MAX PLANCK INSTITUTE
FOR BEHAVIORAL ECONOMICS

Loukas Balafoutas
Helena Fornwagner
Rudolf Kerschbamer
Matthias Sutter
Maryna Tverdostup

Discussion Paper
2023/11

DIAGNOSTIC UNCERTAINTY AND INSURANCE COVERAGE IN CREDENCE GOODS MARKETS. AN EXPERIMENT

Forthcoming in:
Journal of Economic Behavior and Organization

Diagnostic uncertainty and insurance coverage in credence goods markets. An experiment [§]

Loukas Balafoutas[°*], Helena Fornwagner[°++], Rudolf Kerschbamer[*],

Matthias Sutter[*#+], and Maryna Tverdostup[†]

[°]University of Exeter, United Kingdom, [*]University of Innsbruck, Austria,

[++] Austrian Institute of Economic Research (WIFO)

[#]Max Planck Institute for Behavioral Economics, IZA@LISER and CESifo Munich, Germany

[+]University of Cologne, Germany, [†]Vienna Institute for International Economic Studies, Austria

Abstract

Credence goods markets are prone to fraudulent behavior and market inefficiencies due to informational asymmetries between sellers and customers. We examine experimentally the effects of diagnostic uncertainty and insurance coverage on the information acquisition and provision decisions by sellers and the trading decisions by consumers. Our results reveal that diagnostic uncertainty is a major source of inefficiency that decreases efficient service provision. Insurance coverage has a positive net effect on market efficiency, despite making information acquisition and efficient service provision less likely.

JEL-codes: C91, D82, G22

Keywords: credence goods; diagnostic uncertainty; insurance coverage; experiment

Declarations of interest: none

This version: 20 February 2026

[§] We thank Matthias Heinz, Marco Schwarz, an associate editor and two referees, participants at the Annual Conference of the Western Economic Association, CEAR / MRIC Behavioral Insurance Workshop, Workshop in Behavioral and Experimental Health Economics, Annual Meeting of the Austrian Economic Association, and seminar participants at the University of Düsseldorf (DICE), University of Cologne and MPI Bonn for helpful comments. Financial support from the Austrian Science Fund (FWF) through special research area grant SFB F63 and grant numbers P26901 and P27912, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/2– 390838866 is gratefully acknowledged. This study has received a Certificate of Good Standing (IRB approval) from the Board for Ethical Questions in Science of the University of Innsbruck (approval number 50/2020). Corresponding author: Matthias Sutter, Max Planck Institute for Behavioral Economics, Kurt-Schumacher-Strasse 10, D-53113 Bonn, Germany. sutter@coll.mpg.de.

1. Introduction

In markets for credence goods, expert sellers (e.g., doctors, mechanics, legal or financial advisors) are better informed than consumers about the quality of the good, service, or asset that fits consumers' needs best. This informational advantage of sellers creates incentives to cheat consumers – by providing too much (*overtreatment*), too little (*undertreatment*), or charge for more than has been provided (*overcharging*). The incentives for fraudulent behavior may cause large inefficiencies. This is particularly worrisome because credence goods markets have a huge volume and cover a large variety of sectors and industries (Dulleck and Kerschbamer, 2006). For instance, in the U.S.A., health care services account for about 17% of GDP[1], the finance industry represents about 20% of GDP[2], and car repair services, another prominent example for credence goods, generated total revenues of about 70 billion Dollars in 2019.[3]

In analyzing the informational asymmetries between expert sellers and consumers, most of the existing work assumes that experts can identify their consumers' needs perfectly, and often at no cost (Balafoutas and Kerschbamer, 2020). In reality, however, this is seldom the case. Diagnosis is usually costly because it requires the expert to invest time and effort (not all of which may necessarily be compensated, for instance, because it is not verifiable and therefore not contractible), and it may even fail to identify a consumer's problem perfectly (Schneider, 2012). Another common assumption in the literature is that consumers bear the full cost of service. In contrast to this assumption, in many real-world credence goods markets consumers are insured and thus often have zero, or negligible, marginal costs of additional services.

This paper presents an experimental investigation of the role of diagnostic uncertainty and insurance coverage, and their potential interaction, for behavior on credence goods markets. One motivation for focusing on diagnostic uncertainty and insurance coverage is the fact that diagnostic uncertainty introduces risk in some dimension (quality), while insurance conversely reduces risk in another dimension (price), and the combination of this affects incentives. Moreover, because the insurance-diagnosis interaction is hard to identify in the field—owing to selective contracting, unobserved case mix, ethical and compliance constraints, and limited measurement of diagnostic precision—the laboratory setting that we use looks especially suitable for our main research question.

[1] See http://www.oecd.org/els/health-systems/health-expenditure.htm (accessed 29 March 2025).
[2] See the United States' Bureau of Economic Analysis: https://www.bea.gov/data/gdp/gdp-industry (accessed 29 Macrh 2025).
[3] See https://www.ibisworld.com/industry-trends/market-research-reports/other-services-except-public-administration/repair-maintenance/auto-mechanics.html (accessed 29 March 2025).

Ours is the first paper to examine the behavior of expert sellers and consumers in a controlled environment that departs from perfectly informative and costless diagnosis. The interaction between diagnostic uncertainty and insurance coverage may be manifold. For example, insurance coverage may induce consumers to think themselves safe with respect to potential problems arising from diagnostic uncertainty. However, insurance coverage may reduce the expert's incentives to invest in costly diagnosis, leading to more diagnostic uncertainty. Also, for a given level of diagnostic uncertainty, insurance coverage may lead to less efficient service provision. Conversely, higher uncertainty in the diagnosis and less efficient service provision are likely to increase consumer expenditures, creating a stronger need for insurance coverage, thus ultimately inflating expenditures for health care.

Our experiment is built on a theoretical model by Balafoutas et al. (2023). It involves 576 participants and is based on a 3 x 2 factorial design, varying the following factors: (i) *Diagnostic Uncertainty*. Experts receive a costly signal about the consumer's problem that is either 100% precise, 70% precise, or of endogenous precision. (ii) *Insurance*. Consumers are either fully insured, i.e., the full price of service is covered by an insurance in return for a premium, or not insured, in which case they have to pay the full price of service themselves. In addition, we implement three price vectors that imply different incentives for experts. This design allows for a direct test of the model's predictions regarding the impact of diagnostic uncertainty and insurance coverage on information acquisition and efficient service provision.

Our findings largely confirm the main predictions. First, we identify diagnostic uncertainty as a major source of inefficiencies in credence goods markets. Comparing the two treatments that vary the exogenously given diagnostic precision, we find that less precision leads to lower rates of efficient service provision and lower rates of consumer willingness to trade. These effects hold independently of insurance coverage. Second, we find empirical support for the prediction that insurance coverage reduces the rate of efficient service provision (in the treatments with exogenous precision) and the average investment in diagnostic precision (in the treatment where precision is endogenous). As a side result, we can also show that the net impact of insurance on market efficiency is nevertheless positive, thanks to the counterbalancing force of more frequent trade from insured consumers.

Regarding our contributions to the literature on credence goods markets, we start by noting that there is only a small set of papers that theoretically address endogenous information acquisition in credence goods markets (Pesendorfer and Wolinsky, 2003; Dulleck and Kerschbamer, 2009; Bester and Dahm, 2018; Liu et al., 2019), and no controlled empirical evidence on this topic. Also, there is only a handful theoretical papers that analyze the role of diagnostic uncertainty on provision behavior (Inderst and Ottaviani 2012; Fong et al. 2022; Baumann and Rasch 2024), and no laboratory evidence on this issue. Our experiment closes those gaps.[4] Turning to the theoretical literature on the role of insurance, Sülzle and Wambach (2005) show that increasing the degree of coinsurance by the consumer can increase or decrease the likelihood of overcharging by the expert seller, depending on the parameters. There are several experimental studies on how insurance coverage affects sellers' and consumers' behavior in credence goods markets. In a laboratory experiment, Huck et al. (2016) find that insurance coverage lets consumers ask for more treatments, and sellers are more likely to overtreat consumers. Three field experiments confirm these patterns. Lu (2014) studies physicians' prescription behavior and reports that doctors write significantly more expensive prescriptions to insured patients. Kerschbamer et al. (2016) find that notebook owners with an insurance are charged about 80% more than non-insured owners for getting a notebook repaired, and Balafoutas et al. (2017) report a similar pattern in the market for taxi rides. However, none of these papers test predictions arising directly from a theoretical model that builds insurance into a credence goods market. Moreover, the literature has, so far, dealt with the effects of diagnostic uncertainty and insurance coverage in completely separate lines of research, thus ignoring how these two prominent factors might interact with each other in credence goods markets. Here, we opt for a treatment variation that includes both factors and analyze the behavior of sellers and consumers in an environment that is as comprehensive as possible.

The rest of the paper is organized as follows: Section 2 presents the main elements of the model, the experimental parameterization and implementation, and our predictions. Section 3 presents the experimental results and Section 4 concludes the paper.

2. Theoretical Framework and Experimental Implementation

2.1. Main Elements of the Model

Here we summarize the main elements of the model by Balafoutas et al. (2023) and the parameters we used to implement it in our experiment.

Each consumer (he) has either a major problem requiring a high-quality service (HQS) at cost $\overline{c} = 60$, or a minor problem requiring a low-quality service (LQS) at cost $\underline{c} = 20$. The

[4] Note that diagnostic uncertainty can be seen as creating moral wiggle room, because experts can use the excuse that the signal could be wrong, e.g., when receiving a low signal but providing the high-quality service. This aspect of our paper makes it somewhat related to the literature on moral wiggle room (see, e.g., Dana et al., 2007; Feiler, 2014; Grossman et al., 2017), even though the structure and richness of credence goods experiments is much different from the experimental designs used in the literature on moral wiggle room. We cannot provide evidence on the potential influence of moral wiggle room in our credence goods setting, as we do not have data on beliefs about the subjective probability of each state of the world, but mention it as potentially playing a role in expert motivation.

consumer knows that he has an *ex ante* probability $h = 0.4$ of having the major problem and a probability of $1 - h$ of having the minor one. The consumer derives utility $v = 150$ when his problem is solved through a service provided by the expert, and zero utility otherwise. While the HQS solves both problems, the LQS solves only the minor problem. The consumer can observe and verify the kind of service he receives, but he only finds out whether the received quality was the needed one when the expert provides the LQS for the major problem (since in that case his problem remains unsolved). In case a consumer does not receive a service, both he and the expert receive a reservation utility of $o = 15$.

