding an opportunity for reputation building to competition has virtually no effect in comparison to behavior when only competition applies, except for* *condition* **N** *where adding* **R** *to* **C** *increases the volume of trade (without significantly affecting efficiency, however).*

Table 4 presents aggregate data for all 16 experimental treatments. Comparing across the four columns within each of the four panels **N**, **L**, **V**, or **LV** allows checking how behavior is affected by competition and reputation, holding liability and verifiability constant. Again we use two-sided Mann-Whitney U-tests to indicate significant differences between two conditions, each within a given panel. The overall pattern emerging from Table 4 adds the following details to Result 2:

(i) *Trade volume and efficiency*. Reputation has a significant impact on trade volume when neither liability nor verifiability applies (63 percent in **R**/**N** versus 45 percent in **B**/**N**). In all other cases, reputation in itself has no impact on trade volume or market efficiency. The latter finding is consistent with prediction **R**, while the former is not. Competition increases the trade volume significantly in each single panel of Table 4 (comparing **B**/**X** to **C**/**X**, where **X** ∈ {**N**, **L**, **V**, **LV**}). The very high trade volume (ranging from 73 percent in **C**/**N** to 99 percent in **C**/**L**) is largely consistent with prediction **C**. But the impact of competition on efficiency is less clear cut. In particular, efficiency is not increased by introducing **C** when liability and verifiability are both violated, which is roughly consistent with prediction **C**. In all other cases the impact of competition on market efficiency is—on average—positive, albeit only significant in set **V**. Note that our efficiency definition takes into account that the scope of payoffs is larger with competition since one seller can treat several consumers, while sellers without a consumer still get the outside option. In particular, we subtract the outside option from both the actual average profit and the maximum possible average profit when calculating relative efficiency. This approach keeps the efficiency measure perfectly comparable between treatments with competition and those without competition.[17]

(ii) *Undertreatment*. In set **N** the undertreatment rate increases in the course of the experiment, while there is no discernible development in set **V** (see the online Supplement S2). In the latter set reputation appears to reduce the undertreatment rate. Competition, however, has no clear-cut effects on undertreatment.

(iii) *Overtreatment*. When verifiability applies both competition and reputation increase overtreatment while there is no such effect when verifiability is violated, as can be seen in Table 4. Recall from Section IVA that overtreatment need not be a mistake when verifiability applies.

[17] Another (and probably more elegant) way of settling this issue would have been to use an unrestricted random matching instead of the random one-to-one matching between sellers and consumers in the treatments without competition. This would have made it possible that sellers can also treat up to four consumers in sets **B** and **R**, which would have kept experimental conditions even more similar between treatments with and without competition (we thank an anonymous referee for drawing our attention to this interesting alternative design which has been used in Huck, Lünser, and Tyran 2007, for instance). However, we do not believe that the degree of market concentration has an important behavioral impact since otherwise sellers' provision and charging behavior should strongly depend on the number of consumers they have in a given period. But this is not what we observe in the experiment. For instance, the relative frequency of undertreatment is 54 percent on average if sellers have one consumer, 62 percent on average for sellers with two consumers, and 45 percent for sellers with three consumers ($p = 0.21$; Kruskal-Wallis test). Likewise, the relative frequency of overcharging is 61 percent with one consumer, 59 percent with two consumers, 62 percent with three consumers, and 44 percent with four consumers (the latter being a very rare case) ($p = 0.34$; Kruskal-Wallis test, using all observations from sets **C** and **CR**).

Table 4—Overview of Results across All Treatments

	N (*no liability / no verifiability*)				L (*liability / no verifiability*)			
Averages per period	**B/N**	**R/N**	**C/N**	**CR/N**	**B/L**	**R/L**	**C/L**	**CR/L**
Trade on cons. side[1]	0.45[a,b,c]	0.63[a,d,e]	0.73[b,d,f]	0.85[c,e,f]	0.82[b,c]	0.76[d,e]	0.99[b,d]	0.98[c,e]
Avg. # of consumers[2]	1	1	1.54	1.54	1	1	2.29	2.28
Trade on seller side[3]	0.45[a,c]	0.63[a,d]	0.47[d,f]	0.55[c,f]	0.82[b,c]	0.76[d,e]	0.43[b,d]	0.43[c,e]
Efficiency[4]	0.18	0.27	0.13	0.14	0.84	0.72[d,e]	0.92[d]	0.96[e]
Undertreatment[5]	0.53	0.56	0.73	0.64	–	–	–	–
Overtreatment[6]	0.06	0.06	0.08	0.10	0.03[c]	0.08	0.06	0.13[c]
Overcharging[7]	0.88[a,c]	0.62[a]	0.79	0.62[c]	0.75	0.70	0.70	0.66
p^l with trade	4.67[b,c]	5.09[d,e]	3.19[b,d]	3.31[c,e]	5.94[b,c]	5.47[d,e]	3.36[b,d]	3.06[c,e]
p^l without trade	5.17[b]	5.72[d]	4.16[b,d]	4.65	6.12[b]	6.07[d]	4.13[b,d]	4.23
p^h with trade	7.28[b,c]	7.60[d,e]	5.73[b,d]	6.33[c,e]	8.00[b,d]	8.02[d,e]	5.72[b,d]	5.76[c,e]
p^h without trade	7.91	8.09	7.55	7.60	8.77[b]	9.11[d]	7.64[b,d]	7.55
Actually paid price	7.08[b,c]	7.14[d,e]	5.35[b,d]	5.80[c,e]	7.77[b,c]	7.56[d,e]	5.42[b,d]	5.24[c,e]
Profits sellers[8]	2.69[a]	3.21[a,d]	2.36[d]	2.65	3.44[b,c]	2.98[d,e]	1.23[b,d]	1.19[c,e]
Profits consumers[8]	1.00	0.75	1.21	0.94	2.12[b,d]	2.24[d,e]	4.56[b,d]	4.71[c,e]
Number of subjects	96	48	48	48	96	48	48	48

