Journal of Behavioral and Experimental Economics 103 (2023) 101973

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Framing and subject pool effects in healthcare credence goods

Silvia Angerer [a,*], Daniela Glätzle-Rützler [b], Christian Waibel [c]

[a] UMIT TIROL – Private University for Health Sciences and Health Technology, Institute for Management and Economics in Healthcare, Eduard-Wallnöfer-Zentrum 1, Hall in Tirol 6060, Austria
[b] University of Innsbruck, Universitätsstrasse 15, Innsbruck 6020, Austria
[c] ETH Zurich, Zuerichbergstrasse 18, Zurich 8092, Switzerland

ARTICLE INFO

JEL-Codes:
C91
D82
I11

Keywords:
Healthcare economics
Trust
Fraud
Framing effects
Experts
Credence goods
Undertreatment
Overcharging
Laboratory experiment

ABSTRACT

Credence goods markets are characterized by asymmetric information between experts and consumers, specifically relating to quality (what is needed and/or supplied). The functioning of the market is heavily reliant on trust on the side of the consumer and trustworthiness on the side of the expert. However, a large body of empirical and experimental research has demonstrated the existence of overtreatment, undertreatment, and overcharging in several credence goods markets. In this paper, we study two determinants of trust and trustworthiness in experimental credence goods markets: the effects of a health frame (vs. a neutral frame) and an expert subject pool (vs. standard student subject pool). Our results reveal that the subject pool, in combination with the health frame, has a significant impact on the level of trust and trustworthiness, shown by a higher willingness of consumers (patients) to enter the market and a lower rate of undertreatment by experts.

1. Introduction

Trust is a key aspect in credence goods markets, characterized by information asymmetries between providers and consumers. Compared to other goods, where consumers typically learn about their quality during or after consumption, the quality of credence goods may be difficult to assess even after consumption (Nelson 1970; Darby and Karni 1973; Dulleck and Kerschbamer 2006). Therefore, providers serve as experts and may exploit their informational advantage in various ways: They may provide more services than necessary (overtreatment), fewer services than necessary (undertreatment), or the appropriate service but charge a higher price (overcharging) (Dulleck and Kerschbamer 2006).[1] Thus, the consumer has to trust that the expert will provide the appropriate service and charge the correct price.

Empirical evidence of all three types of unethical behavior has been shown in different credence goods markets, including the market for taxi rides (Balafoutas et al. 2013; Balafoutas et al. 2017), computer specialists (Kerschbamer et al. 2016), auto repair services (Schneider 2012; Rasch and Waibel 2018), and health care (Cutler 1995; Pasero and McCaffery 2001; Clemens and Gottlieb 2014; Jürges and Köberlein 2015; Hennig-Schmidt et al. 2019; Gottschalk et al. 2020; Gross et al. 2021).[2] Despite the spread of unethical behavior over different contexts, the extent of fraud may vary based on the market under investigation.

In healthcare markets, experts are subject to particularly high ethical standards by virtue of the Hippocratic ideals that all physicians subscribe to. Therefore, physicians should be expected to be more

Funding by the Nachwuchsförderung at the University of Innsbruck and the Austrian Central Bank (grant number 17805) is gratefully acknowledged.*Competing interests statement:* The authors have no competing interests.

* Corresponding author.
E-mail address: silvia.angerer@umit-tirol.at (S. Angerer).

[1] Depending on the market structure, undercharging or undertaking costly non-price actions could also take place to signal high quality or that the expert's incentives are aligned with those of the patient (Berg and Kim 2019; Berg et al. 2021). However, in the underlying study, we follow the theoretical structure of and focus on investigating overtreatment, undertreatment, and overcharging as possible problems associated with credence goods.

[2] Even though in most health systems prices are regulated, overcharging is still a prevalent phenomenon in hospital settings with a DRG-based reimbursement system (Jürges and Köberlein 2015; Hennig-Schmidt et al. 2019).

https://doi.org/10.1016/j.socec.2022.101973
Received 17 November 2021; Received in revised form 22 December 2022; Accepted 23 December 2022
Available online 24 December 2022

trustworthy than experts in other credence goods markets.[3] The trustworthiness of physicians has also been underlined by the Gallup poll and the Veracity Index by Ipsos on ratings of the honesty and ethical standards of people in different professions, with medical doctors outperforming most other professions, including bankers, lawyers, and car salespeople (Brenan 2018; IPSOS 2020; IPSOS 2021). Since trust involves beliefs about trustworthiness, consumers (patients) in a healthcare market are expected to trust physicians more than other experts.[4] We aim to test this research hypothesis in our paper using a credence goods framework.[5]

Experimental studies investigating the effect of health framing and the use of naturalistic subject pools, such as medical students or physicians, compared to standard subject pools show that a health frame and subject pool have an impact on medical decision-making in laboratory settings (Ahlert et al. 2012; Ahlert et al. 2013; Hennig-Schmidt and Wiesen 2014; Kesternich et al. 2015; Brosig-Koch et al. 2016; Cox et al. 2016; Galizzi and Wiesen 2017; Galizzi and Wiesen 2018; Reif et al. 2020). However, to date, to the best of our knowledge, there have been no investigations on the impact on patient decision-making in a healthcare market.

In this paper, we use a credence goods model to analyze whether a healthcare framing and a subject pool of experts have an impact on the behavior of consumers in a credence goods market. More specifically, we implement the experimental design of Dulleck et al. (2011), replicate their baseline treatment, and extend it by two experimental conditions. In one condition, we introduce a healthcare framing with a conventional subject pool to investigate the impact of the healthcare context on the behavior of consumers (i.e., patients) and experts (i.e., physicians). In the second variation, we introduce medical students playing the physician role, allowing us to study the impact of the expert subject pool on the behavior of students playing the patient role as well as measure the trustworthiness of real experts in a more naturalistic setting.

Our results suggest that the subject pool, in combination with the healthcare context, has a significant impact on both the behavior of consumers (patients) and experts (physicians). In the condition comprising the healthcare setting and medical students playing the physician role, the frequency of market entry by consumers (patients) and, thus, the interaction rate between the experts (physicians) and consumers (patients) increases by 15 percentage points to 65% compared to 49% in the baseline condition. Also, the rate of undertreatment decreases by 25 percentage points to 39% compared to 63% in the baseline condition. Thus, the level of trust and trustworthiness is significantly higher in the context of the healthcare market compared to a neutrally framed credence goods market.

The role of framing and the subject pool has been investigated in various laboratory experiments.[6] In what follows, we will concentrate on the studies most closely related to the underlying research question. In the context of credence goods experiments, Beck et al. (2014) investigated the behavior of professional car mechanics in an experimental credence goods market and compared it with the behavior of a standard student subject pool. They found very similar behavioral responses—except for the level of overtreatment, which was higher among car mechanics. In a healthcare context, the behavior of standard student subjects, medical students, and physicians was compared by Brosig-Koch et al. (2016). Their study analyzed the effect of different payment schemes (fee-for-service vs. capitation) on the provision of medical services in a laboratory experiment framed in a physician decision-making situation. Their results suggest that all three subject pools reacted consistently toward financial incentives and provided more medical services under fee-for-service than under capitation schemes.[7] However, the physicians reacted less toward financial incentives than the medical and non-medical students. In an earlier study, Hennig-Schmidt and Wiesen (2014) compared medical students with non-medical students in a similar setting, showing that the medical students were more patient-regarding and, thus, willing to sacrifice more from their own profit than the non-medical students, regardless of the payment scheme (fee-for-service or capitation).[8]

In line with these results, Ahlert et al. (2012) and Ahlert et al. (2013) provided evidence showing that, in a neutrally and medically framed allocation task, prospective economists were more selfish than prospective physicians. Concerning the framing of the task, they found that professional norms influenced decision-making, with economists moving further away from selfishness in a medical context. Kesternich et al. (2015) employed a simple distribution task and provided additional evidence on framing effects. In their study, they manipulated the salience of professional norms (i.e., Hippocratic Oath reminder vs. no reminder), framing (neutral vs. medical framing), and the subject pool of the recipient (student vs. real-world charity caring for patients). While the provision of the good was, on average, not higher in the medical framing, it increased significantly when the receiver was a charity. Moreover, the salience of professional norms decreased the selfishness of their (medical student) subjects but only in the medical framing. Using a patient–physician–insurance model, Reif et al. (2020) also investigated the effect of a medical framing and subject pool on medical decision-making in a laboratory experiment. They showed that medical doctors behaved similarly to students and that the health framing increased patient-regarding behavior. Using individual prospect theory, Kairies-Schwarz et al. (2017) analyzed the effect of a health framing on decision quality in health insurance choices. Their results showed that the health framing significantly improved decision quality, resulting in fewer inconsistent choice patterns compared to when a neutral framing was employed. Deploying a political economy model to investigate the decision to exit from public to private finance, Buckley et al. (2016) also compared the effects of a neutrally and medically framed decision

[3] Note that we use medical students playing the physician role. Though Attema et al. (2022) show that altruism decreases during medical education (with an increase again during the practical year), they also show that medical students are more altruistic than non-medical students. Therefore, our hypothesis regarding the higher trustworthiness of physicians compared to other experts in credence goods markets should also extend to medical students. We would like to thank an anonymous referee for pointing this out.

[4] Note that the type of profession is just one factor influencing beliefs about the trustworthiness of an expert. There are other individual-specific, market-based, and systemic factors such as communication, reputation, the market environment, or shocks such as the COVID-19 pandemic, that influence trust in experts (Huang et al. 2018; Baker 2020). These other factors, however, are not investigated in the underlying study.