Ex ante the expert (she) has the same information as the consumer on the severity of the consumer's problem. In contrast to the consumer, the expert is able to acquire additional information about the consumer's problem by performing a costly diagnosis that yields a signal about the consumer's problem. The signal is correct with probability σ. We consider two cases, one with exogenously given diagnostic precision (abbreviated as *EXO*) and one with endogenously determined precision (abbreviated as *ENDO*).

In **EXO100**, the expert receives a fully precise signal (σ =1.0) about the consumer's problem. In **EXO70**, the expert receives a signal that is 70% precise (σ =0.7), in the sense that it is correct with a probability of 0.7 and incorrect with a probability of 0.3. In **ENDO**, the expert can choose between six different precision levels σ (in 10% steps from 50% to 100%). The signal is costly for the expert (as capturing time and effort) in all cases and the cost follows the function $D(\sigma) = 40(\sigma - 0.5)^2$.[5] Hence, the price of diagnostic precision ranges from 0 for $\sigma = 0.5$ to 10 for $\sigma = 1$.[6]

If the expert does not acquire any information, she receives a completely uninformative signal (σ =0.5). In both cases (*EXO* and *ENDO*), the expert faces a penalty $t = 50$ whenever she prescribes the LQS to a consumer having the major problem. This payment is a transfer from the expert to the consumer as compensation for service failure.

We fix the price of the LQS across all price vectors at $\underline{p} = 60$. The price for the HQS can take on three different values, which generates three different types of price vectors. For $\overline{p} = 80$, we have what Balafoutas et al. (2023) call an undertreatment (UT) price vector. Here, the LQS

[5] In Balafoutas et al. (2023) the expert receives the diagnosis for free in the EXO environment. Assuming instead that the signal is costly in all cases does not change any of the derivations. We decided to make the signal costly here in order to keep the EXO and ENDO treatments more comparable.

[6] In EXO100 and EXO70, the cost of the corresponding signal is automatically subtracted from the expert's profit. In ENDO, the cost is subtracted only if the expert decides to invest in precision. Otherwise, without any investment, she receives an uninformative, costless signal of 50% precision. Note that consumers don't carry the expert's diagnosis costs in our setting.

markup (i.e., $\underline{p} - \underline{c}$) is larger than the HQS markup (i.e., $\overline{p} - \overline{c}$). Such a price vector creates material incentives for the seller to provide the low quality service. For $\overline{p} = 100$, we have an equal markup (EM) price vector, which makes the seller indifferent with respect to the markup between LQS and HQS. Finally, for $\overline{p} = 120$, we have an overtreatment (OT) price vector, which makes providing HQS more attractive for sellers.

Social preferences are also taken into account in the model and the experiment. We assume that the expert maximizes his own material payoff (weighted by one) plus λ times the consumer's surplus, $\lambda \in [0,1]$. A positive value of λ characterizes a prosocial expert, while $\lambda = 0$ implies that the expert is completely selfish.[7] We will explain how we measure social preferences in the experiment in subsection 2.4 below.

2.2. Predictions for the Experiment

There are three candidates for an efficient strategy by expert sellers. We define that as the strategy that minimizes the direct costs of providing a given service plus the implied utility loss for the case where the service fails:

Strategy A: *Implement the HQS independently of the outcome of the diagnosis*.

Strategy B: *Implement the LQS independently of the outcome of the diagnosis*.

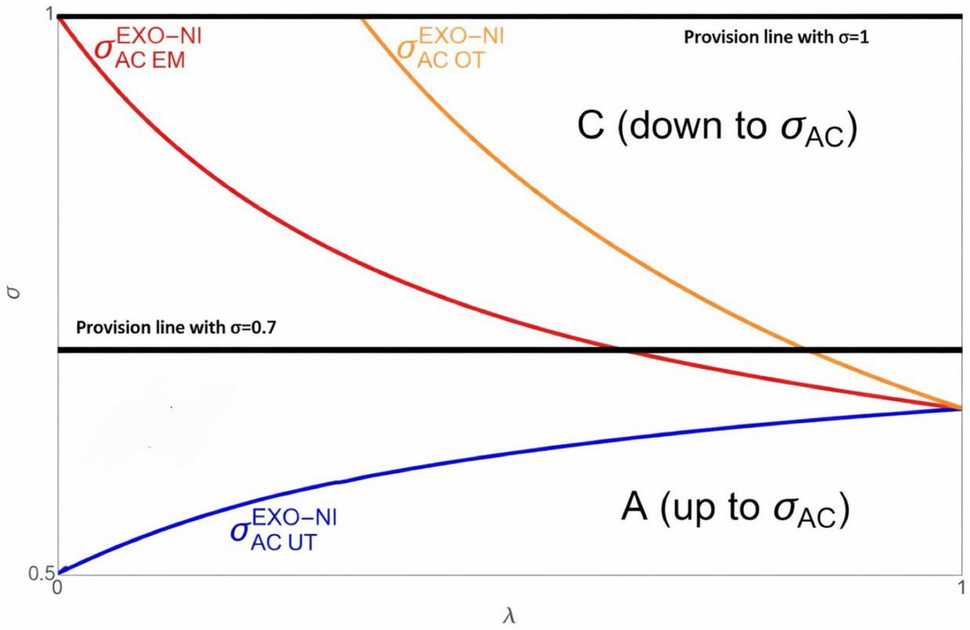
Strategy C: *Implement the LQS if the signal suggests that the problem is minor and implement the HQS if the signal suggests that the problem is major*.[8]

Intuitively, Strategy C is efficient if the diagnosis is sufficiently precise and if the likelihood of suffering from the major problem is neither close to 1 (in this case Strategy A is efficient) nor close to 0 (where Strategy B is efficient).

[7] Negative values of λ correspond to spiteful experts. Empirically, they are very rare (< 2%) in our experiment.

[8] There is a fourth possible pure strategy, in which the expert provides a treatment opposite to the signal she receives. We ignore this strategy in the remainder of the paper, since Balafoutas et al. (2023) show that it is dominated by one of the other three strategies for any given constellation of the parameters.

__Figure 1.__ Expert's provision in EXO-NI



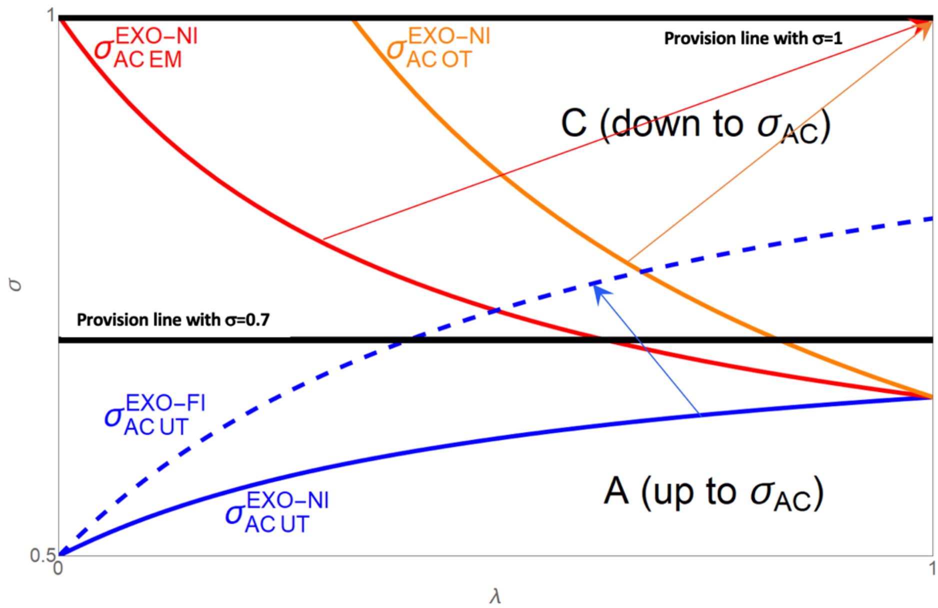
Note: The figure characterizes the predicted and the efficient provision strategy for the EXO treatments of the experiment when no insurance coverage is given (NI). The two black solid horizontal lines (at $\sigma = 1$ and at $\sigma = 0.7$) are the provision lines for the two precision levels implemented in EXO

Figure 1 characterizes the efficient and the predicted provision strategy for the EXO treatments for the case without insurance (NI). The blue line defines the provision areas for the UT price vector, the red line does the same for the EM price vector and the orange line stands for the OT price vector. Specifically, the blue line represents (λ, σ) combinations at which the expert is indifferent between Strategy A and Strategy C. At any point to the Northwest the expert strictly prefers C to A. The same interpretation applies to the red and orange line, just with different price vectors (EM and OT). The efficient provision strategy is the one chosen by the expert with $\lambda = 1$ – this expert maximizes the sum of the payoffs of the two parties meaning that the prices cancel out in her payoff function (the efficient strategy does not depend on the price vector, of course). The figure shows that efficiency demands to implement Strategy A for $\sigma < 0.65$ and Strategy C for higher precision levels. The two black horizontal lines (at $\sigma = 1$ and at $\sigma = 0.7$) are the provision lines for the two precision levels implemented in EXO. Since both precision levels exceed 0.65 efficiency demands to implement Strategy C in both cases.[9] As can be seen in the

[9] Note that the EXO70 case is quite close to the 0.65 threshold which has to be taken into account when interpreting the results.

figure, with the fully precise signal (upper black solid horizontal line) all expert types choose the efficient strategy (Strategy C) under the UT and the EM price vector, but not under the OT price vector. Under the latter, only fairly altruistic experts choose Strategy C, while selfish and moderately altruistic experts choose Strategy A. In the presence of diagnostic uncertainty (lower black solid horizontal line) the UT price vector still induces efficient provision for all expert types, but EM and OT induce most types (except for the very altruistic ones) to choose Strategy A.

Figure 2. *The impact of insurance in the EXO treaments*



Note: The figure characterizes the impact of introducing insurance in the EXO treatments. The solid lines are identical to those in Figure 1 and they represent the case when no insurance coverage is given (NI). The blue, red and orange arrows illustrate the effect of introducing full insurance (FI) under the UT, EM and OT price vectors, respectively.