	V (*verifiability / no liability*)				LV (*liability / verifiability*)			
Averages per period	**B/V**	**R/V**	**C/V**	**CR/V**	**B/LV**	**R/LV**	**C/LV**	**CR/LV**
Trade on cons. side[1]	0.50[b,c]	0.60[d,e]	0.88[b,d]	0.93[c,e]	0.82[b,c]	0.76[d,e]	0.98[b,d]	0.99[c,e]
Avg. # of consumers[2]	1	1	1.63	1.60	1	1	2.27	1.92
Trade on seller side[3]	0.50	0.60	0.54	0.58	0.82[b,c]	0.76[d,e]	0.43[b,d]	0.52[c,e]
Efficiency[4]	0.16[b,c]	0.30	0.34[b]	0.46[c]	0.81	0.71[d,e]	0.88[d]	0.93[e]
Undertreatment[5]	0.60[c]	0.38	0.53	0.36[c]	–	–	–	–
Overtreatment[6]	0.05[a,b,c]	0.12[a]	0.13[b]	0.19[c]	0.03[b,c]	0.07	0.12[b]	0.09[c]
Overcharging[7]	–	–	–	–	–	–	–	–
p^l with trade	5.84[b,c]	5.74[d,e]	4.06[b,d]	3.99[c,e]	6.88[b,c]	6.48[d,e]	4.19[b,d]	4.20[c,e]
p^l without trade	6.21[b]	6.31[d]	4.84[b,d]	4.84	7.46[b,c]	6.75[d,e]	4.59[b,d]	4.43[c,e]
p^h with trade	7.70[b,c]	7.69[d]	6.41[b,d]	6.97[c]	7.96[b,c]	8.04[d,e]	6.19[b,d]	6.38[c,e]
p^h without trade	7.82[a]	8.38[a]	7.72	7.79	8.83[b,c]	9.10[d,e]	7.94[b,d]	7.65[c,e]
Actually paid price	6.44[b,c]	6.60[d,e]	5.19[b,d]	5.58[c,e]	7.46[b,c]	7.39[d,e]	5.44[b,d]	5.49[c,e]
Profits sellers[8]	2.58[b,c]	2.51[e]	1.96[b]	1.90[c,e]	3.10[b,c]	2.83[d,e]	1.15[b,d]	1.31[c,e]
Profits consumers[8]	1.06[b,c]	1.53[e]	2.18[b]	2.59[c,e]	2.37[b,c]	2.37[d,e]	4.51[b,d]	4.49[c,e]
Number of subjects	88	48	48	48	80	48	48	48

[1] relative frequency of consumers trading with a seller
[2] average number of consumers treated by a seller who has at least one consumer
[3] relative frequency of sellers trading with at least one consumer
[4] calculated as (actual average profit − outside option) / (maximum possible average profit − outside option)
[5] relative frequency of consumer needs q^h, but seller provides q^l.
[6] relative frequency of consumer needs q^l, but seller provides q^h.
[7] relative frequency of seller provides q^l, but charges p^h (with $p^h > p^l$ and consumer needs q^l)
[8] in experimental currency units
Mann-Whitney U-tests for pairwise differences between treatments (with matching groups of eight subjects as one independent observation)
[a] B versus R ($p < 0.05$)
[b] B versus C ($p < 0.05$)
[c] B versus CR ($p < 0.05$)
[d] R versus C ($p < 0.05$)
[e] R versus CR ($p < 0.05$)
[f] C versus CR ($p < 0.05$)

(iv) *Overcharging*. Reputation reduces the overcharging rate in set **N**. While overcharging occurs in 88 percent of cases in **B**/**N**, it is reduced to about 60 percent of cases when reputation applies in **R**/**N** and **CR**/**N**. These findings indicate clearly that reputation building of sellers is beneficial for consumers as far as their risk of being overcharged is concerned. But if liability applies, reputation and competition do not influence the overcharging rate (see set

L). The different effects of reputation in sets **N** and **L** show that liability as an institutional safeguard for consumers dominates the effects of reputation as another means to cope with the informational asymmetries prevalent in credence goods markets.