[5] While the trust game more directly measures trust and trustworthiness, the credence goods framework is better suited to model the information asymmetries between patients and physicians in healthcare markets (Huck et al. 2016; Balafoutas and Kerschbamer 2020). Given the informational advantage of physicians and the possibility that physicians might mistreat patients, market efficiency is determined by the amount of trust (and trustworthiness) in the market (for a discussion of trust in the medical marketplace see Maynard and Bloor 2003). Piette et al. (2005) showed that when patients' trust in their physicians is low, they are more likely to forgo medications because of cost pressure than when trust levels are high. However, there might be other motivations for patients' interaction with doctors, such as urgent treatment needs.

[6] See, for instance, the seminal paper by Tversky and Kahneman (1981) or the experimental investigation of framing and subject pool effects in the trust game (Chaudhuri et al. 2016), public goods games (Cookson 2000; Dufwenberg et al. 2011), and the prisoner's dilemma game (Goerg and Walkowitz 2010).

[7] A similar result was shown in Brosig-Koch et al. (2020), which studied the effect of performance pay on the provision of medical services in an online experiment with physicians and medical students. A comparison of the two subject pools found no difference in the effect of performance pay on medical service provision.

[8] Wang et al. (2020) studied subject pool effects across countries in a similar setting as the previous two studies and found a remarkable stability in patient-regarding preferences in comparisons of the behavior of Chinese medical doctors, German medical students, and Chinese medical students.

Table 1
Experimental conditions.

	Description				
	Frame	Experts	Consumers	# sessions	# markets(# subjects)
B	Neutral	Standard students	Standard students	4	12 (96)
HF	Health	Standard students	Standard students	3	9 (72)
HF&MS	Health	Medical students	Standard students	4	10 (80)
TOTAL				11	31 (248)

situation. They found a significant difference between the two frames, with the proportion of exits being significantly higher in the health frame than in a neutral frame.[9]

Given the evidence on the behavioral differences between healthcare providers and providers in other economic sectors, our study addresses an important topic in experimental health economics (Cox et al. 2016) and contributes to this strand of the literature on two key dimensions. First, compared to the studies investigating the effect of a naturalistic framing and subject pool on medical decision-making (Ahlert et al. 2012; Ahlert et al. 2013; Hennig-Schmidt and Wiesen 2014; Kesternich et al. 2015; Brosig-Koch et al. 2016; Reif et al. 2020), where patients play a passive role and, therefore, are not present in the laboratory, a credence goods model allows the investigation of both sides of the market—experts and consumers. Second, by utilizing a credence goods setting, we also contribute to the literature on credence goods. While various studies have utilized a credence goods model with a focus on the healthcare market (Huck et al. 2016; Mimra et al. 2016a; Mimra et al. 2016b; Angerer et al. 2021), none of them investigated the effect of a health framing compared to a neutral framing. Our study, thus, is the first to provide experimental evidence of the effect of a health framing and the use of prospective physicians as experts on market efficiency as well as expert *and* consumer behavior in a credence goods setting.

In line with the research hypothesis, our study shows that trust and trustworthiness in healthcare markets are higher than for a neutrally framed credence goods market with standard student subjects serving as experts. In what follows, we introduce the conceptual framework and the experimental design in section 2. The results are presented in section 3. Section 4 provides a discussion of our results and concludes the paper.

2. Material and methods

The research design is based on the credence goods model proposed by Dulleck and Kerschbamer (2006) and the experiment implemented by Dulleck et al. (2011), in which the authors tested the institutions of liability, verifiability, competition, and reputation to solve the problems associated with the information asymmetry in their credence goods model. We replicate the data from their baseline condition and extend the design to answer our research questions.[10]

In what follows, we briefly present the structure of the baseline model, described in detail in Dulleck et al. (2011), the experimental conditions, the procedure, and our research questions and hypotheses.

2.1. Basic set-up and parameters

The credence goods setup consists of two parties, the expert (he) and the consumer (she). Experts can provide two types of goods, a high-quality good (q^h) at a cost of $c^h = 6$ or a low-quality good (q^l) at a cost of $c^l = 2$. All consumers need either a high- or low-quality good, each with an equal probability of 50%. The probability is common knowledge; however, in the respective period, consumers are not informed about the type of quality they need. A consumer receives a sufficient level of service in case she needs q^l, independent of the type of quality provided, or in case she needs q^h and receives q^h. Experts learn with certainty the quality needed by the consumer. At the beginning of each period, the expert sets two types of prices, p^h for the high-quality good and p^l for the low-quality good (with $1 \leq p^l \leq p^h \leq 11$).[11] However, the expert can decide independently of the chosen quality whether to charge p^h or p^l. A consumer is randomly matched with an expert and can decide whether to interact with the expert or stay out of the market. The only information the consumer receives before deciding to interact with the expert is the menu of prices that the randomly matched expert charges for both types of quality (see Fig. A1 in the Appendix for a representation of the stage game in the extensive form).

In case of interaction (i.e., the consumer decides to enter the market), the payoffs are as follows. The expert receives $p_i - c_j$ (with $i \in (h,l)$ and $j \in (h,l)$), and the consumer receives the value of the provided quality (v) minus the price charged by the expert. The value depends on whether the quality received was sufficient ($v = 10$) or not ($v = 0$). In the event of no interaction, the payoff is 1.6 for both the expert and consumer.

The stage game of the baseline setup is as follows:

1. For each consumer, nature draws the type of quality needed, with a 50% probability the consumer needs q^h and a 50% probability the consumer needs q^l.
2. The expert posts prices p^h and p^l with $1 \leq p^l \leq p^h \leq 11$.
3. The consumer is informed about the prices p^h and p^l that the randomly matched expert charges and decides whether or not to interact. If the consumer decides not to interact with the expert, the period ends. Otherwise:
4. The expert is informed about the quality needed, provides q^h or q^l, and charges p^h or p^l.
5. Each consumer observes her payoff, and each expert observes his payoff.

2.2. Experimental conditions

The basic setup described in the previous section represents the baseline condition ***B***, which was explained in a neutral framing, with the expert being player A and the consumer being player B.[12] Experts could choose between two actions, action I (q^l) and action II (q^h), and the prices for both actions. The subjects participating in the experiment were standard students, primarily economics and business students. In the two experimental conditions, we introduced the following variations. First, in condition ***HF***, we introduced a health framing, with experts serving as physicians and consumers as patients. Each patient has a

[9] Even though both studies (Buckley et al. 2016; Kairies-Schwarz et al. 2017) investigated framing effects on the demand side of a health market (the insurance demand), they were not directly related to our paper as our focus is on health insurance decisions and not on the behavior of patients and physicians in a physician–patient interaction.

[10] In a previous version of the manuscript, the baseline data were taken from Dulleck et al. (2011). Upon the request of a referee to recollect the baseline data, we replicated the results of the baseline treatment of Dulleck et al. (2011) in the spring of 2022 (see Table A1 in the Appendix for a comparison of our baseline data with the baseline treatment of Dulleck et al. (2011)). The data for the other experimental conditions were collected in the spring of 2019.

[11] In healthcare markets, prices are highly regulated, such that physicians (with some exceptions) do not compete over prices. However, the goal of the underlying study was to investigate the effect of a health framing and not the specific market structure and institutions of healthcare markets compared to other credence goods markets.

[12] Note that in the context of our paper, we use the term expert regardless of whether the experts are represented in a neutral framing (player A), as in condition ***B***, or serving as physicians, as in conditions ***HF*** and ***HF&MS***.

mild or severe health problem that can be cured with the mild treatment q^l or the severe treatment q^h. The subjects participating in this experiment were again standard student subjects, primarily economics and business students. In our second variation, ***HF&MS***, we again used a health frame, now with medical students serving as physicians.[13] The patients were again standard student subjects, primarily economics and business students (see Table 1 for an overview of the experimental conditions). Note that for a systematic 2 × 2 design a condition looking at the effect of medical students serving as experts with a neutral frame is missing (***B&MS***). However, since the experiment is motivated by an investigation of trust and trustworthiness in a healthcare credence goods market, compared to a neutrally framed credence goods market, a condition with medical students serving as experts coupled with a neutral frame would have been an unnatural setting. Further, because of the possibility of students inferring a health setting from this unnatural instruction, the interpretation of this condition (***B&MS***), compared to the condition with a health frame (***HF&MS***) or the condition without medical students (***B***), is ambiguous.

2.3. Experimental protocol

The stage game was repeated for 16 periods. The subjects were randomly assigned either the expert (physician) role or that of the consumer (patient) at the beginning of the experiment, except for condition ***HF&MS*** where all medical students were assigned the physician role and all other invited participants serving as patients. Besides, all subjects were randomly matched within a group of four experts (physicians) and four consumers (patients). Within this matching group of eight subjects, each anonymous consumer (patient) was randomly matched with an anonymous expert (physician) in each period. Ultimately, the earnings of all periods were added to determine the participants' payments.

The data collection was conducted at the laboratory for experimental economic research at the University of Innsbruck, with a total of 248 subjects (see Table 1 for an overview of the sessions, markets, and subjects for each experimental condition). The experiment was computerized using zTree (Fischbacher 2007), and students were recruited using hroot (Bock et al. 2014).[14] Moreover, the study was approved by the internal review board of the University of Innsbruck. Following the protocol used in Dulleck et al. (2011), all subjects were instructed to read a detailed description of the game and answer a set of control questions that were incentivized to give each subject an initial endowment designed to offset any resulting negative period profits (see the instructions and control questions in the Appendix). At the end of the session, the subjects were further asked to fill out a short post-experimental questionnaire on their socio-demographic background information (gender, age, studies, and semester; see Table A2 in the Appendix for the descriptive statistics).[15]

2.4. Research questions and hypotheses

The experimental design enables us to answer the following research questions.