In Figure 2 the blue, red and orange arrows illustrate the effect of introducing full insurance (FI) under the UT, EM and OT price vectors, respectively. In the full insurance case, all experts choose Strategy A under the EM and the OT price vector, while for the UT price vector the dashed blue line defines the provision areas (Strategy A below the line and Strategy C above the line).

We begin by characterizing the impact of diagnostic precision on the expert's provision strategy. We see that, for each price vector and for both insurance regimes (NI and FI), a larger range of λ values is in Area C on the upper ($\sigma = 1$) line than on the lower ($\sigma = 0.7$) line.[10] This leads to our first prediction:

[10] Without insurance the relation is strict for the EM and OT price vector, while with full insurance it is strict only for the UT price vector.

Prediction 1 (Impact of diagnostic precision on provision strategy in EXO): *Independently of the price vector and of whether insurance is present or absent, experts choose the efficient provision strategy more frequently under* $\sigma = 1$ *than under* $\sigma = 0.7$.

Next we look at the effect of introducing full insurance – see Figure 2. Comparing – for each price vector and for each provision line – the provision strategy under no insurance with the one under full insurance, we see that, in all comparisons, a strictly smaller range of λ values is in the C-area with than without insurance. This leads to the second prediction:

Prediction 2 (Impact of insurance on provision strategy in EXO): *Independently of the price vector and the level of diagnostic precision, experts choose the efficient provision strategy less frequently when full insurance is in place.*

The final comparison for the EXO case regards the impact of the prosociality parameter λ on the expert's provision strategy. Looking at Figure 2 we see that, without insurance, more prosocial experts are weakly more inclined to choose Strategy C than less prosocial ones. With full insurance, however, the relationship is reversed.[11] We thus conclude:

Prediction 3 (impact of prosocial motivation on provision strategy in EXO): *Independently of the price vector and of the level of diagnostic precision, without insurance more prosocial experts choose the efficient provision strategy more frequently than less prosocial ones, while with full insurance the relationship is reversed.*

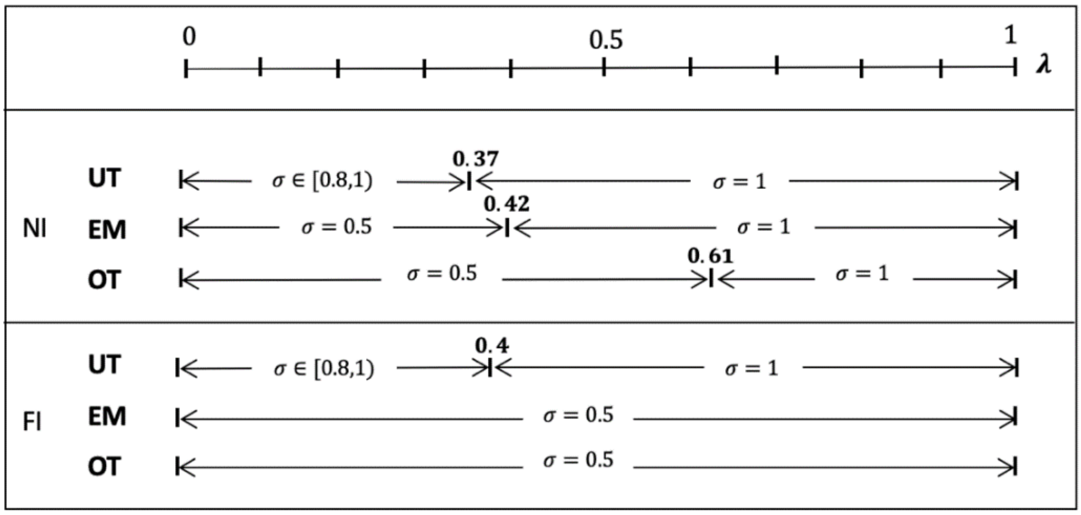
Figure 3 characterizes the expert's investment in information acquisition and her provision strategy for the ENDO treatments where sellers choose the costly precision level themselves. For the parameters implemented in the experiment, efficiency prescribes to acquire a fully precise signal ($\sigma = 1$) and to follow it. In Figure 3 we see that, in the case of no insurance (NI), altruistic experts (with a $\lambda > 0.61$) choose an efficient information acquisition for all three price vectors. This is no longer the case with full insurance (FI), where even very altruistic experts do not acquire information (and choose Strategy A) under the EM and the OT price vector. Comparing the two insurance regimes we see that, for all price vectors and all $\lambda \in [0,1]$ the investment in information acquisition is weakly lower with than without insurance.[12] This leads to the following prediction:

[11] Under NI the relationship is strict for the OT price vector under $\sigma = 1$ and for the OT and the EM price vector under $\sigma = 0.7$, while under FI the (reversed) relationship is strict for the UT price vector under $\sigma = 0.7$.

[12] Under the UT price vector the relation is strict for all $0.37 < \lambda < 0.4$, under the EM price vector it is strict for all $\lambda \geq 0.42$, and under the OT price vector it is strict for all $\lambda \geq 0.61$.

Prediction 4 (Impact of insurance on investment in diagnostic precision in ENDO): *Independently of the price vector, the average investment in diagnostic precision is lower when full insurance is in place.*

Figure 3. *Expert's investment in information acquisition in ENDO*



Note: The figure characterizes the expert's investment in information acquisition and her provision behavior in ENDO. For each insurance regime(NI or FI) and each price vector (UT, EM or OT), the figure characterizes the chosen precision level for each λ in $[0, 1]$. In constellations where $\sigma > 0.5$ the expert invests in diagnostic precision and then follows the signal (i.e., chooses Strategy C). In constellations where $\sigma = 0.5$ the expert abstains from investing and chooses Strategy A.

Our second prediction for the experts' behavior in the ENDO treatments regards the impact of the parameter λ on investment in information acquisition. Figure 3 reveals that, independently of the price vector and the insurance regime, investment in information acquisition is higher for more prosocial experts than for less prosocial ones.[13] We conclude from this:

Prediction 5 (Impact of prosocial motivation on investment in diagnostic precision in ENDO): *Independently of the price vector and of insurance being present or absent, the average investment in diagnostic precision is weakly increasing in the expert's prosociality.*

Finally, we also provide predictions for consumers' behavior. Since we focus mainly on experts' provision behavior, we just state the predictions for consumers in the main text here, but relegate a more extensive description and derivation to Appendix A.

[13] Under NI, the relationship is strict for all price vectors while under FI the relationship is strict only for the UT price vector.

Prediction 6 (factors affecting consumer's trading decision in EXO and ENDO):
(a) In EXO, consumers are more likely to trade on the market if the diagnostic precision is 100% rather than 70%. This prediction holds independently of the price vector and of whether insurance is in place.
(b) Both in EXO and in ENDO, consumers are more likely to trade on the market under full insurance than under no insurance. In both settings, this prediction holds independently of the price vector, and in EXO it also holds independently of the level of diagnostic precision.

2.3. Experimental Implementation

Online Part – Elicitation of Social Preferences

After subjects registered for an experimental session, they received a link for the online part of the experiment that they had to complete between 24 hours and one week before the lab session. In the online part we elicited social preferences by means of the Equality Equivalence Test, henceforth EET (Kerschbamer, 2015). The EET exposes subjects to two choice lists, one located in the domain of advantageous inequality and the other in the domain of disadvantageous inequality. The switching points of a subject in the two lists are used to infer the subject's prosociality, our parameter λ. This parameter is measured separately for each domain. For the domain of disadvantageous inequality, we denote the inferred λ as λ_D, and for the domain of advantageous inequality we denote it as λ_A. The parameters of the EET and the details of the calculation of λ are provided in Appendix B. Decisions were incentivized and participants were paid after the lab part had also been concluded. All experimental instructions are provided in Appendix C.

Lab Part – The Market Experiment and its Treatments

The experimental treatments are summarized in Table 1. We implement three conditions with respect to diagnostic precision in a between-subjects design, and two insurance conditions in a within-subjects design.

In the three **NI** (No Insurance) conditions, the price for the service is fully paid by the consumer if he trades on the market; in the three **FI** (Full Insurance) conditions, the price for the service is entirely covered by an insurance institution in return for a premium paid by the consumer. This premium is set to $P = 80$, which was calibrated based to ensure zero profits for the insurance institution in expectation.[14]

[14] To achieve this calibration, we ran one pilot session – that was never intended to be part of the main experiment – to find out at which premium one could expect zero profits for the insurance institution in expectation. Note that this

Table 1. *Summary of experimental design*

Treatment	Diagnostic precision	Insurance	Price vectors
EXO100	100%	12 periods with and 12 periods without insurance	8 periods for each of UT, EM, OT
EXO70	70%	12 periods with and 12 periods without insurance	8 periods for each of UT, EM, OT
ENDO	expert chooses among {50%, 60%, 70%, 80%, 90%, 100%}	12 periods with and 12 periods without insurance	8 periods for each of UT, EM, OT

The two insurance conditions are varied within subjects such that a group played 12 periods with insurance and 12 periods without, in blocks of six periods and with balanced ordering. Hence, within each of the three diagnostic precision treatments, we vary two insurance schemes (NI, FI) and three types of price vectors (UT, OT, and EM). This results in 6 possible combinations of insurance and price vector. We implemented each of these combinations four times within one session, yielding a total of 24 periods. We find no order effects with regards to the sequence of the different combinations; therefore, we pool all orders in the analysis. In addition, we vary the type of consumer problem within each sequence (minor or major, with 0.4 of cases being of the latter type).

The timeline and decision structure in the lab experiment were as follows. At the beginning of each period, both players were informed about the prevailing price vector and insurance condition. Each period then consisted of three different stages. In Stage 1, the consumer decided whether or not to trade on the market. At the same time the expert received a signal on the consumer's problem in EXO100 and EXO70, while in ENDO she decided how much to invest in diagnostic precision and afterwards received the signal with the chosen precision. In Stage 2, the consumer was inactive while the expert decided on the quality of the service without knowing whether the consumer accepted in Stage 1 (this decision was only implemented if the consumer had accepted to trade; so we applied a strategy method). In Stage 3, both players were informed about their own payoff.[15] From the second period onwards, experts and consumers could see the history of their own decisions and payoffs from previous periods, but had no information about their previous interaction partners. After the 24 periods of the market game, we elicited participants' risk attitudes using a validated survey measure (Dohmen et al., 2011), based on a

had to be done in a pilot in order to calibrate the premium based on human behavior, and not on the game-theoretic predictions.