(v) *Pricing*. The most prominent price vectors are {6, 8} in set **R** with 18.4 percent, and {3, 7} both in set **C** (10.9 percent) and set **CR** (9.4 percent). While vector {6, 8} does not match prediction **R** in set **R**, the prominence of vector {3, 7} is consistent with predictions **C** and **CR** (though this price vector is chosen less often than theoretically predicted). Also in line with predictions **C** and **CR** we find that prices p^l and p^h are significantly lower with *c*ompetition than without. On average, seller competition drives down prices by about two units. This can be clearly observed in Table 4 (and in panels E and F of Figure S1 in the online Supplement S2). Table 5 takes a closer look at the determinants of consumers' choice of a seller when competition holds by looking at the properties of accepted price vectors. Theory predicts that consumers visit the seller with the lowest p^h if verifiability is violated, but that consumers' choice of a seller depends on both prices if **V** holds. From Table 5 it becomes clear that with verifiability the share of accepted price vectors that include the lowest price p^h goes down, whereas price vectors with the minimum price p^l (that do not also have the minimum price p^h) become more often accepted.[18] Thus, consumers' choice of a seller is largely in line with predictions **C** and **CR**.

(vi) *Profits*. With reputation, the profits of sellers and consumers are basically the same as in set **B** without reputation. This implies that consumers earn less than their outside option when liability does not hold in **R**/**N** and **R**/**V**. Introducing competition shifts the gains from trade (almost completely) from sellers to consumers, as expected from predictions **C** and **CR**. However, competition actually drives down sellers' profits below their outside option whenever liability applies.[19] Although from the theoretical prediction sellers should sell their credence goods also when competition plus liability apply, the competitive pressure on prices (to attract consumers) is so strong that staying out of the market would have been better for sellers on average.

C. *Estimating the Effects of Liability, Verifiability, Reputation, and Competition*

In Table 6 we report the coefficients from random effects probit regressions where we examine the impact of liability, verifiability, reputation building and competition on the relative frequency of trade, undertreatment, overtreatment, and overcharging.[20]

[18] Applying a χ^2-test, we find that the distribution of accepted price vectors between rows a) and b) in Table 5 (both for **C** and for **CR**) is significantly different between **N** and **V**, and between **L** and **LV** ($p < 0.05$ in both cases). Another way of stating this result is that the share of accepted price vectors that include the minimum price p^h is significantly higher in **N** than in **V**, and also higher in **L** than in **LV** ($p < 0.05$ in both cases).

[19] For calculating sellers' average profits in sets **C** and **CR** we subtract excessive outside options (that exceed the difference between 4 and the number of actual trades in a matching group) in order to make sellers' profits comparable to the sets without competition (i.e., sets **B** and **R**).

[20] With respect to model selection in Table 6 note that none of the third-order or fourth-order interaction effects of the main treatment variables, *liability*, *verifiability*, *reputation*, and *competition* (as well as the interaction of these higher-order effects with period and prices), had been found significant when including them as independent variables. The model fit was also improved by dropping any interaction terms where either the period or any of

TABLE 5—CONSUMERS' CHOICE OF SELLER WHEN COMPETITION APPLIES (*sets C and CR*)

	C (*competition / no reputation*)			
	C/N	**C/L**	**C/V**	**C/LV**
Relative frequency with which a consumer chooses a seller with				
a) the minimum price p^h (irrespective of order of p^l)	0.68	0.90	0.52	0.77
b) the minimum price p^l (but not the minimum price p^h)	0.13	0.02	0.23	0.15
c) neither of the above	0.19	0.08	0.25	0.08
Number of observations	281	381	337	377
	CR (*competition / reputation*)			
	CR/N	**CR/L**	**CR/V**	**CR/LV**
Relative frequency with which a consumer chooses a seller with				
a) the minimum price p^h (irrespective of order of p^l)	0.64	0.87	0.44	0.76
b) the minimum price p^l (but not the minimum price p^h)	0.17	0.08	0.33	0.13
c) neither of the above	0.19	0.05	0.23	0.11
Number of observations	326	378	359	382

Column 1 of Table 6 considers the likelihood of *trade* for a seller.[21] Looking at the main treatment effects one can see that liability and competition have a significant effect, whereas verifiability and reputation per se are insignificant. All other things being equal, liability has a strong positive effect on the likelihood of trade, because consumers can be sure to receive a sufficient quality. Competition has a significant negative main effect on the probability of a particular seller attracting a consumer for trade, since it leads to a concentration of several consumers interacting with the same seller, leaving other sellers without any consumer. The row "average # of consumers" in Table 4 shows that sellers that attract at least one customer serve, on average, between 1.54 and 2.29 consumers in treatments with seller competition. This concentration leaves other sellers without any consumers, hence the negative coefficient for competition in column 1 of Table 6. Contrary to the theoretical prediction, verifiability has no significant main effect. Reputation also lacks a significant main effect, consistent with prediction **R**. Both prices, p^l and p^h, have a significant negative effect on the likelihood of trade. Recall that predictions **B** and **R** had implied for sets **N** and **L** that the low price p^l has no impact on a consumer's decision to enter the market. This is obviously not what we observe. Rather, the low price lures consumers into the market. Another noteworthy effect is the interaction of the high price with verifiability. If verifiability holds, the negative main effect of the high price p^h on the likelihood of trade is reduced because an increase in the high price renders undertreatment less attractive for sellers. Similarly, reputation has a positive interaction effect with the high price p^h, indicating that reputation allows for a (modest) increase in the high price without endangering the consumer's willingness to trade with the seller in set **N**. Finally, note that the likelihood of trade is significantly declining across periods—which is an indication of learning—and that this downward trend is only partially offset when liability or verifiability applies.

the two prices (p^l or p^h) had been interacted with a second-order interaction of the main treatment variables (for example, "Period × liability × verifiability").