1. Expert (physician) behavior:
   I What is the impact of a health framing on expert behavior in a credence goods setting? (***HF*** vs. ***B***)
   2. Does the expert behavior of prospective physicians differ from that of standard student subjects serving as physicians in a credence goods setting? (***HF&MS*** vs. ***HF***)
2. Consumer (patient) behavior: What is the impact of (i) a health framing (***HF*** vs. ***B***) and (ii) the use of prospective physicians playing the physician role (***HF&MS*** vs. ***HF***) on consumer (patient) behavior in a credence goods setting?
3. Market efficiency: What is the impact of (i) a health framing (***HF&MS*** vs. ***HF***) and (ii) the use of prospective physicians serving as physicians (***HF&MS*** vs. ***HF***) on market efficiency in a credence goods setting?

Assuming the rationality of market participants, risk neutrality and utility functions that depend only on the monetary payoffs of decision-makers, the theoretical predictions for the stage game in the three experimental conditions provide a unique Perfect Bayesian Nash equilibrium in which the experts always provide the low-quality good and charge the price for the high-quality good. Anticipating the behavior of the experts, the consumers do not interact with them, and thus, the market breaks down. The results from Dulleck et al. (2011) showed that the market did not break down. However, there was a significant amount of undertreatment (53%) and overcharging (81%), and only 45% of consumers interacted with an expert. Given the results from the extant literature on the effect of health framing and the subject pool on medical decision-making (i.e. higher patient-regarding behavior; Ahlert et al. 2012; Ahlert et al. 2013; Hennig-Schmidt and Wiesen 2014; Brosig-Koch et al. 2016; Reif et al. 2020), we propose the following hypotheses concerning the behavior of subjects playing the physician role.[16]

Hypotheses 1. Expert (physician) behavior

1a (***HF*** vs. ***B***): The level of undertreatment and overcharging is lower with a health frame than without a health frame.
1b (***HF&MS*** vs. ***HF***): The level of undertreatment and overcharging is lower when prospective physicians play the physician role than with standard student subjects.

Given our expectations regarding the higher level of trustworthiness of subjects playing the physician role in both ***HF*** and ***HF&MS***, we expect subjects playing the patient role to show a higher level of trust and, thus, interact more often.

Hypotheses 2. Consumer (patient) behavior

2a (***HF*** vs. ***B***): The frequency of market interaction between patients and physicians is higher with a health frame than without a health frame.
2b (***HF&MS*** vs. ***HF***): The frequency of market interaction between patients and physicians is higher when prospective physicians play the physician role than with standard student subjects.

Combining the previous hypotheses on expert behavior and consumer behavior leads to the following hypotheses concerning the level of

[13] Note that this information was provided in the experimental instructions to all participants.

[14] Students who had previously participated in a credence goods game more than twice were excluded from the pool of subjects as well as medical students in conditions ***B*** and ***HF***.

[15] In condition ***B***, the subjects also participated in four additional games to measure risk preferences, altruism, trust and lying (see experimental instructions in the Appendix). These additional games were conducted in the baseline condition only because we originally planned to compare our experimental variations with the baseline data from Dulleck et al. (2011), where these data were not collected. These additional games, however, were announced at the end of the session so as to exclude any potential influence of the data collection on the behavior of the subjects compared to the other two experimental conditions.

[16] Note, however, that Kesternich et al. (2015) did not find a significant impact of a health framing on medical decision-making in their setting and that Reif et al. (2020) found no difference between medical doctors and non-medical students.

Table 2
Overview of the means with standard deviations of the baseline condition and difference in means with p-values between conditions.

	B Mean (SD)	*HF vs. B(HF - B)[p-value]*	*HF&MS vs. HF(HF&MS - HF)[p-value]*	*HF&MS vs. B (HF&MS - B)[p-value]*
Overtreatment[1]	0.08 (0.26)	-0.06 [0.042]	-0.01 [0.480]	-0.07 [0.008]
Undertreatment[2]	0.63 (0.48)	-0.15 [0.545]	-0.10 [0.461]	-0.25 [0.048]
Overcharging[3]	0.68 (0.47)	0.07 [0.188]	0.00 [0.837]	0.07 [0.692]
Interaction[4]	0.49 (0.50)	0.08 [0.522]	0.08 [0.347]	0.15 [0.038]
Efficiency[5]	0.55 (0.47)	0.06 [0.522]	0.07 [0.289]	0.12 [0.027]
p^l with interaction	4.44 (1.67)	0.59 [0.406]	-0.42 [0.886]	0.17 [0.668]
p^l without interaction	4.90 (2.26)	0.02 [0.439]	-0.55 [0.165]	-0.53 [0.462]
p^h with interaction	7.47 (1.19)	0.20 [0.482]	0.08 [0.153]	0.28 [0.087]
p^h without interaction	8.12 (1.48)	-0.15 [0.699]	-0.25 [1.000]	-0.40 [0.624]
Actually charged price	7.12 (1.47)	0.32 [0.177]	0.02 [0.683]	0.35 [0.075]
Profit patients/consumers[6]	0.86 (3.44)	-0.00 [0.776]	0.22 [0.327]	0.22 [0.391]
Profit physicians/experts[6]	2.94 (2.00)	0.30 [0.337]	0.15 [0.595]	0.45 [0.041]

Notes: p-values in square brackets from Mann–Whitney U tests (MWU) for pairwise differences between conditions, with matching groups of eight subjects as one independent observation.
[1] The consumer needs q_l, but the expert provides q_h.
[2] The consumer needs q_h, but the expert provides q_l.
[3] The expert provides q_l but charges p_h (with $p_h > p_l$ and consumer needing q_l).
[4] The relative frequency of market interaction.
[5] Calculated as $\frac{\text{actual average profit (across patients and physicians)}}{\text{maximum possible profit (across patients and physicians)}}$.
[6] In experimental currency units.

Table 3
Random effects panel probit/OLS.

	(1)	(2)	(3)	(4)
VARIABLES	OC[1,a]	UT[2,a]	INT[3,a]	EFF[4,b]
HF (=1)	0.817 (0.793)	-0.933 (0.828)	0.260 (0.237)	0.056 (0.053)
p-values	*[0.303]*	*[0.260]*	*[0.272]*	*[0.291]*
HF&MS (=1)	0.753 (0.750)	-1.567 (0.791)	0.502 (0.205)	0.118 (0.045)
p-values	*[0.316]*	*[0.047]*	*[0.015]*	*[0.009]*
Period	0.116 (0.027)	0.074 (0.029)	-0.046 (0.008)	-0.003 (0.002)
p-values	*[0.000]*	*[0.010]*	*[0.000]*	*[0.081]*
Constant	0.184 (0.419)	0.303 (0.355)	0.359 (0.094)	0.578 (0.025)
p-values	*[0.662]*	*[0.393]*	*[0.000]*	*[0.000]*
Observations	618	505	1,984	1,984
Number of subjects	122	122	124	124
Wald test (p-values)				
$H_0 : \beta_{HF} = \beta_{HF\&MS}$	0.941	0.504	0.400	0.339

Notes: Coefficients with robust standard errors clustered on the matching group level in parentheses and p-values in square brackets.
[1] Overcharging (OC): the expert provides q_l but charges p_h (with $p_h > p_l$ and consumer needing q_l).
[2] Undertreatment (UT): the consumer needs q_h, but the expert provides q_l.
[3] The relative frequency of market interaction (INT).
[4] Market efficiency (EFF) calculated as $\frac{\text{actual average profit (across patients and physicians)}}{\text{maximum possible profit (across patients and physicians)}}$.
[a] Random effects panel probit.
[b] Random effects panel OLS.

market efficiency, measured as the sum of patient and physician surplus divided by the maximum possible market surplus.

Hypotheses 3. Market efficiency

3a (***HF*** vs. ***B***): The level of market efficiency is higher with a health frame than without a health frame.
3b (***HF&MS*** vs. ***HF***): The level of market efficiency is higher when prospective physicians play the physician role than with standard student subjects.

From our hypotheses, we can derive the following ordering for the expected expert (physician) behavior, consumer (patient) behavior, and market efficiency concerning our three experimental conditions.

1. Expert (physician) behavior, i.e., undertreatment and overcharging: ***HF&MS*** < ***HF*** < ***B***
2. Consumer (patient) behavior, i.e., interaction rate: ***HF&MS*** > ***HF*** > ***B***
3. Market efficiency, i.e., the sum of consumer (patient) and expert (physician) surplus: ***HF&MS*** > ***HF*** > ***B***