[15] The consumer was not informed about the quality that was actually needed, but in case of undertreatment he could infer from the payoff that insufficient quality had been provided.

simple question that asked them to report their risk tolerance on a scale from 0 ("completely risk averse") to 10 ("completely risk-seeking"). Moreover, we collected data on their gender, age, study program, and highest educational degree.

We conducted our experiment in the Innsbruck EconLab with students enrolled at the University of Innsbruck. All experimental sessions were computerized with oTree (Chen et al., 2016) and we recruited subjects via H-ROOT (Bock et al., 2014). A total of 576 students participated in the experiment. The number of participants in a session was either 16 or 24. Given that the stage game is repeated for 24 periods, the matching protocol of participants is important. We used matching groups of eight participants, with four experts and four consumers (in fixed roles). Consumers and experts were randomly re-matched within a matching group after each period. This means that the entire interaction in a session over the course of the experiment took place within a matching group, while there were no interdependencies across groups. Accordingly, we use matching groups as independent observations in our statistical tests and cluster for matching groups in our regressions. Table 2 shows the number of participants, sessions, and matching groups per treatment.

Table 2. *Observations*

Treatment	Subjects	Observations	Sessions	Matching Groups
EXO100	200	4,800	9	25
EXO70	192	4,608	8	24
ENDO	184	4,416	8	23
Total	576	13,824	26	72

All parameters, the underlying treatment condition, as well as the matching procedure were made common knowledge to participants by reading them out aloud at the beginning of each session. The average session duration was 1.5 hours and all payoffs were stated in ECU (experimental currency units). Participants received 80 ECU as starting endowment, and earned on average a total of €18.34 (exchange rate 80 ECU = €1) from the online part and all 24 periods in the lab.

3. Results

Table 3 provides summary statistics about participants as well as for mean earnings, by treatment and overall. Importantly, randomization across the three treatments was successful in every dimension ($p > 0.05$ for all variables; χ^2 tests and Kruskal-Wallis tests).

Table 3. Summary statistics, by experimental treatment

	All	N	EXO100	N	EXO70	N	ENDO	N
Age (in years)	22.66 (3.51)	576	22.47 (3.26)	200	22.86 (4.00)	192	22.66 (3.22)	184
% Female	51.56	576	56.00	200	48.96	192	49.46	184
λ_D	0.10 (0.36)	247	0.08 (0.37)	86	0.05 (0.38)	83	0.17 (0.29)	78
λ_A	0.39 (0.36)	247	0.41 (0.39)	86	0.42 (0.37)	83	0.33 (0.31)	78
Risk Measure	5.59 (2.11)	528	5.33 (2.10)	176	5.72 (2.04)	192	5.72 (2.20)	160
Average payment experts in €	14.41 (1.39)	288	13.94 (1.29)	100	14.74 (1.25)	96	14.57 (1.51)	92
Average payment consumers in €	22.26 (2.35)	288	23.78 (1.83)	100	20.70 (1.92)	96	22.24 (2.19)	92

Notes: Means for all variables, except *Female*, which refers to the percentage of female subjects. *Risk Measure* is constructed from responses to a survey question on a scale from 0 "completely risk averse" to 10 "fully risk seeking" and is available for participants in all but one session. λ_D and λ_A are elicited in the EET and range from -5/6 to 5/6, with higher values indicating stronger prosociality. These variables are reported for experts and are missing for 41 subjects, for whom choices in the online part of the experiment could not be matched to the lab part. Standard deviations in parentheses.

3.1. Impact of Diagnostic Uncertainty and Insurance on Provision Policy (EXO)

We begin by testing predictions 1 and 2 for the EXO treatments. The main variable of interest is the rate of *efficient service provision*: This is the frequency (or, in the regressions, the likelihood) with which experts follow the signal they received.[16,17]

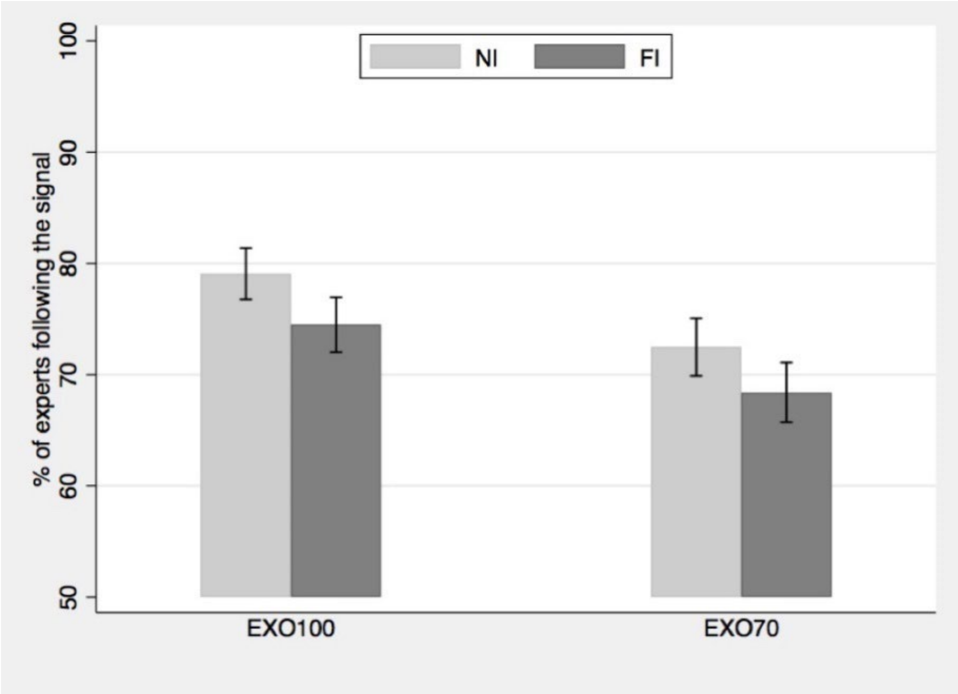
Figure 4 shows the rate of efficient service provision broken down by diagnostic precision and insurance coverage. The treatment differences support Predictions 1 and 2. In both insurance regimes (NI and FI), efficient service provision is more frequent in EXO100 than in EXO70 (pooled rates 76.8% vs. 70.4%, $p < 0.01$, χ^2 test). This result confirms the negative impact of

[16] Here note that following the signal for both possible signal realizations (Strategy C) is the efficient provision strategy for our parametrization. In the experiment we observe the quality choice of the expert for only one possible signal realization in each period. We therefore use in the experimental part the term 'rate of efficient service provision' (and do not refer to the frequency with which the efficient strategy is chosen). If Strategy C is adopted then the rate of efficient service provision is 1, while for Strategy A (Strategy B, respectively) it is 0.4 (0.6, respectively).

[17] In terms of our model, with the parameters in our experiment Strategy C is the efficient service strategy and Strategy A is inefficient or even fraudulent. This interpretation could be questioned as too narrow. In our model the expert is risk neutral. In the experiment the decision for Strategy A might reflect risk aversion on top of prosocial concerns of the expert, avoiding the risk of undertreatment for the client and penalty payment for the expert. This is especially true for the EXO70 case which is close to the border where Strategy A becomes the efficient strategy.

diagnostic uncertainty on the rate at which expert sellers follow the signal that they receive.[18] In line with Prediction 1, this impact is negative independently of whether insurance is present or absent, and independently of the price vector.[19] Turning to Prediction 2, the data confirm that the rate of efficient service provision declines when insurance is in place (FI) compared to when this is not the case (NI), dropping from 75.9% to 71.5% on aggregate ($p = 0.04$, χ^2 test). This effect is slightly more pronounced when diagnostic precision is high, but it generally does not retain its significance when disaggregated by the level of precision ($p = 0.08$ in EXO70; $p = 0.14$ in EXO70, χ^2 tests) – probably since the sample sizes become much smaller.

Figure 4. *Rate of efficient service provision, by treatment*



Note: All bars in the figure include 95% confidence intervals.

We also test predictions 1-3 in the multivariate OLS regressions shown in Table 4.[20] The dependent variable in the regressions is whether the expert followed the signal. We first provide a parsimonious specification with only price vector, insurance, and the EXO70 treatment dummy as independent variables. We then add a specification that further includes the interaction term between insurance and treatment, as well as a third one that includes additional control variables

[18] Remember, however, that the EXO70 case is close to the border where Strategy A becomes the efficient strategy.
[19] Differences in efficient provision rates across EXO100 and EXO70 are significant both under NI and FI ($p < 0.01$ for both comparisons, χ^2 tests). These differences are significant under UT and EM price vectors ($p < 0.01$, χ^2 tests), but not for OT vectors ($p = 0.18$).
[20] The regression results are robust to the use of Probit as the estimation method.

(period, experts' demographics, risk attitudes, and the two social preference parameters λ_D and λ_A). To test Prediction 3, in the last two columns we run the column (3) regressions separately for the treatments without insurance (NI) and with insurance (FI). All regressions in this section include subject random effects and report standard errors clustered at the matching group level.

Table 4. *Determinants of efficient service provision*

	(1) NI & FI	(2) NI & FI	(3) NI & FI	(4) NI	(5) FI
EXO70	-0.063***	-0.066***	-0.066***	-0.071***	-0.060**
	(0.019)	(0.021)	(0.022)	(0.025)	(0.026)
Insurance	-0.043***	-0.045***	-0.049***		
	(0.012)	(0.012)	(0.013)		
UT price vector	0.116***	0.116***	0.121***	0.088***	0.154***
	(0.017)	(0.017)	(0.019)	(0.024)	(0.026)
OT price vector	-0.081***	-0.081***	-0.068***	-0.111***	-0.029
	(0.021)	(0.021)	(0.021)	(0.029)	(0.024)
Insurance x EXO70		0.004			
		(0.024)			
Period			-0.002**	-0.001	-0.004**
			(0.001)	(0.001)	(0.002)
Risk measure			-0.007	-0.002	-0.013*
			(0.006)	(0.006)	(0.007)
λ_D			0.042	0.040	0.046
			(0.032)	(0.032)	(0.041)
λ_A			0.017	0.060*	-0.023
			(0.025)	(0.034)	(0.027)
Female			0.018	0.026	0.010
			(0.025)	(0.028)	(0.026)
Age			0.001	0.002	0.000
			(0.003)	(0.004)	(0.004)
N	4704	4704	3792	1896	1896

Notes: Dependent variable is *efficient service provision*, equal to 1 if the expert followed the received signal and 0 otherwise. The table presents estimates from OLS regressions with subject random effects. Standard errors are clustered at the matching group level and stated in parentheses. *EXO70* is equal to 1 for the EXO70 treatment. *Insurance* is equal to 1 if the consumer was insured in a particular period. *UT price vector* is equal to 1 if the price for the LQS is 60 and the price for the HQS is 80. *OT price vector* is equal to 1 if the price for the LQS is 60 and the price for the HQS is 120. The omitted benchmark is the EM vector with the price of 60 for LQS and of 100 for HQS. *Risk Measure* is constructed from responses to a survey question on a scale from 0 'fully risk averse' to 10 'fully risk seeking' and is available for participants in all but one session. λ_D and λ_A are elicited in the EET and range from -5/6 to 5/6, with higher values indicating stronger prosociality. These variables are reported for experts and are missing for 41 subjects, for whom choices in the online part of the experiment could not be matched to the lab part. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The regression results in the first three columns confirm predictions 1 and 2. The negative effects of diagnostic uncertainty and insurance coverage on efficient service provision are captured by the statistically and economically significant coefficients for *EXO70* and *Insurance*. Diagnostic uncertainty decreases the likelihood of efficient service provision by over 6 percentage points and insurance decreases it by up to 5 percentage points. These are average effects for each dimension, while the interaction between the two dimensions is insignificant in column (2).