[21] We use the frequency of trade on the seller (and not on the consumer) side as the dependent variable here because among the independent variables there are the prices p^h and p^l and because in the competition treatments (where consumers can choose from four price vectors) it would not be clear on which prices the frequency of trade on the consumer side should be conditioned upon.

Table 6—Panel Probit Regressions Using Data from All Conditions

Independent variables	Trade on seller side [1]	Undertreatment [2]	Overtreatment [3]	Overcharging [4]
Liability (**L** = 1)	5.45**	NA	0.30	−0.64
Verifiability (**V** = 1)	−0.55	0.19	0.70	NA
Reputation (**R** = 1)	0.02	−1.69*	−0.08	−0.71
Competition (**C** = 1)	−0.85*	−0.13	0.22	−1.25**
Liability × Verifiability	−0.00	NA	0.04	NA
Liability × Reputation	−0.41**	NA	0.28	0.60*
Liability × Competition	−2.16**	NA	−0.07	0.34
Verifiability × Reputation	−0.11	−0.63*	−0.53	NA
Verifiability × Competition	0.26*	−0.12	−0.66*	NA
Reputation × Competition	−0.04	1.84*	0.84	1.25
Price p^l	−0.13**	0.03	−0.38**	−0.61**
Price p^h	−0.46**	−0.56**	0.27**	0.39**
Price p^l × Liability	0.03	NA	−0.11	0.21*
Price p^h × Liability	−0.56**	NA	−0.09	−0.10
Price p^l × Verifiability	−0.05	0.23**	−0.60**	NA
Price p^h × Verifiability	0.11**	−0.11	0.35**	NA
Price p^l × Reputation	−0.06*	−0.11	0.04	−0.18
Price p^h × Reputation	0.08*	0.31**	0.02	0.13
Price p^l × Competition	0.00	0.10	0.08	0.01
Price p^h × Competition	0.04	−0.30**	−0.18	−0.19
Period	−0.08**	0.05**	−0.08**	0.06**
Period × Liability	0.04**	NA	−0.03	−0.02
Period × Verifiability	0.03**	−0.05**	0.06**	NA
Period × Reputation	0.02	0.00	0.02	−0.03
Period × Competition	−0.01	−0.01	−0.01	0.03
Intercept	4.75**	3.75**	−1.61**	0.96
Number of Observations	7,488	1,225	2,902	1,205

Notes: NA not applicable, i.e., a given action (dependent variable) is not possible in a particular condition (independent variable).

**Significant at the 1 percent level.

*Significant at the 5 percent level.

Column 2 of Table 6 reports an estimation of the likelihood of *undertreatment*. The most interesting finding from the viewpoint of our predictions is the significant negative effect of the high price p^h, while the low price p^l is not significant per se. Note, however, that the interaction terms between both prices and verifiability show the predicted sign, although only the term price "p^l × verifiability" is significant. Undertreatment also becomes more likely with more experience across periods. This is an indication that some sellers learn how to exploit the opportunities arising in credence goods markets. Column 3 of Table 6 reveals that the likelihood of *overtreatment* depends on the low and the high price (as such and in the interaction with verifiability) in the predicted direction. The negative effect of price p^l and the positive effect of price p^h indicate that overtreatment increases on average in the price difference $p^h - p^l$. *Overcharging* (see column 4 of Table 6) increases significantly across periods and depends otherwise mainly on a seller's prices. The likelihood of overcharging decreases with the low price p^l and increases with the high price p^h, which is consistent with the theoretical predictions.

V. Discussion and Conclusion

In this paper we have analyzed in a large-scale experiment with 936 participants the behavior of experts and consumers on credence goods markets. These markets are prone—both in reality and in our experiment—to problems of undertreatment, overtreatment, and overcharging. We start the conclusion by noting that even under the most disadvantageous market conditions for consumers, the experimental sellers of credence goods have provided an appropriate quality in about half of the cases. Furthermore, some sellers have done so very consistently. For instance, in treatment **B/N** we have observed that 27 percent of the sellers have provided the appropriate quality throughout the whole 16 periods of the experiment. Hence, even in the absence of liability laws, a substantial fraction of sellers is honest. This finding might be explained by social or moral norms (like not to exploit or cheat on others; see Uri Gneezy 2005), by guilt aversion (Gary Charness and Dufwenberg 2006), or by experts having nontrivial distributional preferences in the sense that they consider not only their own but also their customers' material payoffs when making their provision decisions (see also our discussion below).