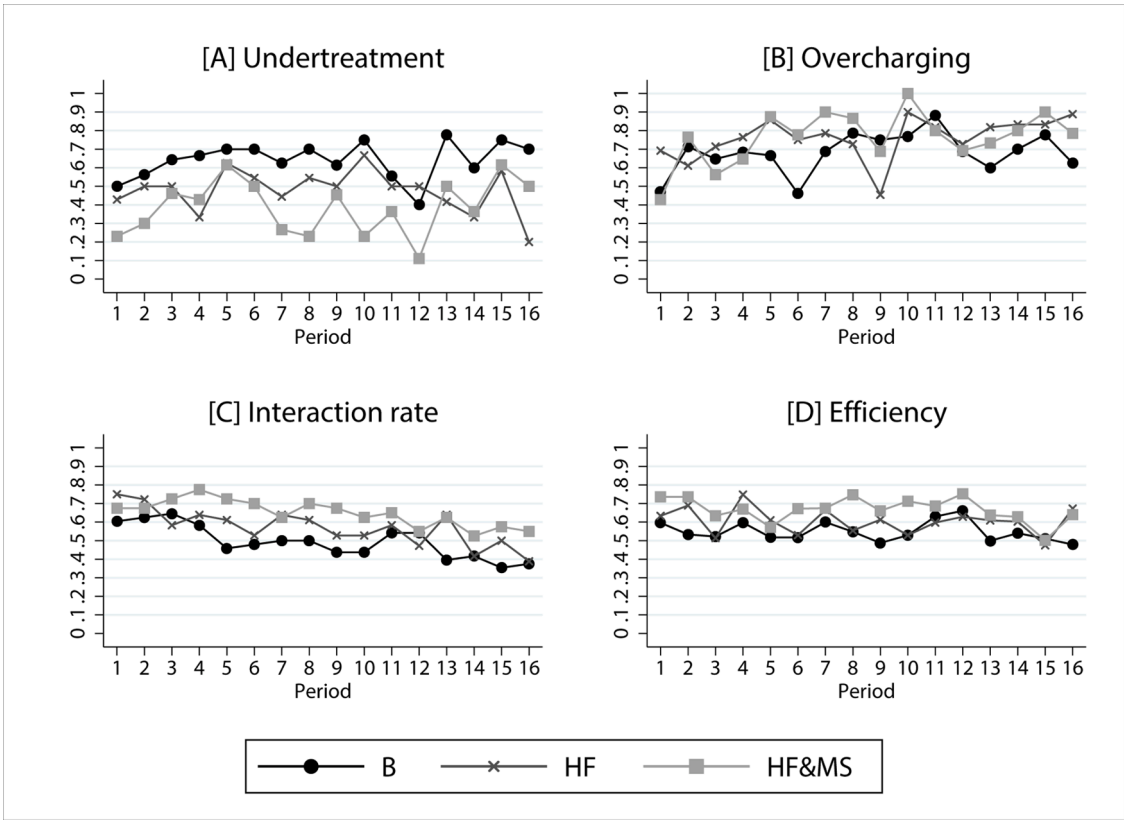


Fig. 1. Provision, charging, market interaction, and efficiency across rounds by conditions ***B, HF***, and ***HF&MS***.

3. Results

In this section, we provide an overview of our main outcome and decision variables for each of the three experimental conditions and discuss the dynamics over the periods and the differences between the conditions to answer our research questions. Table 2 displays the results for the baseline condition ***B*** in column two, and columns three to five provide the binary differences between the three experimental conditions, with the exact p-values in parentheses.[17] Table 3 provides the results from the random effects probit and linear regressions for the differences between the three experimental conditions.[18]

In line with the theoretical prediction, the rate of overtreatment is close to or equal to zero for ***HF*** and ***HF&MS***, respectively.[19] The rate of undertreatment amounts to 39% for ***HF&MS***, 49% for ***HF***, and 63% for the baseline condition. Thus, as expected, the level of undertreatment is lower in a healthcare context with medical students playing the physician role than in a neutrally framed credence goods setting, and the ordering of conditions is as anticipated. However, the difference is only significant between the baseline condition and ***HF&MS*** (see Table 2 for the non-parametric tests and model 2 in Table 3 for the regression results). Concerning overcharging, we find no statistically significant differences between the three treatments (see Table 2 and model 1 in Table 3).[20]

Result 1 (physician/expert behavior): The level of undertreatment in the health framing with medical students playing the physician role is significantly lower than in the baseline condition. This is due to a combination of the health framing and the medical students playing the physician role. The framing (***HF*** vs. ***B)*** and subject pool alone (***HF&MS*** vs. ***HF)*** do not affect the level of undertreatment. The level of overcharging does not differ between the experimental conditions.

The analysis of the interaction rate between the consumers and experts and, thus, the level of consumer trust shows the expected ordering between the treatments with the least amount of trust in the baseline condition (49%), followed by the healthcare setting (57%), and the highest in the healthcare setting with medical students playing the physician role (65%). This difference is significant for the health framing with medical students compared to the baseline condition (see Table 2 and model 3 in Table 3). The binary comparisons between ***HF*** and ***B*** and ***HF*** and ***HF&MS*** show no statistically significant difference (see Table 2 and model 3 in Table 3).

Result 2 (patient/consumer behavior): The frequency of interactions between the patients and physicians is significantly higher for the health framing with medical students playing the physician role compared to the baseline condition. The framing (***HF*** vs. ***B)*** and subject pool alone (***HF&MS*** vs. ***HF)*** have no significant effect on the frequency of the

[17] The baseline data from Dulleck et al. (2011) were successfully replicated, except for the frequency of overcharging, which was significantly higher in Dulleck et al. (2011) than in our baseline data.

[18] The results remain qualitatively unchanged when multilevel mixed-effects panel and linear regressions are used with two random effects equations at the market and individual levels. We decided to show the random effects probit regressions because the multilevel mixed-effects probit model for undertreatment does not converge (see Table A4 in the Appendix).

[19] In our setting, overtreatment is a strictly dominated strategy since overcharging yields a higher payoff for the expert and does not affect the payoff for the consumer.

[20] The rates of overtreatment and overcharging are calculated by considering only periods where the consumer needs the low-quality good. The rate of undertreatment is calculated on the basis of periods where the consumer needs the high-quality good.

market interactions.

In terms of the comparison of the levels of market efficiency, the ordering between the conditions is as expected, and we find a statistically significant increase in market efficiency between ***B*** and ***HF&MS***. However, no statistically significant difference is found between ***HF*** and ***B*** or ***HF&MS*** and ***HF*** (see Table 2 and model 4 in Table 3).

Result 3 (market efficiency): Market efficiency is significantly higher for the health framing with medical students playing the physician role compared to the baseline condition. The framing (***HF*** vs. ***B)*** and subject pool alone (***HF&MS*** vs. ***HF)*** have no significant effect on market efficiency.

Fig. 1 shows the development of our main outcome variables over the 16 periods for our three experimental conditions. Panel C shows that the interaction rate decreases significantly throughout the 16 periods. Nevertheless, the difference in the interaction rate between our conditions persists throughout the 16 periods. Besides, panels A and B also show a significant increase in the level of undertreatment and overcharging for condition ***HF&MS*** as well as a significant increase in the level of overcharging for the baseline condition. No dynamic trends are found for market efficiency (see also the coefficients for the variable "period" in Table 3 and Table A3 in the Appendix for the analysis of period effects for each condition separately).

To investigate the dynamics and determinants of physician and patient behavior in greater detail, Tables A4 and A5 in the Appendix provide random effects panel probit regressions with additional experimental and demographic background variables for our baseline condition. The results show that the behavior of the experts in the credence goods game is significantly related to the level of trustworthiness of the receivers in the trust game; that is, the subjects who are more trustworthy in the trust game engage at a significantly lower rate in undertreatment in the credence goods game (see the coefficient and p-value for the relationship between trustworthiness and undertreatment in model 2 in Table A5). Concerning consumer behavior, we do not find a direct association with the level of trust measured in the trust game; however, we find a significant association between the experience of undertreatment and the decision to interact with an expert. The lower the experience of undertreatment in the market, the more trustworthy the experts are and, thus, the higher the interaction frequency of the consumers (see first coefficient and p-value of Table A6).

4. Discussion and Conclusion

We analyze the impact of a health framing and the use of prospective physicians serving as experts in a credence goods setting on the level of trust, trustworthiness, and the resulting market efficiency compared to a neutrally framed market with standard student subjects serving as experts.

In line with our hypothesis, we find a significant increase in the level of trust and trustworthiness—measured as the frequency of market interactions between patients and physicians and the frequency of undertreatment by experts—with the introduction of a health frame coupled with medical students playing the physician role compared with a neutral framing with standard student subjects. In a healthcare setting with prospective physicians serving as experts, the interaction rate increases by 15 percentage points to 65%, and the undertreatment rate decreases by 25 percentage points to 39% compared to a neutrally framed setting with standard students serving as experts. Furthermore, compared with the baseline condition, there is a significant increase in market efficiency when a health frame is introduced along with medical students playing the physician role. This effect is driven by the combination of a health frame and the use of medical students playing the physician role. The framing and subject pool in isolation have no significant impact on patient and expert behavior when we compare condition ***B*** with ***HF*** and ***HF&MS*** with ***HF***.

Our results regarding the effect of a health framing and medical students playing the physician role on expert behavior support those of previous experiments investigating the effect of a health framing and subject pool on medical decision-making (Ahlert et al. 2012; Ahlert et al. 2013; Hennig-Schmidt and Wiesen 2014; Brosig-Koch et al. 2016). However, in our setting, the effect of the framing or subject pool alone is not significant, and only a combination of the two leads to a significant effect. This could be due to two important differences.

First, in the other experiments, the authors used an individual allocation task where only subjects playing the physician role made a decision. In our setting, due to our primary research question, subjects playing the patient role also make decisions, and thus, the behavior of the physicians (experts) also depends on their beliefs regarding the behavior of future patients (i.e., whether future patients decide to interact). Given the fact that a significant proportion of the subjects in the patient role does not interact (even though the interaction rate increases in conditions ***HF*** and ***HF&MS***), we speculate that there might have been a stronger strain on the subjects in the expert role so as to guarantee an adequate profit at the end of the experiment and, thus, engage in more undertreatment and overcharging compared to a situation where the period profit is independent of the decision of other individuals, as in other studies.

Second, compared to Brosig-Koch et al. (2016) and Hennig-Schmidt and Wiesen (2014), the patients in our setting were standard student subjects participating in experiments, and the health benefits generated by the experts in the healthcare frame were translated into money paid to the subjects serving as patients. Brosig-Koch et al. (2016) and Hennig-Schmidt and Wiesen (2014) used passive real patients as receivers of the health benefits created by the subjects playing the physician role by transferring the monetary value to a charity caring for ophthalmic patients. This procedure increased the salience of the professional norm, which could be an explanation for the different findings (Kesternich et al. 2015). However, given our primary research question, the creation of real health benefits outside the laboratory was not feasible and represents an additional change compared to Dulleck et al. (2011), which is worthy of future investigation.

One concern with our credence goods setting, which was raised by an anonymous referee, is the interpretation of the treatment effects regarding patient behavior as differences in trust and treatment effects regarding physician behavior as differences in trustworthiness. Besides the level of trust toward physicians, the decision to interact in healthcare markets depends on additional factors, such as the urgency of the treatment. However, in our controlled experimental environment, the interaction frequency mainly depends on expert behavior. Given that our results show that expert behavior is significantly associated with the trustworthiness measured in the trust game and that the decision to interact depends on the experience of undertreatment in the market, we are confident that trust and trustworthiness are the main drivers of behavior in our setting.