For testing Prediction 3, we focus on λ_D because experts earn much less than consumers on average, which makes the degree of disadvantageous inequality salient to experts. We see in columns (4) and (5) of Table 4 no significant coefficients for λ_D, implying that we don't find support for Prediction 3.[21]

In terms of further explanatory variables, the regressions reveal a strong positive (negative) effect of UT (OT) price vectors on efficient service provision compared to the omitted category of EM prices, with effect sizes being around 7 to 8 percentage points for the OT vector and larger than 12 percentage points for the UT vector. All other control variables are insignificant, with the exception of *Period*, suggesting a slight downward time trend.

Result 1 (Provision behavior): *Predictions 1 and 2 are supported by the experimental data. Experts are more likely to choose the efficient service provision when diagnostic precision is higher, and they are less likely to do so when insurance is in place. We do not find support for Prediction 3 and the role of prosociality.*

3.2. Impact of Insurance on Investment in Diagnostic Precision (ENDO)

We now turn to predictions 4 and 5 regarding investment in diagnostic precision by experts in ENDO. In this treatment the expert chooses between six different levels of diagnostic precision. The distribution of choices is shown in Table 5 (along with the costs in ECU for each choice). Overall (i.e., pooling ENDO-NI and ENDO-FI), a precision level of 80% is both the median value and the modal choice of experts. At this level, diagnostic precision costs are 3.6 ECU. By contrast, efficiency would prescribe to acquire the 100% precise signal at the cost of 10 ECU.

Prediction 4 states that insurance coverage for consumers leads to a reduction in experts' investment in diagnostic precision. Indeed, comparing the distribution of chosen precision between the third and fourth column of Table 5 reveals that experts choose significantly lower precision levels when insurance is in place ($p = 0.04$, χ^2 test). Table 5 further reveals that a zero investment in precision is more frequent in the presence of insurance (20.7% vs. 16.1%, $p < 0.01$, χ^2 test).

[21] A more comprehensive test of the effects of prosociality on efficient service provision should consider the fact that these effects vary by price vector according to theory, as described in footnote 11. For this reason, we have run versions of the regressions in columns (4) and (5) of Table 4 separately for each price vector. This analysis reveals no systematic effect of the prosociality parameters – and especially λ_D – on service provision. This aligns with our conclusion that we have insufficient evidence to support Prediction 3.

Table 5. *Investment in diagnostic precision*

Diagnostic precision (in %)	Cost (in ECU)	ENDO-NI	ENDO-FI	Pooled
50	0	16.12	20.74	18.43
60	0.4	4.44	5.89	5.16
70	1.6	17.93	16.39	17.16
80	3.6	28.99	26.27	27.63
90	6.4	18.03	17.75	17.89
100	10	14.49	12.95	13.72

Notes: The table reports the relative frequencies (in %) with which the six different levels of diagnostic precision were chosen by experts.

Table 6. *Determinants of investment in diagnostic precision*

	(1)	(2)	(3)	(4)
Insurance	-0.133**	-0.148**	-0.288**	-0.327*
	(0.059)	(0.067)	(0.141)	(0.170)
UT price vector	0.163	0.195	0.237	0.273
	(0.115)	(0.120)	(0.275)	(0.309)
OT price vector	0.183***	0.186***	0.508***	0.525***
	(0.050)	(0.056)	(0.122)	(0.142)
Period		-0.023***		-0.046***
		(0.006)		(0.014)
Risk measure		-0.027		-0.084
		(0.050)		(0.120)
λ_D		0.533**		1.672***
		(0.240)		(0.629)
λ_A		0.265		1.004
		(0.334)		(0.840)
Female		-0.138		-0.415
		(0.219)		(0.571)
Age		0.025		0.053
		(0.020)		(0.046)
Constant			3.703***	3.055**
			(0.244)	(1.355)
N	2208	1584	2208	1584

Notes: Dependent variable is the number of invested ECU in diagnostic precision, ranging from 0 to 10. The table presents subject random effects regressions, using ordered Probit in columns (1) and (2) and Ordinary Least Squares in columns (3) and (4). Standard errors are clustered at the matching group level and stated in parentheses. The independent variables are defined as in Table 3. * p < 0.10, ** p < 0.05, *** p < 0.01.

The regressions in Table 6 provide further support for the effect of insurance on endogenously chosen diagnostic precision. The dependent variable in these regressions is the investment by experts into precision in ENDO, ranging from 0 to 10 ECU. The independent variables are the same as in the regressions of Table 4 (except for EXO70 and its interaction with insurance). The first two columns present results from ordered Probit regressions and the last two columns from OLS regressions. Regarding Prediction 4, the key aspect to note is the significantly negative coefficient on *Insurance*. Moreover, the significantly positive coefficients for the expert's λ_D-parameter shows that more prosocial experts in the domain of disadvantageous inequality invest more in a precise signal. This provides direct support for Prediction 5.[22]

Result 2 (Investment in diagnostic precision): *Predictions 4 and 5 are confirmed: Insurance leads to a reduction in experts' investments and to lower precision, while more prosocial experts invest more in diagnostic precision.*

3.3. Trading Decisions of Consumers

Rates of accepting to trade are generally very high, with a mean of 92.4% across all periods and treatments, but there is still a substantial amount of variation across conditions. The effect of diagnostic precision on the willingness to trade supports Prediction 6(a): A higher diagnostic precision leads to significantly higher trading rates (94.7% in the aggregate in EXO100 vs. 90% in EXO70, $p < 0.01$, χ^2-test). Prediction 6(b) on the role of insurance also finds support in the data: Under FI, trading rates are at 98% both in EXO and ENDO, while under NI they are 87% in EXO and 89% in EXO, with all differences significant at $p < 0.01$ (χ^2-tests).

The Probit regressions on willingness to trade shown in Table 7 report the results from two specifications for the exogenous treatments (differing in the number of control variables), followed by two specifications for treatment ENDO. Starting with the effects of diagnostic uncertainty in the exogenous treatments, the significant coefficient for EXO70 indicates that consumers are about 4 percentage points less likely to trade when the expert receives an imprecise exogenous signal about his problem, thus corroborating Prediction 6(a). Similarly, the data provide strong support for Prediction 6(b), since consumers are significantly more likely (by almost 10 percentage points) to trade on the market when they are insured, in the exogenous treatments as well as in ENDO.

Finally, we find that trade is most likely with the UT price vector, intermediate with the EM price vector, and lowest with the OT price vector.[23] This holds for all treatments. The differences between EM and OT price vectors are always significant ($p < 0.01$, χ^2-tests), while the differences between UT and EM price vectors are not. This is consistent with our theoretical framework: With our parametrization (especially given the penalty payment of 50) undertreatment

[22] The coefficient on λ_A is also positive, but insignificant. The fact that Prediction 5 is more strongly borne out in the domain of disadvantageous inequality (λ_D) is not unexpected, given that experts generally earn substantially less than consumers in all treatments (see Table 3).

[23] In the (pooled) EXO treatments, the rates are 96.8%, 95%, and 85.5% for UT, EM, and OT price vectors respectively, and for ENDO the corresponding rates are 98%, 96.6%, and 86.8%.

is never an issue. In the EXO-NI regime the UT price vector together with the penalty payment induces the expert to follow Strategy C independently of her prosociality and even if precision is low. By contrast, the EM and the OT price vector often yield overtreatment especially for rather selfish experts and for the low precision level (see Figure 1 for details). In addition, as we move from the UT to the EM and then to the OT price vector, the price for the HQS increases while the price for the LQS stays constant. Together, those facts imply that consumers are more likely to trade on the market under the UT than under the EM price vector, and more likely under the EM than the OT price vector.[24] The argument for the EXO-FI case and for the ENDO treatments is similar.

Table 7. *Determinants of willingness to trade*

	(1)	(2)	(4)	(5)
	EXO70 & EXO100	EXO70 & EXO100	ENDO	ENDO
EXO70	-0.041***	-0.043***		
	(0.010)	(0.011)		
Insurance	0.092***	0.085***	0.072***	0.057***
	(0.010)	(0.017)	(0.016)	(0.015)
UT price vector	0.017**	0.017**	0.014	0.005
	(0.008)	(0.008)	(0.009)	(0.009)
OT price vector	-0.075***	-0.074***	-0.080***	-0.075***
	(0.012)	(0.012)	(0.015)	(0.016)
Insurance x EXO70		0.012		
		(0.021)		
Period		0.002***		0.001***
		(0.001)		(0.001)
Risk measure		0.004*		0.006
		(0.002)		(0.004)
λ_D		0.012		-0.012
		(0.014)		(0.013)
λ_A		0.022		0.050***
		(0.015)		(0.016)
Female		-0.024**		0.012
		(0.010)		(0.012)
Age		0.000		0.004*
		(0.001)		(0.002)
N	4704	3936	2208	1608

Notes: Dependent variable is *Willingness to Trade*, equal to 1 if the consumer decided to trade and 0 otherwise. The table presents marginal effects estimates from Probit regressions with subject random effects. Standard errors are clustered at the matching group level and stated in parentheses. The independent variables are defined as in the notes to Table 4. * p < 0.10, ** p < 0.05, *** p < 0.01.