The main purpose of this paper has been to study which other means—besides an intrinsic motivation, a moral commitment, or a concern for the (material) well-being of others—may yield efficient interaction in credence goods markets. More precisely, we have examined in a $2 \times 2 \times 2 \times 2$ experimental design the role of *liability*, *verifiability*, *reputation*, and *competition*. Contrary to standard theory's prediction, *verifiability* has been found to have almost no effect (compared to a situation without institutional safeguards against fraud). It seems safe to conclude, therefore, that forcing sellers of credence goods to charge only for the product quality they have actually provided does not approach the roots of the problems.[22] At first glimpse the main reason for why verifiability does a poor job in enhancing efficiency seems to be the very rare use of equal markup vectors. This interpretation does not really explain the data, however, for two (related) reasons. First, it only defers the question mark to the next level, as it leaves unanswered the question why sellers choose most often undertreatment vectors in the first place. Secondly, as we have seen in Section IVA, experts' provision behavior under equal markup prices in condition **B/V** is far from the predicted one, with overtreatment (instead of providing appropriate quality) being observed in up to 40 percent of cases. Below we argue that experts having (heterogeneous) distributional preferences can explain both, why (certain) equal markup price vectors work particularly poorly and why equal markup vectors are not chosen in the first place.

Liability to provide a sufficient quality has been found to have a very strong effect not only on the likelihood of trade on our experimental credence goods markets, but also on its efficiency. It seems straightforward that liability makes it very attractive for consumers to trade with sellers, as undertreatment is precluded by definition. However, overtreatment (which may also create inefficiencies) and overcharging (which leads to inefficiencies if the fear of getting overcharged deters consumers from interacting with expert sellers) are still possible in **B/L,** and they might—at least in theory—also impose substantial costs on consumers. As we have seen, in the experiment they do not: Overtreatment is not a very important phenomenon in any of the experimental conditions and overcharging is not very costly for consumers

[22] Hence, it cannot be an effective remedy of the problems on credence goods markets to ask a car mechanic, for example, to put the replaced parts in the boot of the car in order to verify the presumed action. Somewhat ironically, this was one of the conclusions offered by Dulleck and Kerschbamer (2006) in their theoretical treatment of credence goods markets. Our experimental study clearly shows the need to test the behavioral relevance of theories, as in our experiment we have found that verifiability is practically of no help.

since experimentally observed price differences between different qualities are typically small. However, as we have seen, an institutional safeguard against undertreatment has another important cost for consumers as they pay, on average, far higher prices with than without liability. This connection between liability and high costs might also be one of the factors that increase the costs of providing health care when doctors are required to provide a sufficient quality.[23]

Reputation as implemented in the experiment has been found to be effective in increasing trade in credence goods markets, but only when none of the other effective means (liability and competition, as it turns out) is present. In this case (**R**/**N**), reputation building benefits only the sellers, though, as they attract more trade for each price level (since consumers are more likely to trust them), but do not change their behavior; i.e., the undertreatment rate is not decreasing in the aggregate when reputation building is possible. This may be interpreted as evidence that an opportunity to build up reputation can even deteriorate, rather than improve, the situation for customers as they might become too credulous. The absence of a reputation effect in the other conditions may—at least in part—be driven by the fact that the possibility for reputation formation is rather limited in our design since buyers could only recall their own experience, but not how other buyers fared with a particular seller. It is an open question for future research whether considering a public information condition—where the consumers in the market get to know the experiences of all other consumers with a particular seller—might yield stronger reputation effects. It is not clear, however, which of the two informational conditions is empirically the more relevant one. In reality, consumers can always be sure about their own experience with a particular expert. Of course, they might get some fuzzy information from neighbors and friends, but only rarely they observe the whole history of a particular seller.

Competition among sellers has been found to be very influential, first of all, in bringing down the prices for credence goods (compared to a situation of a bilateral matching of sellers and consumers), and, second, in increasing the trade volume on the market. This is not to say that competition solves all of the problems on credence goods markets, because we have also found that the likelihood of undertreatment, overtreatment, or overcharging, is not reduced in the aggregate (but interestingly not increased either!) through seller competition. This means that the effects of competition on trade are first and foremost driven by price cuts through competition, while the overall level of efficiency is not increased through competition (except when verifiability applies).

As noted in the introduction, our paper is most closely related to two experimental papers by Huck, Lünser, and Tyran (2007, 2008) where they study how reputation and competition affect the efficiency on markets for experience goods. Our model of a credence goods market can be seen as a generalization of their experience goods market model. In particular, if we set the probability of needing the high quality in our

[23] It has to be noted, though, that strict liability is difficult to impose in real world markets. On the one hand, it requires a form of verifiability of the outcome. Especially in the medical realm, treatment success is often impossible or very costly to measure for a court, while still being observed by the consumer (how can one prove the presence/absence of pain, for instance?). On the other hand, even in cases where the outcome is verifiable it often not only depends on the quality provided by the seller but also on consumer behavior. In both cases issues of moral hazard on the consumer side may rule out strict liability. Dulleck and Kerschbamer (2006) discuss these and other issues in a section on the theory of credence goods and real world examples.

credence goods market equal to one ($h = 1$ in our notation), then our **N**-treatments (**B**/**N**, **R**/**N**, and **C**/**N**) corresponded exactly to their design of a market for experience goods. Huck, Lünser, and Tyran (2008) compare the equivalents of our **B**/**N**- and **R**/**N**-treatments, and Huck, Lünser, and Tyran (2007) relate analogously **B**/**N** to **C**/**N**. Huck, Lünser, and Tyran (2007, 2008) report pretty large effects of reputation and competition. Whereas competition has also very strong effects in our design, we find much weaker effects of reputation. One possible explanation is that cheating is always detected in the set up of Huck, Lünser, and Tyran (2008)—which makes a good reputation all the more valuable—whereas in our design cheating may remain undetected. The latter possibility weakens the importance, and ultimately the effects, of reputation.