Overall, our results confirm that the level of trust and trustworthiness in a healthcare credence goods market with medical students playing the physician role is higher than that in a neutrally framed credence goods market with standard student subjects serving as experts.

Data availability

Data will be made available on request.

Acknowledgements

We thank Thomas Rittmannsberger for the assistance in data collection and two anonymous referees for their helpful comments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.socec.2022.101973.

Appendix

Experimental Instructions (condition HF and condition HF&MS in italics)

Dear participants,

Welcome to today's experiment!

Please read the instructions for the experiment carefully. All statements in the instructions are true. Your payoff at the end of the experiment depends on how well you have understood the instructions. All data gathered during the experiment will be treated confidentially and evaluated anonymously.

We ask you to remove all items, including other reading materials and writing utensils from the table, and switch off your mobile phone, as well as any other electronic devices. If you have a question, raise your hand and one of the experimenters will come to you to answer your question privately.

All personal designations in this experiment refer equally to men and women.

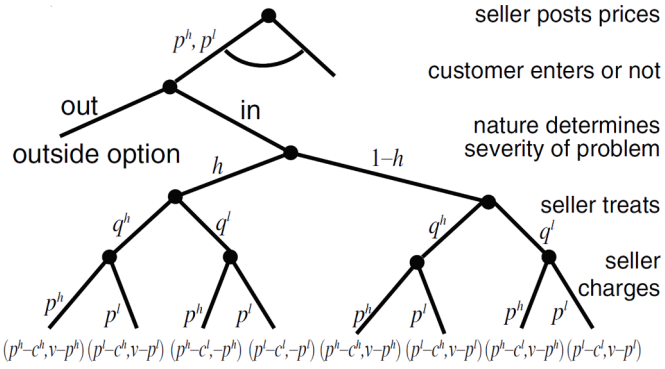


Fig. A1. Extensive Form of the Stage Game.
Source: Dulleck et al. (2011)

Table A1
Overview of the means with standard deviations of our baseline condition and the baseline condition of Dulleck et al. (2011) and difference in means with p-values between conditions.

	B Mean (SD)	*B-Dulleck Mean (SD)*	*B vs. B-Dulleck(B – B-Dulleck)[p-value]*
Overtreatment[1]	0.08 (0.26)	0.06 (0.24)	0.01 [0.430]
Undertreatment[2]	0.63 (0.48)	0.53 (0.50)	0.10 [0.174]
Overcharging[3]	0.68 (0.47)	0.81 (0.39)	-0.13 [0.038]
Interaction[4]	0.49 (0.50)	0.45 (0.50)	0.04 [0.402]
Efficiency[5]	0.55 (0.47)	0.56 (0.44)	-0.01 [0.644]
p^l with interaction	4.44 (1.67)	4.67 (1.48)	-0.23 [0.199]
p^l without interaction	4.90 (2.26)	5.17 (1.70)	-0.27 [1.000]
p^h with interaction	7.47 (1.19)	7.28 (1.15)	0.19 [0.317]
p^h without interaction	8.12 (1.48)	7.91 (1.19)	0.21 [0.647]
Actually charged price	7.12 (1.47)	7.08 (1.29)	0.04 [0.707]
Profit patients/consumers[6]	0.86 (3.44)	1.00 (3.05)	-0.14 [0.488]
Profit physicians/experts[6]	2.94 (2.00)	2.69 (1.80)	0.25 [0.119]

Notes: p-values in square brackets from Mann–Whitney U tests for pairwise differences between conditions, with matching groups of eight subjects as one independent observation.

[1] The consumer needs q_l, but the expert provides q_h.

[2] The consumer needs q_h, but the expert provides q_l.

[3] The expert provides q_l but charges p_h (with $p_h > p_l$ and consumer needing q_l).

[4] The relative frequency of market interaction.

[5] Calculated as $\frac{\text{actual average profit (across patients and physicians)}}{\text{maximum possible profit (across patients and physicians)}}$.

[6] In experimental currency units.

Table A2
Descriptive statistics subject pool.

	B		***HF***		***HF&MS***	
	Non-medical students (expert role)	Non-medical students (customer role)	Non-medical students (physician role)	Non-medical students (patient role)	Medical students (physician role)	Non-medical students (patient role)
Male (=1)	40%	50%	61%	50%	35%	45%
Age in years	22.48 (2.10)	22.83 (2.50)	22.89 (3.21)	23.08 (3.55)	23.53 (3.19)	23.7 (6.63)
Study progress	4.27 (2.07)	4.42 (1.97)	3.64 (2.17)	3.58 (2.71)	6.25 (3.18)	3.95 (3.06)

Table A3
Random effects panel probit / OLS for each condition separately.

VARIABLES	(1) OC[1,a]	(2) UT[2,a]	(3) INT[3,a]	(4) EFF[4,b]
Condition ***B***				
Period	0.128 (0.049)	0.040 (0.052)	-0.045 (0.012)	-0.002 (0.003)
p-values	*[0.008]*	*[0.439]*	*[0.000]*	*[0.543]*
Constant	0.015 (0.428)	0.392 (0.347)	0.359 (0.102)	0.566 (0.035)
p-values	*[0.972]*	*[0.259]*	*[0.000]*	*[0.000]*
Observations	213	166	768	768
Number of subjects	47	48	48	48
Condition ***HF***				
Period	0.226 (0.527)	0.064 (0.041)	-0.052 (0.011)	-0.003 (0.002)
p-values	*[0.668]*	*[0.122]*	*[0.000]*	*[0.190]*
Constant	7.100 (61.713)	-0.548 (0.901)	0.671 (0.205)	0.634 (0.061)
p-values	*[0.908]*	*[0.543]*	*[0.001]*	*[0.000]*
Observations	177	152	576	576
Number of subjects	35	35	36	36
Condition ***HF&MS***				
Period	0.091 (0.035)	0.136 (0.046)	-0.040 (0.018)	-0.005 (0.004)
p-values	*[0.010]*	*[0.003]*	*[0.027]*	*[0.214]*
Constant	0.646 (0.667)	-2.093 (5.467)	0.823 (0.113)	0.710 (0.052)
p-values	*[0.333]*	*[0.702]*	*[0.000]*	*[0.000]*
Observations	228	187	640	640
Number of subjects	40	39	40	40

Notes: Coefficients with robust standard errors clustered on the matching group level in parentheses and p-values in square brackets.
[1] Overcharching (OC): the expert provides q_l, but charges p_h (with $p_h > p_l$ and consumer needing q_l).
[2] Undertreatment (UT): the consumer needs q_h, but the expert provides q_l.
[3] The relative frequency of market interaction (INT).
[4] Market efficiency (EFF) calculated as $\frac{\text{actual average profit (across patients and physicians)}}{\text{maximum possible profit (across patients and physicians)}}$.
[a] Random effects panel probit.
[b] Random effects panel OLS.

Thank you very much for your participation in today's experiment.

Instructions for the experiment

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

2. Roles and 16 rounds

This experiment consists of **16 rounds**, each with the same sequence of decisions. The sequence of decisions is explained in detail below.

There are 2 roles in the experiment: **Doctor** and **patient**. At the beginning of the experiment, you will be randomly assigned one of these roles and keep it for the whole experiment. On the first screen of the experiment, you will see which role you have. This role remains the same for all 16 rounds. (*HF&MS: All participants with the participant numbers 1-4, 9-12, and 17-20 are in the role of doctors. All other participants (participant numbers 5-8, 13-16, and 21-24) are in the role of patients. Participants in the role of doctors are medical students, whereas participants in the role of patients are non-medical students (all fields of study except medicine))*.

A doctor always interacts with a patient. However, the pairs **change** after each round. This means that you interact with a **new** player (of the other role) in each round.

All experiment participants receive the same information regarding the rules of the game, including costs and payoffs to both players.

Overview of the decisions in a round

Each round consists of a maximum of 4 decisions, which are made successively. Decisions 1, 3, and 4 are made by the doctor; decision 2 is made by the patient.

The sequence of decisions in every round (concise)

1. The doctor chooses prices for the minor and the major treatment.

Table A4
Multilevel mixed-effects probit (linear) regression models.

VARIABLES	(1) OC[1,a]	(3) INT[2,a]	(4) EFF[3,b]
HF (=1)	1.022 (0.923)	0.264 (0.237)	0.056 (0.053)
p-values	*[0.268]*	*[0.265]*	*[0.291]*
HF&MS (=1)	1.080 (1.492)	0.503 (0.204)	0.118 (0.045)
p-values	*[0.469]*	*[0.014]*	*[0.009]*
Period	0.117 (0.030)	-0.045 (0.008)	-0.003 (0.002)
p-values	*[0.000]*	*[0.000]*	*[0.080]*
Constant	0.563 (1.257)	0.359 (0.093)	0.578 (0.025)
p-values	*[0.654]*	*[0.000]*	*[0.000]*
Observations	618	1,984	1,984
Number of subjects	122	124	124
Wald test (p-values)			
$H_0 : \beta_{HF} = \beta_{HF\&MS}$	0.945	0.406	0.338

Notes: Coefficients with robust standard errors in parentheses and p-values in square brackets. Model 2 with undertreatment as dependent variable is not included in this Table because of convergence problems.

[1] Overcharching (OC): the expert provides q_l, but charges p_h (with $p_h > p_l$ and consumer needing q_l).

[2] The relative frequency of market interaction (INT).

[3] Market efficiency (EFF) calculated as $\frac{\text{actual average profit (across patients and physicians)}}{\text{maximum possible profit (across patients and physicians)}}$.

[a] Multilevel mixed-effects probit with two random effects equations at the market and individual level.