[24] While this is obvious for the high precision level, it also holds for the low precision level (see Appendix A for details).

Result 3 (willingness to trade): *Consumers' willingness to trade is higher when insurance is in place, regardless of whether diagnostic precision is exogenous or not. In the exogenous case, a higher precision increases the willingness to trade. Hence, the data fully support Predictions 6(a) and 6(b).*

3.4. Market Efficiency

As a final part of our results section, and to complete the picture, we examine what the behavior of experts and consumers implies for market efficiency. The efficiency index used in this analysis measures (in %) which fraction of the highest possible efficiency gain is achieved in an expert-consumer interaction. The highest possible efficiency gain is the difference between the *first-best joint payoff* and the *minimal joint payoff*. In the EXO treatments the first-best joint payoff is achieved if the consumer trades on the market and the expert follows the signal for both signal realizations, while in the ENDO treatments it is achieved if the consumer trades and the expert acquires the fully informative signal and follows the signal. In all treatments the minimal joint payoff is achieved if the consumer refuses interaction on the market.[25] The efficiency index is then calculated as follows: *Efficiency Index = (actual joint payoff – minimal joint payoff) / (first-best joint payoff – minimal joint payoff).*[26] Note that our normalized efficiency index accounts for the fact that there is a mechanical efficiency loss when moving from EXO100 to EXO70 due to the lower signal precision: While in EXO100 the first-best joint payoff is 104, it is only 92 in EXO70 (see footnote 255); thus, if the consumer trades on the market and the expert follows the signal, then the efficiency index is 1 even in EXO70.

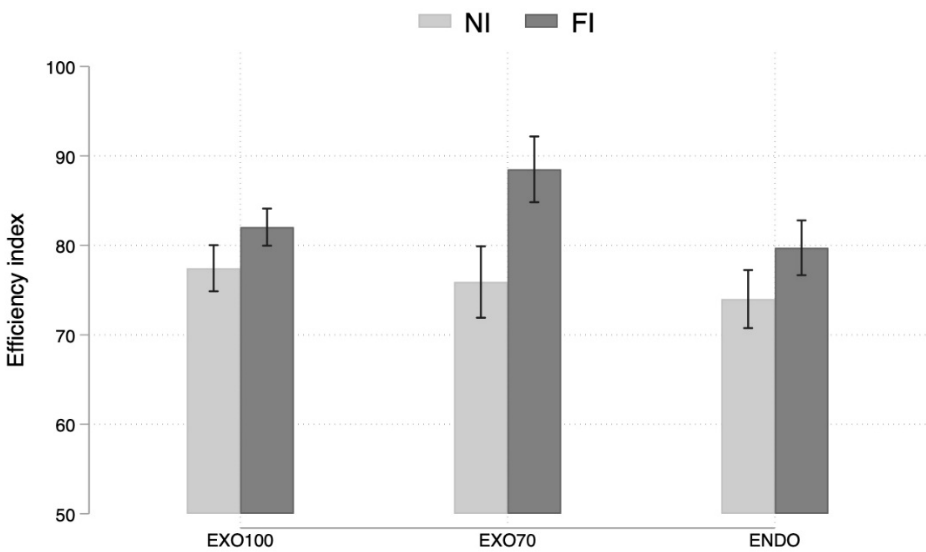
Figure 5 shows efficiency levels across treatments. The figure reveals that efficiency is always higher under full insurance (dark bars) than under no insurance (light bars). The difference is statistically significant for all three treatments pooled ($p < 0.01$, Wilcoxon signed-ranks test) and in the two exogenous treatments separately ($p < 0.01$ in EXO70, $p < 0.05$ in EXO100, Wilcoxon signed-ranks tests), while it is insignificant in ENDO ($p = 0.12$).[27]

[25] The first-best joint payoff equals $v - h\overline{c} - (1-h)\underline{c} - D(1) = 150 - 0.4 * 60 - 0.6 * 20 - 10 = 104$ in treatments EXO100 and ENDO, and $[1 - h(1-\sigma)]v - (1 - h - \sigma + 2h\sigma)\overline{c} - (h + \sigma - 2h\sigma)\underline{c} - D(\sigma) = 0.88 * 150 - 0.46 * 60 - 0.54 * 20 - 1.6 = 92$ in EXO70. The minimal joint payoff is 30 (the sum of the two outside payoffs in the case of no trade) in all treatments.

[26] Our definition of the actual joint payoff includes the transfers relating to insurance, i.e., the insurance premium paid by the consumer to the insurer and the price paid by the insurer (instead of the consumer) to the expert in the periods when full insurance is in place.

[27] Comparing efficiency across treatments is less meaningful, given the way our efficiency index is constructed.

Figure 5. *Efficiency rates by treatment and insurance*



Note: All bars in the figure include 95% confidence intervals.

Table 8 presents estimates from linear regressions with the efficiency index as dependent variable. In these regressions, each consumer-expert interaction is one data point, since efficiency is measured at the market interaction level. We separate between cases with exogenous (columns 1 and 2) and endogenous (column 3) diagnostic precision. This distinction allows us to cleanly identify the effects of diagnostic precision on efficiency, but also to include the chosen level of investment in column (3).

In line with the non-parametric test results, insurance has a sizeable and significant positive impact on the efficiency index (ranging from 5 to 9 percentage points) in the exogenous as well as in the endogenous treatments. This effect is stronger in the presence of diagnostic uncertainty, as evinced by the positive and significant interaction term in (2). This analysis suggests that the drop in efficient service provision as a result of insurance documented in section 5.1 is outweighed by the increase in the frequency of trade documented in section 5.3, leading to a beneficial net impact of insurance coverage on efficiency.

In treatment ENDO, experts' investments into diagnostic precision have a highly significant positive effect on efficiency. The intuition here is that experts who have achieved a higher precision are more likely to follow the signal and, thus, more likely to provide the correct quality. We further document an efficiency-diminishing effect of OT price vectors, which is of considerable magnitude. This finding is aligned with the way prices affect the behavior of experts in the exogenous treatments, with OT price vectors yielding lower rates of efficient service provision than EM price vectors.

Table 8. *Regressions on market efficiency*

	(1)	(2)	(3)
Treatment(s)	EXO100 & EXO70	EXO100 & EXO70	ENDO
Insurance	0.086***	0.048***	0.062**
	(0.018)	(0.018)	(0.029)
EXO70	0.025	-0.015	
	(0.017)	(0.029)	
UT price vector	0.020	0.020	-0.020
	(0.018)	(0.018)	(0.032)
OT price vector	-0.117***	-0.116***	-0.076***
	(0.018)	(0.018)	(0.022)
Insurance x EXO70		0.078**	
		(0.035)	
Investment			0.019***
			(0.004)
Period	0.003***	0.003***	0.004**
	(0.001)	(0.001)	(0.002)
constant	0.745***	0.765***	0.650***
	(0.025)	(0.026)	(0.031)
N	4704	4704	2208

Notes: Dependent variable: *Efficiency Index*. The table presents random effects OLS regressions. Standard errors are clustered at the matching group level and stated in parentheses. The independent variables are defined as in the notes to Table 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4. Concluding Remarks

In a laboratory experiment conducted with 576 participants we have tested how diagnostic uncertainty and insurance coverage, as well as their interaction, affect behavior on credence goods markets. Based on the theoretical model of Balafoutas et al. (2023), our experimental implementation has confirmed systematic effects that have gone unnoticed in previous work. For the parameter constellations used in the experiment, for which efficiency requires expert sellers to follow the signal they receive about a consumer's needs, lower diagnostic precision leads to less efficient provision. Insurance has the same predicted effect of reducing efficient service provision, and in addition it leads to a reduction in experts' investment in diagnosis and therefore to lower precision (when this is endogenous). This latter finding has been obtained in an environment where insurance does not reimburse diagnostic effort, and its implications are therefore also bound to this kind of environment. Finally, the willingness of consumers to trade on the market increases with diagnostic precision and insurance, implying that higher precision as well as insurance increase efficiency.

We hope that our findings – although we would not want to draw too strong conclusions from a laboratory experiment – can provide some evidence-based input into the policy debate about the organization and potential regulation of credence goods markets. For instance, our finding that higher diagnostic precision increases market efficiency through reducing incentives for over- or undertreatment suggests that technological progress to improve diagnostic precision is socially welcome from an efficiency point of view. Seen from this angle, there are additional and hitherto overlooked benefits when advances in medical science make diagnoses less error-prone (Gottschalk et al., 2020) or when websites help consumers to identify their needs (Kerschbamer et al., 2023). Diagnostic precision is also linked to insurance coverage, but this relationship has never been investigated before. Importantly, insurance coverage provides negative incentives for diagnostic precision, implying that regulatory policies that introduce a requirement to buy insurance may have unintended side effects on experts' effort provision to diagnose consumers' problems properly (at least in a coverage environment where experts bear the effort costs of diagnosis themselves, as in our setting). For this reason, an important avenue for future research may be to look more deeply into what happens when consumers can only get partial insurance – meaning that they have to pay deductibles even when being insured – or when they can endogenously choose the extent of insurance. It would also be interesting to study whether markets may split up in consumers who ask for insurance while others remain (voluntarily) uninsured, and how experts would respond with their diagnosis effort and tailor their offers to both types of consumers.

References

Balafoutas, L., Fornwagner, H., Kerschbamer, R., Sutter, M. and Tverdostup, M. (2023) 'Diagnostic uncertainty and insurance coverage in credence goods markets', Working Papers in Economics and Statistics, 2023-14, University of Innsbruck.

Balafoutas, L. and Kerschbamer, R. (2020) 'Credence goods in the literature: What the past fifteen years have taught us about fraud, incentives, and the role of institutions', *Journal of Behavioral and Experimental Finance*, *26*, 100285.

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

Baumann, F. and Rasch, A. (2024) 'Second opinions and diagnostic uncertainty in expert markets', *Journal of Institutional and Theoretical Economics*, 180(1), 74-105.

Bester, H. and Dahm, M. (2018) 'Credence goods, costly diagnosis and subjective evaluation', *The Economic Journal*, 128(611), 1367–1394.

Bock, O., Baetge, I. and Nicklisch, A. (2014) 'Hroot: Hamburg registration and organization online Tool', *European Economic Review*, 71, 117–120.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2016) 'Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision', *Journal of Economic Behavior and Organization*, 131, 17–23.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N. and Wiesen, D. (2017) 'The effects of introducing mixed payment systems for physicians: Experimental evidence', *Health Economics*, 26(2), 243–262.