Regarding the issue of parallelism of our experimental results and real-world credence goods markets, we note that we have identified in our experiment undertreatment (and to a lesser extent overcharging) as the main source of inefficiencies. In markets for medical treatments—a prime example of a credence goods market—overtreatment seems to be a key problem, however. Our experiment differs from real-world medical markets in some important dimensions which might explain this difference. First, prices in the health care sector are typically exogenous to the sellers of health care services. If those exogenously given price vectors are of the overtreatment type (which is very rarely the case in our experiment with endogenous prices), then overtreatment may become more likely due to financial incentives. For example, in some countries the markups for Cesarean deliveries seem to be higher than those for normal child births, increasing the likelihood of Cesarean deliveries (Gruber, Kim, and Mayzlin 1999). Second, diagnosis is assumed to be perfect (in the sense of revealing the appropriate quality for sure) in our experiment while it is almost always imperfect in reality. If providers of credence goods can be sued for insufficient quality (meaning that some form of liability applies), imperfect diagnosis creates incentives for overtreatment. Third, in the realm of medical services patients typically do not pay the fees for health care provision themselves, but insurance companies do. This fact reduces the incentives for patients to take care of themselves, and it makes patients less worried about overtreatment as they do not carry the full social costs. These differences may explain why in the market for health care services overtreatment seems to be a bigger problem than undertreatment and overcharging. However, many other markets for credence goods do suffer from problems created by undertreatment and overcharging which have been identified as main problems in our experiment. For instance, Henry Schneider (2006) finds in his field study on fraudulent behavior in the auto repair business substantial evidence of all three kinds of fraud: (i) completely unnecessary repairs or repairs that use more labor or parts than necessary (i.e., overtreatment); (ii) neglect of defects that require urgent attention (undertreatment); and (iii) billing for parts and labor not provided (overcharging). Also, taxi rides in an unknown city may not only be prone to attempts of overtreatment (by taking a circuitous route) but also result in overcharging (by manipulating the taximeter and therewith the distance traveled, for example) and even undertreatment (failure to reach the desired destination).[24]

[24] See a recent test of the ADAC (Allgemeiner Deutscher Automobil Club, Germany's largest automobile club) on http://www1.adac.de/Tests/Mobilitaet_und_Reise/taxitest/ergebnisse/default.asp?ComponentID = 234050&SourcePageID = 235884.

We conclude by noting that our experimental results might be an important stimulus for the further development of the theory of credence goods markets. Our theoretical analysis of the experimental behavior has been guided by a model that assumes standard economic preferences (along the lines developed in Dulleck and Kerschbamer 2006). It is noteworthy that despite its simplicity the model has been able to predict several of our key findings at least in a qualitative way (for example, the importance of liability for efficiency and trade, the impact of markup differences on provision behavior under verifiability, the lack of an influence of reputation, and the tremendous effect of competition on prices and volume of trade, as well as consumers' reaction to different price vectors). Our simple model has failed, however, in predicting the null effect of verifiability on trade volume and efficiency on credence goods markets. More specifically, we have found that market performance in conditions with verifiability (but without liability) is substantially worse than standard theory's prediction, while market performance in conditions without verifiability (and without liability) is much better than predicted, leading to both conditions performing equally (well or poor) in the aggregate. What causes those results? In a companion paper (Kerschbamer, Sutter, and Dulleck 2009) we argue that experimental experts' heterogeneous distributional preferences are the key to understanding the failure of verifiability to induce trade and increase efficiency in experimental credence goods markets. There we also argue that experts' heterogeneous distributional preferences can not only explain the failure of verifiability but also the fact that market performance under **N** is better than predicted and the fact that the performance under **L** is almost exactly as predicted. Before presenting the arguments, it is important to recall from the results section that distributional concerns seem to play an important role in our experiment. The price vector {6, 8}—which has no special appeal from standard theory's perspective—is by far the most prominent price vector at the price posting stage of the game in all four basic treatments of set **B** (accounting for 38 percent of observations in **B**/**V**, for instance). As mentioned earlier, this price vector would split the gains from trade equally between sellers and consumers—*if* sellers always provided the appropriate quality and always charged for the delivered quality—and its prevalence is therefore an indication that relative payoffs matter for experts' behavior. Also, as discussed earlier, the most often chosen equal markup vector in **B**/**V** (i.e., {4, 8}) induces a considerable amount of overtreatment which contradicts the standard theory's prediction, but which is exactly what theories of inequality aversion (e.g., Ernst Fehr and Klaus Schmidt 1999, Gary E. Bolton and Axel Ockenfels 2000) would predict.[25] Finally, as we have seen earlier, there is a considerable amount of appropriate treatment even in the condition without institutional safeguards against fraud (i.e., even in set **N**), which is an indication that a taste for efficiency (Charness and Matthew Rabin 2002) might play a role for experts' behavior.