[b] Multilevel mixed-effects linear regression with two random effects equations at the market and individual level.

Table A5
Determinants of expert behavior in condition *B*, random effects panel probit.

VARIABLES	(1) OC[1,a]	(2) UT[2,a]
Past interaction[3]	0.222 (0.086)	0.099 (0.093)
p-values	*[0.010]*	*[0.288]*
Male (=1)	0.858 (0.641)	0.682 (0.413)
p-values	*[0.181]*	*[0.099]*
Age in years	0.080 (0.128)	0.157 (0.184)
p-values	*[0.531]*	*[0.392]*
Study progress	0.056 (0.146)	-0.160 (0.126)
p-values	*[0.699]*	*[0.206]*
Trustworthiness[4]	-2.002 (1.407)	-4.765 (1.416)
p-values	*[0.155]*	*[0.001]*
Altruism[5]	0.133 (0.127)	0.052 (0.125)
p-values	*[0.295]*	*[0.678]*
Liar (=1)[6]	1.184 (1.447)	0.112 (0.663)
p-values	*[0.413]*	*[0.865]*
Constant	-2.860 (3.422)	-1.550 (4.043)
p-values	*[0.403]*	*[0.702]*
Observations	213	166
Number of subjects	47	48

Notes: Coefficients with robust standard errors clustered on the matching group level in parentheses and p-values in square brackets.

[1] Overcharching (OC): the expert provides q_l, but charges p_h (with $p_h > p_l$ and consumer needing q_l).

[2] Undertreatment (UT): the consumer needs q_h, but the expert provides q_l.

[3] Sum of past interactions with consumers.

[4] Trustworthiness measured with a trust game (see experimental instructions below).

[5] The amount donated to charity in a dictator game (DG) as a measure for altruism (see experimental instructions below).

[6] Classified as a liar (=1 when reporting 4 or more correct dices out of 12 in a lying task) (see experimental instructions below).

2. The patient learns the prices chosen by the doctor. Then the patient decides if he/she wants to visit (interact with) this doctor. If not, this round ends.

   If yes ...

1. The doctor (but **not** the patient) receives information about whether the patient has a minor or a major illness. Then the doctor chooses a minor treatment or a major treatment.
2. The doctor asks the patient to pay the price for one of the two treatments (minor or major) set in decision 1. The price charged need not be the same as the price for the treatment chosen in decision 3 but maybe the price of the other treatment.

Detailed presentation of decisions and their consequences in terms of payoffs

Decision 1. In decision 3 the **doctor** chooses between two treatments, a minor treatment, and a major treatment.

The **minor treatment costs** the doctor **2 points** (= experimental currency unit)

The **major treatment costs** the doctor **6 points**.

For both treatments, the doctor charges prices to the patient. In **decision 1** the doctor has to set **prices for both treatments**. Only (strictly) positive prices in full points from 1 point to a maximum of 11 points are possible. I.e. the allowed prices are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11.

Please note that the price for the minor treatment may not exceed the price for the major treatment.

Decision 2. The **patient** learns the prices set by the doctor in decision 1. Then the patient decides if he/she wants to see (interact with) this doctor.

If yes, this means that the doctor in decisions 3 and 4 chooses a treatment and charges a price for it (see below). However, the patient will **not** be able to observe which treatment the doctor chooses.

If no, this round ends and both players receive a **payoff of 1.6 points for this round**.

Decision 3. Before decision 3 (if the patient has chosen "yes" in decision 2) an illness is randomly determined for the patient. The **patient** may have 2 types of illnesses: A **minor illness** or a **major illness**. The illness is randomly determined every new round.

The patient suffers with a **50% chance** from a **minor illness** and with a **50% chance** from a **major illness**. Imagine a coin toss in each round – if the coin shows "head", then the patient suffers from a minor illness, if it shows "tails", the patient suffers from a major illness.

The **doctor learns** about the **patient's illness** before making his/her decision 3. Then the doctor chooses a treatment, either the minor treatment or the major treatment.

The **treatment** cures the patient's illness under the following conditions:

a) The patient has a minor illness and the doctor chooses either the minor treatment or the major treatment.
b) The patient has a major illness and the doctor chooses the major treatment.

The treatment **does not cure the disease** if the patient has a major illness, but the doctor chooses the minor treatment.

The patient receives 10 points if the treatment chosen by the doctor cures his/her illness. The patient receives 0 points if the treatment chosen by the doctor does not cure his/her illness.

At **no time** is the **patient** informed on the computer screen whether he/she has had a minor illness or a major illness in any round or which treatment the doctor has chosen.

Decision 4. The doctor **asks** the patient to pay the **price,** set in decision 1, for one of the two treatments. The asked price does **not have to be** equal to the price of the treatment chosen in decision 3 but can be the price of the other treatment.

Payoffs

If the patient in decision 2 ends the round (*decision "**No**" of the patient*), then both players receive **1.6 points** in this round.

Otherwise (*decision "Yes" of the patient*) the payoffs are as follows:

The **doctor** receives the **price** (in points) chosen in decision 4 **minus the costs** for the treatment chosen in decision 3.

For the **patient**, the payoff depends on whether the treatment chosen by the doctor in decision 3 has cured the patient's illness.

a) The doctor's treatment cured the illness. The **patient** receives **10 points minus the price** of the treatment chosen in decision 4.
b) The doctor's treatment did not cure the illness. The **patient** must pay the price chosen in decision 4.

At the beginning of the experiment, you will receive an **initial endowment of 6 points**. Additionally, you will receive another **10 points** for answering the control questions. From this initial endowment, you can pay for possible losses in individual rounds. Losses can also be compensated by winnings from other rounds. If you have made a total loss at the end of the experiment, you must pay this loss to the experimenter. By participating in the experiment, you agree to this condition. Please note, however, that there is **always** a way to avoid losses with certainty in this experiment.

For the payoff, the initial endowment and the winnings of all rounds are added together and converted into cash at the end of the experiment using the following exchange rate:

1 point = 25 Euro-Cent
(i.e. 4 points = 1 euro).

Control Questions

It is important to make sure that all participants have fully understood the experiment. Should something have remained unclear, please ask the experimenter. You will receive 10 points (= 2.5 Euro) for answering the questions correctly. Please answer the following questions:

Question	Correct Answer
1. How many decisions does a physician maximally make per period?	3
2. How many decisions does a patient make per period?	1
Assess whether the statements below are true or false.	
3. "The price for the minor treatment can be higher than the price for the major treatment."	F
4. "If the patient in decision 2 chooses "No", then both players receive 1.6 points in this round."	T
5. "The patient is shown on a computer screen which illness he suffers from in a particular round."	F
6. "If the physician chooses a treatment that cures the illness, the total payoff for the patient in that round is exactly 10 points."	F
7. "The maximum payoff for a patient in a round is 9 points, namely if the physician chooses a treatment that cures the illness and charges a price of 1 point."	T

(*continued on next page*)

(*continued*)

Question	Correct Answer
8. "The initial endowment of 6 points is worth 1.5 euros."	T
Please calculate the payoffs (in points) for the patient and the physician in the following examples:	
9. The physician sets the following prices: the price for the minor treatment = 2 points, the price for the major treatment = 8 points.	Patient: 1.6
The patient chooses "No" in decision 2.	Physician: 1.6
10. The physician sets the following prices: the price for the minor treatment = 3 points, the price for the major treatment = 7 points.	Patient: 3
The patient chooses "Yes" in decision 2.	Physician: 1
The patient suffers from a major illness.	
The physician chooses the major treatment and charges the price for the major treatment.	
11. The physician sets the following prices: the price for the minor treatment = 7 points, the price for the major treatment = 9 points.	Patient: 3
The patient chooses "Yes" in decision 2.	Physician: 5
The patient suffers from a mild illness.	
The physician chooses the mild treatment and charges the price for the mild treatment.	

The experiment continues as soon as all participants have answered the questions correctly.

Experimental Instructions for additional games and questionnaire (condition B only)

Part 2

The experiment is not yet over. There are 4 more parts following. At the end of the experiment, one of these parts (part 2, part 3, part 4, or part 5) is randomly selected for payment.

In part 2, you have to decide regarding your payoff as well as the payoff of another person. This person is a patient who is supported by the organization "Licht für die Welt". The organization "Licht für die Welt" is known worldwide for preventing and curing preventable blindness. It enables **eye surgery** and **supplies people with eyeglasses and medicines for eye diseases** in South America, Africa, and Asia. You have an endowment of € 12 and you need to decide how you want to divide the money. There are two fields on your screen. One field is marked "amount for me" and the other field is marked "amount for Licht für die Welt". The amounts you enter always have to add up to € 12, in units of € 0.1 (i.e., 10 cents). The transfer will be made online at the end of the experiment. To be able to donate to the organization "Licht für die Welt" correctly, we kindly ask the participant with ID 1 to confirm that the money has been transferred to the organization after the online transfer has been made. As a reminder, this part will only be paid if part 2 is randomly selected for payment at the end of the experiment. This also applies to the donation to "Licht für die Welt".

Part 3

As a reminder, this part will only be paid if part 3 is randomly selected for payment at the end of the experiment. Part 3 consists of 20 decisions. Below, you are asked to decide for each situation. Each of your choices is a selection between "Option A" and "Option B".

"Option A" always offers an uncertain payoff: with a 50% probability, you will receive € 12, and a 50% probability you receive € 0.