Chen, D. L., Schonger, M. and Wickens, C. (2016) 'oTree-An open-source platform for laboratory, online, and field experiments', *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Dana, J., Weber, R. A., & Kuang, J. X. (2007) 'Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness', *Economic Theory*, 33, 67-80.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011) 'Individual risk attitudes: Measurement, determinants, and behavioral consequences', *Journal of the European Economic Association*, 9(3), 522-550.

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

Dulleck, U. and Kerschbamer, R. (2009) 'Experts vs. discounters: Consumer free-riding and experts withholding advice in markets for credence goods', *International Journal of Industrial Organization*, 27(1), 15–23.

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

Feiler, L. (2014) 'Testing models of information avoidance with binary choice dictator games', *Journal of Economic Psychology*, 45, 253-267.

Fong, Y. F., Liu, T. and Wright, D. J. (2014) 'On the role of verifiability and commitment in credence goods markets', *International Journal of Industrial Organization*, 37, 118–129.

Fong, Y.-F., X. Meng, and Zhao, L. (2022) 'Laboratory Kickback to Doctor, Consumer Awareness, and Welfare Implications', Available at SSRN: https://ssrn.com/abstract=4246596 or http://dx.doi.org/10.2139/ssrn.4246596

Gottschalk, F., Mimra, W. and Waibel, C. (2020) 'Health services as credence goods: A field experiment', *The Economic Journal*, 130(629), 1346–1383.

Grossman, Z., & Van Der Weele, J. J. (2017) 'Self-image and willful ignorance in social decisions', *Journal of the European Economic Association*, 15(1), 173-217.

Huck, S., Lünser, G., Spitzer, F. and Tyran, J. R. (2016) 'Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment', *Journal of Economic Behavior and Organization*, 131, 78–105.

Inderst, R., and Ottaviani, M. (2012) 'Competition through commissions and kickbacks', *American Economic Review*, 102(2), 780-809.

Kerschbamer, R. (2015) 'The geometry of distributional preferences and a non-parametric identification approach: The Equality Equivalence Test', *European Economic Review*, 76, 85–103.

Kerschbamer, R., Neururer, D. and Sutter, M. (2016) 'Insurance coverage of customers induces dishonesty of sellers in markets for credence goods', *Proceedings of the National Academy of Sciences*, 113(27), 7454–7458.

Kerschbamer, R., Neururer, D. and Sutter, M. (2023) 'Credence goods markets, online information and repair prices: A natural field experiment', *Journal of Public Economics*, 222, 104891.

Kerschbamer, R., Sutter, M. and Dulleck, U. (2017) 'How social preferences shape incentives in (experimental) markets for credence goods', *Economic Journal*, 127(600), 393–416.

Liu, T. and Ma, C. (2013) 'Health insurance, treatment plan, and delegation to altruistic physician', *Journal of Economic Behavior and Organization*, 85, 79–96.

Liu, F., Rasch, A., Schwarz, M. and Waibel, C. (2019) 'The role of diagnostic ability in markets for expert services', Available at SSRN 3491607.

Lu, F. (2014) 'Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in China', *Journal of Development Economics*, 106, 156–167.

Lyu, H., Xu, T., Brotman, D., Mayer-Blackwell, B., Cooper, M., Daniel, M., Wick, E., Saini, V., Brownlee, S. and Makary, M. (2017) 'Overtreatment in the United States', *PLoS One* 12.9: e0181970.

Pesendorfer, W. and Wolinsky, A. (2003) 'Second opinions and price competition: Inefficiency in the market for expert advice', *Review of Economic Studies*, 70(2), 417–437.

Schneider, H. S. (2012) 'Agency problems and reputation in expert services: evidence from auto repair', *Journal of Industrial Economics*, 60(3), 406–433.

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

Online Appendix

Appendix A. Details on the predictions for consumer behavior

For a given environment (EXO100, EXO70 or ENDO), given insurance regime (NI vs FI) and given price vector (UT, EM or OT) the consumer can infer for each $\lambda \in [0,1]$ the expert's investment in diagnostic precision (in the ENDO environment) and her provision behavior (in all environments). He then uses his prior on λ to decide whether trade on the market is profitable. In this decision he considers that – for a given strategy of the expert – his payoff depends on the precision level, the insurance regime and the price vector. That is, each of those three variables influences both, the behavior of the expert and – for a given behavior of the expert – the payoff of the consumer. Let us start with the payoffs for a given strategy of the expert.

In the EXO-NI environment the payoff of the consumer under Strategy A is $v - \overline{p} = 150 - \overline{p}$ for both diagnostic precision levels, while his payoff under Strategy C is $v - (1-h)\underline{p} - h\overline{p} = 150 - 0.6\underline{p} - 0.4\overline{p}$ for $\sigma = 1$ and $[1 - h(1-\sigma)]v + h(1-\sigma)t - (h + \sigma - 2h\sigma)\underline{p} - (1 - h - \sigma + 2h\sigma)\overline{p} = 138 - 0.54\underline{p} - 0.46\overline{p}$ for $\sigma = 0.7$. Thus, since $\overline{p} > \underline{p}$ for any given price vector, the consumer unambiguously prefers Strategy C over Strategy A if the signal is perfect; however, under $\sigma = 0.7$ the consumer prefers Strategy C over Strategy A only if $0.54\left(\overline{p} - \underline{p}\right) \geq 12$. Given our parametrization, this latter inequality is satisfied for the EM and the OT price vector, but not for the UT vector.

We now move on to the EXO-FI environment. In this environment, the payoff of the consumer under Strategy A is $v = 150$ for both diagnostic precision levels, while his payoff under Strategy C is $v = 150$ for $\sigma = 1$ but only $[1 - h(1-\sigma)]v + h(1-\sigma)t - (h + \sigma - 2h\sigma)\underline{p} = 138$ for $\sigma = 0.7$. Thus, the consumer is indifferent between Strategy A and Strategy C if the signal is perfect but strictly prefers A over C if the signal is imperfect. What can we conclude from this?

Let us first consider the impact of diagnostic precision in the EXO-NI environment. With OT and EM prices there are two effects to consider: First, the range of λ values for which Strategy C is implemented is narrower when the precision is $\sigma = 0.7$ instead of $\sigma = 1$ (see Figure 1 for details). Second, the payoff under strategy C is lower under $\sigma = 0.7$ than under $\sigma = 1$, while the payoff under Strategy A does not depend on σ but is lower than the payoff under C under both precision levels. Together, these facts imply that the consumer is unambiguously better off under $\sigma = 1$ compared to $\sigma = 0.7$. Still looking at the EXO-NI environment, consider now the case of

UT prices. Here all experts always provide Strategy C under both precision levels and the consumer's payoff under C is lower under $\sigma = 0.7$ than under $\sigma = 1$. So, again, consumers are more likely to trade on the market if the diagnostic precision is 100% rather than 70%.

Turning to the EXO-FI regime, we observe (in Figure 1) that with OT and EM prices all experts provide Strategy A under both precision levels. Furthermore, the consumer's payoff under these price vectors does not depend on the precision level. Thus, consumers are indifferent between the two precision levels. Under UT prices all experts provide Strategy C under $\sigma = 1,$ while some experts implement A and others implement C under $\sigma = 0.7.$ In terms of payoffs, this implies that consumers receive a payoff of 150 under full precision, while under $\sigma = 0.7$ they reach that payoff level only if Strategy A is implemented and they end up with 138 under C. This means that consumers are again more likely to trade on the market if the diagnostic precision is 100% rather than 70%. Using similar arguments for the impact of insurance, and also considering the case where the diagnostic precision is endogenous, we arrive predictions 6a and 6b as stated in the main text.

The model also allows to predict how prices will affect market entry. First consider the EXO environment with $\sigma = 1.$ This case yields a clear prediction: As we move from the UT to the EM and then to the OT price vector, the range of λ values for which Strategy C is provided decreases and at the same time $\overline{p}$ increases. This is unambiguously bad news for the consumer. However, when moving from the UT to the EM price vector for $\sigma = 0.7,$ we have two opposing effects. On the one hand, the range of λ values for which Strategy C is implemented decreases, which is now good news for the consumer. On the other hand, the price $\overline{p}$ increases, which is again bad news for him. Which of the two price vectors is preferred by the consumer depends on a comparison between his payoff from Strategy C under the UT price vector and his payoff from Strategy A under the EM price vector. This comparison shows that the consumer strictly prefers Strategy C under the UT price vector over Strategy A under the EM price vector, since $138 - [0.54] \cdot 60 - [0.46] \cdot 80 > 150 - 100.$

Using similar arguments for the case where the diagnostic precision is endogenous, we arrive at the prediction that both in EXO and in ENDO, consumers are more likely to trade on the market under the UT than under the EM price vector, and more likely under the EM than the OT price vector. In both settings this prediction holds independently of whether insurance is in place, and in EXO it also holds independently of the level of diagnostic precision.

Appendix B. Parameters of the EET and translation of EET results in parameter values

Parameters in the Equality Equivalence Test

Advantageous Inequality Block

LEFT		RIGHT	
decision maker's payoff	passive person's payoff	decision maker's payoff	passive person's payoff
8 ECU	7 ECU	10 ECU	10 ECU
9 ECU	7 ECU	10 ECU	10 ECU
10 ECU	7 ECU	10 ECU	10 ECU
11 ECU	7 ECU	10 ECU	10 ECU
12 ECU	7 ECU	10 ECU	10 ECU

Disadvantageous Inequality Block

LEFT		RIGHT	
decision maker's payoff	passive person's payoff	decision maker's payoff	passive person's payoff
8 ECU	13 ECU	10 ECU	10 ECU
9 ECU	13 ECU	10 ECU	10 ECU
10 ECU	13 ECU	10 ECU	10 ECU
11 ECU	13 ECU	10 ECU	10 ECU
12 ECU	13 ECU	10 ECU	10 ECU

Note: Decision makers had to indicate, for each of the ten decision situations, whether they wanted to implement the LEFT or the RIGHT allocation

Translation of Decisions in the Equality Equivalence Test into λ Values

Denote the own material payoff by m and the other's payoff by o. Suppose subjects in the lab decide in line with the utility function $u(m,o) = m + \lambda o$, as assumed for experts in our model. What can we infer about λ from their choices in the EET?