[25] Kerschbamer, Sutter, and Dulleck (2009) show that the overtreatment prediction under the equal markup vector {4, 8} does not depend on the functional form in which inequality-averse preferences are modelled (let alone on specific parameterizations of a specific functional form), but is a consequence of inequality aversion per se. That is, *any* model of inequality aversion that has the equal split as the reference point *must* predict overtreatment under this vector.

In Kerschbamer, Sutter, and Dulleck (2009) we provide a thorough theoretical analysis of the impact of different types of distributional preferences for experts' provision behavior under verifiability. Based on the theoretical analysis we then develop a theoretical test that allows classifying experimental experts in different distributional types according to their provision behavior in a market with exogenously given prices (in contrast to the endogenous determination of prices in the present paper). In the implementation of the test we find that only a minority of experts behave according to standard theory's prediction while a clear majority of the subject population exhibits nontrivial distributional preferences.

How do these findings relate to the main results in the present study? The crucial insight is that one and the same heterogeneity in distributional preferences can lead to a positive or a negative deviation from standard theory's prediction, depending on subtle features of institutional design. The positive deviation potentially comes either from efficiency-loving experts who are willing to help their customers if the "price for helping" is not too high or from inequality-averse experts who care for customers in the domain of advantageous inequality. The negative deviation is caused by inequality-averse experts who have a propensity to harm their customers in the domain of disadvantageous inequality or by competitive experts willing to hurt customers if the "price for hurting" is not too high.

Consider set **N** first. Under this condition the downside of distributional preferences cannot manifest itself in the market outcome since the standard prediction (undertreatment and overcharging under each price vector) is already a worst case scenario. By contrast, the positive side of distributional preferences immediately manifests itself in a better market outcome than predicted under standard preferences. Under set **V**, by contrast, we get almost the opposite result (where the "almost" disappears in the case of equal markup price vectors, where experts can change their customer's material payoff at no personal cost): Under equal markup price vectors the standard prediction (appropriate treatment independent of the level of markups) is already a best case scenario, so the positive side of distributional preferences cannot improve it. By contrast, the downside of distributional preferences easily manifests itself in the market outcome because hurting the customer involves no cost under equal markup vectors (and little cost under vectors that do not deviate too far from the equal markup rule). Taking the evidence from the **N**-condition and the **V**-condition together, it follows that none of these market institutions is robust against the coexistence of experts with heterogeneous distributional preferences. But this is not the case for liability.

In set **L** the downside of distributional preferences cannot lead to deviations from the standard prediction since undertreatment is ruled out by design and since overcharging is already the standard prediction. Hence, only overtreatment remains as a possible harm to efficiency, but overtreatment (in comparison to overcharging) only reduces the expert's material payoff without affecting that of the customer and is therefore unattractive for experts independent of their distributional type. The positive side of distributional preferences cannot have a deep impact either, since the benchmark prediction is already full efficiency—so the only positive impact can be a distributional one, and indeed we observe overcharging rates that are high, but far below the predicted 100 percent.

Hence, we can explain the (unexpected) differences in the effects of liability and verifiability by an *asymmetry in the robustness of the two institutions to the* *coexistence of experts with heterogeneous distributional preferences*.[26] There are other asymmetries which might contribute to the difference. First, there is an *asymmetry in worst-case payoffs*: while in the case of undertreatment (which is only possible in the treatments without liability, i.e., in **N** and **V**) consumers necessarily experience actual losses, overtreatment (possible in all treatments but dominated by overcharging in the treatments without verifiability) and overcharging (possible only in the treatments without verifiability, i.e., in **N** and **L**) do not imply real losses in practice, but rather "only" lower profits (which may be perceived as less severe by consumers). If consumers care predominantly about losses, then they are more likely to abstain from the market when losses are possible than when they are not. Note, however, that this asymmetry (in worst-case payoffs) is only a practical, but not a theoretical one, since theoretically losses are also possible in treatments with liability when sellers post and charge prices that exceed consumers' valuation for a successful intervention. However, in the experiment price vectors with $p^h > v$ are very rarely posted by sellers and almost never (i.e., only in 0.3 percent of cases) accepted by customers. Also note that the asymmetry in worst-case payoffs is less a peculiarity of our experimental design and more an immanent characteristic of the good under consideration.[27] Verifiability and liability may also have had different effects because of an *asymmetry in the precision of feedback* given after an interaction. Undertreatment can be identified by consumers in our experiment, while overtreatment and overcharging can not. If consumers care predominantly about (not) being defrauded by experts, then they are more likely to abstain from the market when fraud is visible (as it is in the case of undertreatment in sets **N** and **V**) than when it is not (as it is in the case of overcharging in set **L**). Again, we do not consider this asymmetry (in the precision of feedback) as a peculiarity of our experimental design but rather as an idiosyncratic feature of the good under consideration. In sum, we regard the different explanations for the different effects of liability and verifiability as complementary to each other. While the asymmetry in robustness of institutions story gives a (market supply side) explanation for why we observe more fraud than predicted under standard assumptions in sets **N** and **V** and less fraud than predicted in set **L**, the other two asymmetries focus on the demand side of the market, i.e., on consumers' reactions to fraud.

Our overall conclusion is twofold: First, the results of our experimental investigation clearly indicate the need to test the behavioral relevance of theoretical results, especially if they are used to derive policy conclusions. Secondly, more theoretical research is needed on the impact of experts with heterogeneous other-regarding preferences on the performance of markets for credence goods.