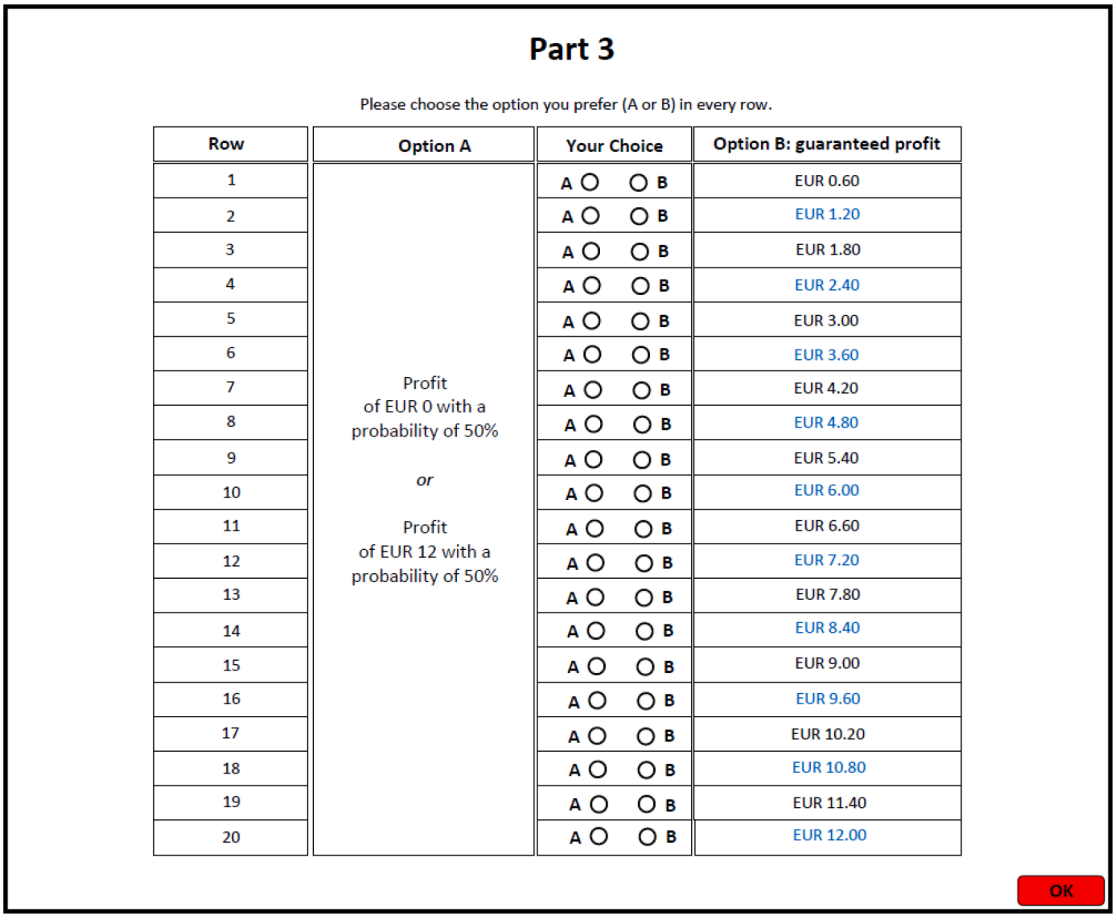
"Option B" always offers a safe payoff: with 100% probability you receive an amount that varies from decision to decision (that is, you receive the guaranteed payoff of that row).

The decision situation will be presented to you on the screen as follows:

Table A6
Determinants of consumer behavior in condition *B*, random effects panel probit.

VARIABLES	INT[1]
Experience of Undertreatment[2]	-0.478 (0.071)
p-value	*[0.000]*
Male (=1)	0.001 (0.302)
p-value	*[0.997]*
Age in years	0.001 (0.042)
p-value	*[0.983]*
Study progress	-0.044 (0.070)
p-value	*[0.529]*
Trust[3]	0.015 (0.123)
p-value	*[0.900]*
Riskaversion[4]	0.023 (0.017)
p-value	*[0.193]*
Constant	0.352 (0.943)
p-value	*[0.709]*
Observations	768
Number of subjects	48

Notes: Coefficients with robust standard errors clustered on the matching group level in parentheses and p-values in square brackets.
[1] The relative frequency of market interaction (INT).
[2] Sum of previous periods experience with undertreatment.
[3] Trust measured with a trust game (see experimental instructions below).
[4] Number of risk averse choices in risk elicitation task with 20 binary choices between safe amount and lottery (see experimental instructions below).



Part 3

Please choose the option you prefer (A or B) in every row.

Row	Option A	Your Choice	Option B: guaranteed profit
1	Profit of EUR 0 with a probability of 50% *or* Profit of EUR 12 with a probability of 50%	A ○ ○ B	EUR 0.60
2		A ○ ○ B	EUR 1.20
3		A ○ ○ B	EUR 1.80
4		A ○ ○ B	EUR 2.40
5		A ○ ○ B	EUR 3.00
6		A ○ ○ B	EUR 3.60
7		A ○ ○ B	EUR 4.20
8		A ○ ○ B	EUR 4.80
9		A ○ ○ B	EUR 5.40
10		A ○ ○ B	EUR 6.00
11		A ○ ○ B	EUR 6.60
12		A ○ ○ B	EUR 7.20
13		A ○ ○ B	EUR 7.80
14		A ○ ○ B	EUR 8.40
15		A ○ ○ B	EUR 9.00
16		A ○ ○ B	EUR 9.60
17		A ○ ○ B	EUR 10.20
18		A ○ ○ B	EUR 10.80
19		A ○ ○ B	EUR 11.40
20		A ○ ○ B	EUR 12.00

OK

If part 3 happens to be paid out, one of the 20 decisions (lines) will be randomly selected for payment. Additionally, it will be randomly determined if you won the lottery (you receive € 12) or if you lost the lottery (you receive € 0) (if you have chosen the lottery option).

When you have made all decisions, please confirm with "OK".

Part 4

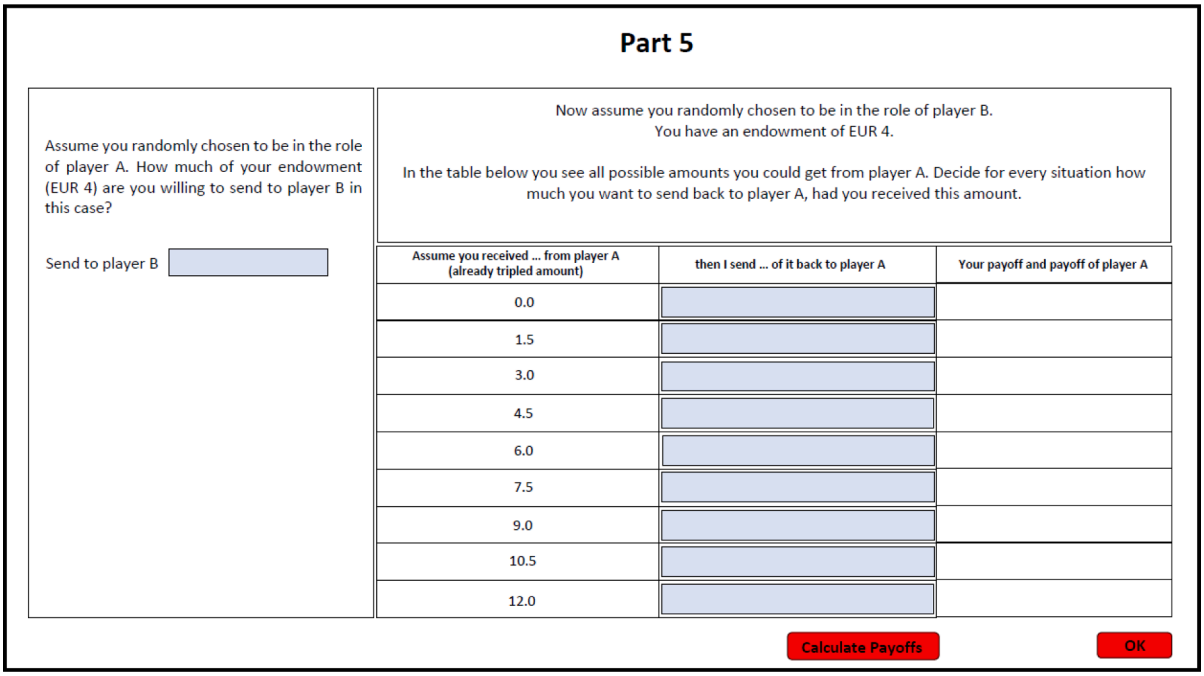
As a reminder, this part will only be paid if part 4 is randomly selected for payment at the end of the experiment. Part 4 is about guessing the outcome of a die roll in a situation marked by randomness. You play 12 rounds of a dice guessing game. Thereby you should guess the number shown on the dice. The more outcomes you guess correctly, the more money you earn. Each round of the game works as follows:

1. First, guess what number will result from the die roll. If you have a number in your head, press the "Next" button.
2. Now you see a dice rolled randomly by the computer. Below the dice, you have to enter what number you have guessed.

For each correctly guessed dice you receive 1 €. For each wrongly guessed die roll you receive 20 cents. The profits of all 12 rounds are added up at the end.

Part 5

As a reminder, this part will only be paid if part 5 is randomly selected for payment at the end of the experiment. Part 5 works as follows: There are two roles, the role of player A and player B. Both players have an initial endowment of € 4 each. Player A has to decide how much of this endowment (between € 0 and € 4, in 50-cent increments) he wants to send to player B. The total amount sent to player B is tripled. The rest is kept by player A (without tripling). Player B may then decide how much of the tripled amount he wants to send back to player A. You have to decide in the role of player A (see the left side of the decision situation on the screenshot below) as well as in the role of player B (for all possible situations, see the right side of the decision situation on the screenshot below). Only at the end of the game, it will be randomly determined in which role you are in. Besides, you will be assigned to a partner playing the other role. You receive the payoff for your decisions in the role chosen for you at random, in combination with the behavior of your randomly assigned partner.