Advantageous Inequality Block (Y-List)

subject chooses LEFT for the first time in row	inference regarding λ_A	attributed λ_A
1 (always left)	$\lambda_A \leq -2/3$	$\lambda_A = -5/6$
2	$-2/3 \leq \lambda_A \leq -1/3$	$\lambda_A = -3/6$
3	$-1/3 \leq \lambda_A \leq 0$	$\lambda_A = -1/6$
4	$0 \leq \lambda_A \leq 1/3$	$\lambda_A = 1/6$
5	$1/3 \leq \lambda_A \leq 2/3$	$\lambda_A = 3/6$
never (always right)	$2/3 \leq \lambda_A$	$\lambda_A = 5/6$

Disadvantageous Inequality Block (X-List)

subject chooses LEFT for the first time in row	inference regarding λ_D	attributed λ_D
1 (always left)	$2/3 \leq \lambda_D$	$\lambda_D = 5/6$
2	$1/3 \leq \lambda_D \leq 2/3$	$\lambda_D = 3/6$
3	$0 \leq \lambda_D \leq 1/3$	$\lambda_D = 1/6$
4	$-1/3 \leq \lambda_D \leq 0$	$\lambda_D = -1/6$
5	$-2/3 \leq \lambda_D \leq -1/3$	$\lambda_D = -3/6$
never (always right)	$\lambda_D \leq -2/3$	$\lambda_D = -5/6$

Appendix C. Experimental Instructions

The following sections contain the instructions translated from German, which were used in the online and lab part of the experiment. We read out all instructions out loud in the lab. If instructions were different between treatments, this is indicated by squared brackets, including the respective treatment condition and the according text in italic letters. To match the observations from the online and the lab part, participants created a unique in the online part and had to state this ID at the beginning of the lab part.

B.1. Instructions for Online Part

The first part of the experiment will be conducted online and consists of 10 decisions. In each of these 10 decisions, the computer randomly assigns another participant to you. In the following, we will call this participant "Your Passive Person". You will never learn the identity of your passive person. You will see below why we call this person a "passive person".

We state the payoffs in ECU (experimental currency units).
The exchange rate is 5 ECU = 1 Euro

Each of your 10 choices is a choice between the LEFT and the RIGHT option. Each option has consequences for your payoff and the payoff of your passive person.

Example:

You will be asked if you prefer to choose the option LEFT, where you will receive 8 ECU and your passive person 13 ECU, or option RIGHT, where you will receive 10 ECU. Your passive person will also receive 10 ECU. You have to choose one of the two options by clicking on the corresponding circle.

This decision problem would be presented on the screen as follows:

Your decision

Option 'Left' **Option 'Right'**

Your Payoff	*Payoff passive person*			*Your Payoff*	*Payoff passive person*
8,00 ECU	13,00 ECU	O	O	10,00 ECU	10,00 ECU

You have to make a total of 10 decisions (5 are shown per screen). Your total earnings from this part are determined as follows:

Payout as Active Person: In the end, one out of the 10 decision-making situations is selected individually and randomly for each participant, and the option chosen in this decision situation is then actually paid out. E.g., if you have chosen the decision situation described above and you have decided on the option RIGHT in this decision situation, you would receive 10 ECU as an active person. In contrast, your passive person would receive 10 ECU as a passive person.

Payout as Passive Person: Just as your Passive Person receives ECU from your decision without doing anything, you receive ECU from another participant without doing anything. This means you are the passive person for this other participant. It is assured that you will not be redeemed twice as an active and passive person. That is, if person X is your passive person, then you are certainly not the passive person of person X.

You will see 5 of all 10 decisions on one screen. You can make corrections to your choices as long as you have not clicked "Next".

After the 10 decisions, this part of the experiment ends. Which of your 10 decisions is relevant for your payoff and how much you have earned, you will get to know in the EconLab, when you participate in the lab part. Also, this online part will be paid at the EconLab in cash.

B.1. Instructions for Lab Part

Instructions

Thank you for your participation in the experiment. Please do not talk to other participants until the end of the experiment.

In this experiment, the payoffs are stated in ECU (experimental currency units).
The exchange rate is 80 ECU = 1 €.

Different roles:

There are two roles in the experiment: client and expert. At the beginning of the experiment, you will be randomly assigned one of these roles and keep that role for the entire experiment, which means for all periods. At the first decision screen (period 1 of 24), you will see your role.

In each period, one expert is randomly assigned to one client (and vice versa). It is ensured in each period that the same client-expert pair is never formed in two consecutive periods. This means that a client always interacts with an expert, and you will randomly get a partner in each period.

24 Periods

This experiment consists of 24 periods, each with the same sequence of decisions described below.

Suppose you have been assigned a client's role. In that case, you have a big problem with a probability of 40% and a small problem with a probability of 60%. As a client, you will never be informed about the problem you actually have. As a client, you can decide if you enter the market, which means you want to interact with the expert randomly assigned to you. By entering the market, your problem may be solved by the expert. If your problem is solved, you will receive 150 ECU in this period. If your problem is not solved, the expert has to pay you a compensation of 50 ECU (more on this below).

[EXO100]

- If you have been assigned the role of an expert, you will receive in each period a signal about what problem your client (randomly matched with you) has. This signal is 100% accurate. That is, if e.g., the client has a small problem, then, with 100% probability, you get the signal that the problem is small; and similar to the big problem. However, this information costs 10 ECU. These costs will be deducted automatically in each period.

[EXO70]

- If you have been assigned the role of an expert, you will receive in each period a signal about what problem your client (randomly matched with you) has. This signal is 70% accurate. If e.g., the client has a small problem, then, with 70% probability, you get the signal that the problem is small; and similar to the big problem. However, this information costs 1.6 ECU. These costs will be deducted automatically in each period.

[ENDO]

If you have been assigned an expert's role, you will receive a signal each period about the client's problem randomly. As an expert, you decide on the precision of the received signal. The costs associated with each precision level are given in the following table:

Precision of the signal	*Costs*
50%	*0 ECU*
60%	*0.4 ECU*
70%	*1.6 ECU*
80%	*3.6 ECU*
90%	*6.4 ECU*
100%	*10 ECU*

For example, suppose you choose a 70% precise signal. In that case, you pay 1.6 ECU and get a signal that identifies the problem correctly with a 70% probability. If you select a 50% precision, you will not be charged. However, your signal is not informative (because it identifies the actual problem with the same probability as right or wrong).

As an expert, your job is to treat the client. You can choose between a low quality treatment and a high quality treatment. The low quality treatment only solves the small problem of the client. The high quality treatment solves both the small and the big problem. You always have to pay ECU 20 for a low quality treatment and 60 ECU for a high quality treatment. If your treatment has not solved the client's problem, then as an expert, you have to pay a compensation fee of 50 ECU.

For the treatment of low quality, you will receive from the client a price of 60 ECU. For high quality treatment, the price is either 80 ECU, 100 ECU, or 120 ECU: this price varies over periods, and you will find out the respective price at the beginning of each period.

- Treatment can only be carried out if the client has decided to enter the market. If the client chooses not to enter the market, that is, if he /she decides not to interact with the expert, both client and expert receive a payoff of 15 ECU in this period.

Overview of the Decisions of one Period

Therefore, each period consists of decisions of the expert and the client, which are made simultaneously and independently of each other.

Client:	**Expert:**
The client decides if he/she wants to enter the market.	*[ENDO]* *1. The expert decides on the signal precision.* *2. The expert receives the signal about the client's problem and decides if he/she wants to use the high or low quality treatment.* *[EXO100] and [EXO70]* *The expert receives the signal about the client's problem and decides whether he /she wants to implement the high or low quality treatment.*

Only after both made their decisions it is announced whether the client has entered the market or not. If the client has not entered, the expert's decisions are irrelevant. In summary, this leads to the following payoffs:

Payoffs

CLIENT DOES NOT ENTER THE MARKET: both the client and the expert receive 15 ECU.

CLIENT ENTERS THE MARKET:

Client:

If the problem was solved: 150 ECU minus the treatment price

If the problem was NOT solved: 50 ECU minus the treatment price

Expert:

[EXO70] If the problem was solved: treatment price minus treatment cost minus 1,6 (costs diagnostic precision 70%)

[EXO100] If the problem was solved: treatment price minus treatment cost minus 10 (costs diagnostic precision 100%)

[EXO70] If the problem was not solved: treatment price minus treatment cost minus 50 ECU (compensation) minus 1,6 (costs diagnostic precision 70%)

[EXO100] If the problem was not solved: treatment price minus treatment cost minus 50 ECU (compensation) minus 10 (costs diagnostic precision 100%)

[ENDO]

If the problem was solved: treatment price minus treatment cost minus the cost for the selected signal precision level

If the problem was not solved: treatment price minus treatment cost minus 50 ECU (compensation) minus cost for the selected signal precision level

Insurance

Please note that there are two possible market situations in each period: either the client is insured or not. You play a total of 12 periods with and 12 periods without insurance, and you will see at the beginning of each period on the screen if the particular period is a period with insurance or not.

In a period without insurance, the situation is the same as previously described. In a period with insurance, the client is insured. This means that at the beginning of such a period, the client pays an insurance premium of 80 ECU (deducted automatically, even if the client does not enter the market). In return, the price for the client's treatment is covered by the insurance. All decisions taken by the client and the expert remain the same as in periods without insurance. The payoffs of the two parties in periods with insurance are as follows:

Payoffs in periods with insurance:

Client:

CLIENT DOES NOT ENTER THE MARKET: 15 ECU minus 80 ECU (insurance premium) = -65 ECU

CLIENT ENTERS THE MARKET:

If the problem has been solved: 150 ECU minus 80 ECU (insurance premium) = 70 ECU

If the problem was NOT solved: 50 ECU minus 80 ECU (insurance premium) = -30 ECU

Expert:

For the expert, the payoff does not change. It is calculated as in periods without insurance.

Information and Feedback

In each period, the expert and the client get to know whether the client is insured in this period and the prices for the low and high quality. The expert also receives the signal about the problem of the client. After the second period, you can see the results of all past periods at the bottom of your screen.

Total Payoff

At the beginning of the experiment, you will receive an endowment of 80 ECU. From this endowment, you can also cover possible losses in individual periods. Profits from other periods also compensate for losses.

For the final payoff, each period's endowment and the payoffs are added together and paid in cash at the end of the experiment, using the exchange rate 80 ECU = 1 €. Also, you will receive your payoff from the online part.