[26] It is important to note that our arguments do not depend on *any* assumptions on the distribution of different social-preference types in the population (i.e., on which fraction of the subjects exhibit inequality aversion, which fraction a taste for efficiency, etc.) and on how specific social preferences are modelled. What is important, though, is that there is some heterogeneity in preference types.

[27] Undertreatment refers to a situation where a consumer pays a price for a good that does not work properly—a situation which has great potential to be a loss situation also in reality. By contrast, overtreatment and overcharging can also in real markets only lead to losses if the consumer accepts prices that exceed his valuation for a successful intervention.

REFERENCES

Afendulis, Christopher C., and Daniel P. Kessler. 2007. "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care." *American Economic Review*, 97(3): 1013–20.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics*, 84(3): 488–500.

Alger, Ingela, and Francois Salanié. 2006. "A Theory of Fraud and Overtreatment in Experts Markets." *Journal of Economics and Management Strategy*, 15(4): 853–81.

Baksi, Soham, and Pinaki Bose. 2007. "Credence Goods, Efficient Labelling Policies, and Regulatory Enforcement." *Environmental and Resource Economics*, 37(2): 411–30.

Berg, Joyce, John W. Dickhaut, and Kevin A. McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior*, 10(1): 122–42.

Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review*, 90(1): 166–93.

Charness, Gary, and Martin Dufwenberg. 2006. "Promises and Partnership." *Econometrica*, 74(6): 1579–1601.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117(3): 817–69.

Darby, Michael R., and Edi Karni. 1973. "Free Competition and the Optimal Amount of Fraud." *Journal of Law and Economics*, 16(1): 67–88.

Dulleck, Uwe, and Rudolf Kerschbamer. 2006. "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods." *Journal of Economic Literature*, 44(1): 5–42.

Dulleck, Uwe, Rudolf Kerschbamer, and Matthias Sutter. 2009. "The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition." IZA Discussion Paper 4030.

Emons, Winand. 1997. "Credence Goods and Fraudulent Experts." *RAND Journal of Economics*, 28(1): 107–19.

Feddersen, Timothy J., and Thomas W. Gilligan. 2001. "Saints and Markets: Activists and the Supply of Credence Goods." *Journal of Economics and Management Strategy*, 10(1): 149–71.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics*, 114(3): 817–68.

Fischbacher, Urs. 2007. "Z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10(2): 171–78.

Fong, Yuk-fai. 2005. "When Do Experts Cheat and Whom Do They Target?" *RAND Journal of Economics*, 36(1): 113–30.

Gneezy, Uri. 2005. "Deception: The Role of Consequences." *American Economic Review*, 95(1): 384–94.

Greiner, Ben. 2004. "An Online Recruitment System for Economic Experiments." In *Forschung und wissenschaftliches Rechnen: Beiträge zum Heinz-Billing-Preis 2003*, ed. Kurt Kremer and Volker Macho, 79–93. Göttingen: Klartext.

Gruber, Jonathan, and Maria Owings. 1996. "Physician Financial Incentives and Cesarean Section Delivery." *RAND Journal of Economics*, 27(1): 99–123.

Gruber, Jonathan, John Kim, and Dina Mayzlin. 1999. "Physician Fees and Procedure Intensity: The Case of Cesarean Delivery." *Journal of Health Economics*, 18(4): 473–90.

Hubbard, Thomas N. 1998. "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market." *RAND Journal of Economics*, 29(2): 406–26.

Huck, Steffen, Gabriele K. Lünser, and Jean-Robert Tyran. 2007. "Competition Fosters Trust." ELSE Working Paper 235.

Huck, Steffen, Gabriele K. Lünser, and Jean-Robert Tyran. 2008. "Pricing and Trust." ELSE Working Paper 245.

Hughes, David, and Brian Yule. 1992. "The Effect of Per-Item Fees on the Behaviour of General Practitioners." *Journal of Health Economics*, 11(4): 413–37.

Iizuka, Toshiaki. 2007. "Experts' Agency Problems: Evidence from the Prescription Drug Market in Japan." *RAND Journal of Economics*, 38(3): 844–62.

Kerschbamer, Rudolf, Matthias Sutter, and Uwe Dulleck. 2009. "The Impact of Distributional Preferences on (Experimental) Markets for Expert Services." University of Innsbruck Faculty of Economics and Statistics Working Paper 2009-28.

Nelson, Phillip. 1970. "Information and Consumer Behavior." *Journal of Political Economy*, 78(2): 311–29.

Pesendorfer, Wolfgang, and Asher Wolinsky. 2003. "Second Opinions and Price Competition: Inefficiency in the Market for Expert Advice." *Review of Economic Studies*, 70(2): 417–37.

Schneider, Henry. 2006. "A Field Experiment to Measure Agency Problems in Auto Repair." Unpublished.

Wolinsky, Asher. 1993. "Competition in a Market for Informed Experts' Services." *RAND Journal of Economics*, 24(3): 380–98.

Wolinsky, Asher. 1995. "Competition in Markets for Credence Goods." *Journal of Institutional and Theoretical Economics*, 151(1): 117–31.