References

Ahlert, M., Felder, S., & Vogt, B. (2012). Which patients do I treat? An experimental study with economists and physicians. *Health Economics Review, 2*(1), 1. https://doi.org/10.1186/2191-1991-2-1

Ahlert, M., Funke, K., & Schwettmann, L. (2013). Thresholds, productivity, and context: an experimental study on determinants of distributive behaviour. *Social Choice and Welfare, 40*(4), 957–984. https://doi.org/10.1007/s00355-012-0652-8

Angerer, S., Glätzle-Rüetzler, D., & Waibel, C. (2021). Monitoring institutions in health care markets: experimental evidence. *Health Economics, 30*(5), 951–971. https://doi.org/10.1002/hec.4232

Attema, A. E., Galizzi, M. M., Groß, M., Hennig-Schmidt, H., Karay, Y., l'Haridon, O. and Wiesen, D. (2022). The Formation of Physician Altruism. *Available at SSRN:* https://ssrn.com/abstract=4079393 or https://dx.doi.org/10.2139/ssrn.4079393.

Baker, D. W. (2020). Trust in health care in the time of covid-19. *JAMA, 324*(23), 2373–2375. https://doi.org/10.1001/jama.2020.23343

Balafoutas, L., Beck, A., Kerschbamer, R., & Sutter, M. (2013). What drives taxi drivers? A field experiment on fraud in a market for credence goods. *The Review of Economic Studies, 80*(3), 876–891. https://doi.org/10.1093/restud/rds049

Balafoutas, L., & 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*, Article 100285. https://doi.org/10.1016/j.jbef.2020.100285

Balafoutas, L., Kerschbamer, R., & Sutter, M. (2017). Second-degree moral hazard in a real-world credence goods market. *Economic Journal, 127*(599), 1–18. https://doi.org/10.1111/ecoj.12260

Beck, A., Kerschbamer, R., Qiu, J., & Sutter, M. (2014). Car mechanics in the lab—Investigating the behavior of real experts on experimental markets for credence goods. *Journal of Economic Behavior & Organization, 108*, 166–173. https://doi.org/10.1016/j.jebo.2014.09.008

Berg, N., & Kim, J.-Y. (2019). A Good Advisor. *Bulletin of Economic Research, 71*(3), 558–572. https://doi.org/10.1111/boer.12180

Berg, N., Kim, J.-Y., & Seon, I. (2021). A performance-based payment: Signaling the quality of a credence good. *Managerial and Decision Economics, 42*(5), 1117–1131. https://doi.org/10.1002/mde.3295

Bock, O., Baetge, I. and Nicklisch, A. (2014). *Hroot: Hamburg Registration and Organization Online Tool*.

Brenan, M. (2018). "Nurses Again Outpace Other Professions for Honesty, Ethics." Retrieved March 31, 2020, from https://news.gallup.com/poll/245597/nurses-again-outpace-professions-honesty-ethics.aspx.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J. and Wiesen, D. (2020). Physician performance pay: Experimental evidence. *Available at SSRN:* https://ssrn.com/abstract=3467583 or https://dx.doi.org/10.2139/ssrn.3467583.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & 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 & Organization, 131*, 17–23. https://doi.org/10.1016/j.jebo.2015.04.011

Buckley, N., Cuff, K., Hurley, J., Mestelman, S., Thomas, S., & Cameron, D (2016). Should I stay or should I go? Exit options within mixed systems of public and private health care finance. *Journal of Economic Behavior & Organization, 131*, 62–77. https://doi.org/10.1016/j.jebo.2016.05.013

Chaudhuri, A., Li, Y., & Paichayontvijit, T. (2016). What's in a frame? Goal framing, trust and reciprocity. *Journal of Economic Psychology, 57*, 117–135. https://doi.org/10.1016/j.joep.2016.09.005

Clemens, J., & Gottlieb, J. D. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review, 104*(4), 1320–1349. https://doi.org/10.1257/aer.104.4.1320

Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics, 3*(1), 55–79. https://doi.org/10.1023/A:1009994008166

Cox, J. C., Green, E. P., & Hennig-Schmidt, H. (2016). Experimental and behavioral economics of healthcare. *Journal of Economic Behavior & Organization, 131*, A1–A4. https://doi.org/10.1016/j.jebo.2016.10.011

Cutler, D. M. (1995). The incidence of adverse medical outcomes under prospective payment. *Econometrica, 63*(1), 29–50. https://doi.org/10.2307/2951696

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

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior, 73*(2), 459–478. https://doi.org/10.1016/j.geb.2011.02.003

Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature, 44*(1), 5–42. https://doi.org/10.1257/002205106776162717

Dulleck, U., Kerschbamer, R., & 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. https://doi.org/10.1257/aer.101.2.526

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Galizzi, M. M., & Wiesen, D (2017). Behavioural experiments in health: An introduction. *Health Economics, 26*(S3), 3–5. https://doi.org/10.1002/hec.3629

Galizzi, M. M., & Wiesen, D (2018). *Behavioral Experiments in Health Economics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190625979.013.244

Goerg, S. J., & Walkowitz, G. (2010). On the prevalence of framing effects across subject-pools in a two-person cooperation game. *Journal of Economic Psychology, 31*(6), 849–859. https://doi.org/10.1016/j.joep.2010.06.001

Gottschalk, F., Mimra, W., & Waibel, C. (2020). Health services as credence goods: A field experiment. *The Economic Journal, 130*(629), 1346–1383. https://doi.org/10.1093/ej/ueaa024

Gross, M., Jürges, H., & Wiesen, D (2021). The effects of audits and fines on upcoding in neonatology. *Health Economics, 30*(8), 1978–1986. https://doi.org/10.1002/hec.4272

Hennig-Schmidt, H., Jurges, H., & Wiesen, D (2019). Dishonesty in health care practice: A behavioral experiment on upcoding in neonatology. *Health Economics, 28*(3), 319–338. https://doi.org/10.1002/hec.3842

Hennig-Schmidt, H., & Wiesen, D (2014). Other-regarding behavior and motivation in health care provision: an experiment with medical and non-medical students. *Social Science & Medicine, 108*, 156–165. https://doi.org/10.1016/j.socscimed.2014.03.001

Huang, E. C., Pu, C., Chou, Y. J., & Huang, N. (2018). Public trust in physicians-health care commodification as a possible deteriorating factor: Cross-sectional analysis of 23 countries. *Inquiry, 55*, Article 46958018759174. https://doi.org/10.1177/0046958018759174

Huck, S., Lünser, G., Spitzer, F., & Tyran, J.-R. (2016). Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. *Journal of Economic Behavior & Organization, 131*, 78–105. https://doi.org/10.1016/j.jebo.2016.06.009

IPSOS. (2020). "Ipsos Veracity Index 2020." from https://www.ipsos.com/en-uk/ipsos-mori-veracity-index-2020-trust-in-professions.

IPSOS. (2021). "Ipsos Veracity Index 2021." from https://www.ctcinfohub.org/2021-ipsos-mori-veracity-index-of-most-trusted-professions-clergy-priests-trusted-by-58-of-population/.

Jürges, H., & Köberlein, J. (2015). The roles of financial incentives and infant health. *Journal of Health Economics, 43*, 13–26. https://doi.org/10.1016/j.jhealeco.2015.06.001

Kairies-Schwarz, N., Kokot, J., Vomhof, M., & Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory – an experiment. *Journal of Economic Behavior & Organization, 137*, 374–397. https://doi.org/10.1016/j.jebo.2017.03.012

Kerschbamer, R., Neururer, D., & 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. https://doi.org/10.1073/pnas.1518015113

Kesternich, I., Schumacher, H., & Winter, J. (2015). Professional norms and physician behavior: Homo oeconomicus or homo hippocraticus? *Journal of Public Economics, 131*, 1–11. https://doi.org/10.1016/j.jpubeco.2015.08.009

Maynard, A., & Bloor, K. (2003). Trust and performance management in the medical marketplace. *Journal of the Royal Society of Medicine, 96*(11), 532–539. https://doi.org/10.1258/jrsm.96.11.532

Mimra, W., Rasch, A., & Waibel, C. (2016a). Price competition and reputation in credence goods markets: Experimental evidence. *Games and Economic Behavior, 100*, 337–352. https://doi.org/10.1016/j.geb.2016.09.012

Mimra, W., Rasch, A., & Waibel, C. (2016b). Second opinions in markets for expert services: Experimental evidence. *Journal of Economic Behavior & Organization, 131*, 106–125. https://doi.org/10.1016/j.jebo.2016.03.004

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy, 78*(2), 311–329.

Pasero, C., & McCaffery, M. (2001). Pain control: The undertreatment of pain. *The American Journal of Nursing, 101*(11), 62–65.

Piette, J. D., Heisler, M., Krein, S., & Kerr, E. A. (2005). The role of patient-physician trust in moderating medication nonadherence due to cost pressures. *Archives of Internal Medicine, 165*(15), 1749–1755. https://doi.org/10.1001/archinte.165.15.1749

Rasch, A., & Waibel, C. (2018). What drives fraud in a credence goods market? – evidence from a field study. *Oxford Bulletin of Economics and Statistics, 80*(3), 605–624. https://doi.org/10.1111/obes.12204

Reif, S., Hafner, L., & Seebauer, M. (2020). Physician behavior under prospective payment schemes—evidence from artefactual field and lab experiments. *International Journal of Environmental Research and Public Health, 17*(15), 5540. https://doi.org/10.3390/ijerph17155540

Schneider, H. S. (2012). Agency problems and reputation in expert services: Evidence from auto repair. *The Journal of Industrial Economics, 60*(3), 406–433. https://www.jstor.org/stable/23324443.

Tversky, A., & Kahneman, D (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453. https://doi.org/10.1126/science.7455683

Wang, J., Iversen, T., Hennig-Schmidt, H., & Godager, G. (2020). Are patient-regarding preferences stable? Evidence from a laboratory experiment with physicians and medical students from different countries. *European Economic Review, 125*, Article 103411. https://doi.org/10.1016/j.euroecorev.2020.103411