

The Value of Rating Systems in Healthcare Credence Goods Markets*

Silvia Angerer, Daniela Glätzle-Rützler, Wanda Mimra, Thomas
Rittmannsberger, and Christian Waibel[†]

* We thank participants attending the 7th Workshop in Behavioral and Experimental Health Economics in Innsbruck, the 5th ATHEA Conference in Vienna, the SABE 2020 Annual Conference, the EuHEA 2020 Seminar Series, and the dggö Workshop for their helpful comments. Especially, we want to thank Geir Godager and Nadja Kairies-Schwarz for their discussion and input. Financial support from the Nachwuchsförderung at the University of Innsbruck as well as from the Austrian Central Bank (Jubilaefonds Project 17805) is gratefully acknowledged.

[†] Angerer: UMIT: Private University for Health Sciences, Medical Informatics and Technology; *silvia.angerer@umit-tirol.at*

Glätzle-Rützler: University of Innsbruck; *daniela.ruetzler@uibk.ac.at*

Mimra: ESCP Business School; *wmimra@escp.eu*

Rittmannsberger: University of Innsbruck; *thomas.rittmannsberger@uibk.ac.at*

Waibel: ETH Zürich ; *cwaibel@ethz.ch*

Abstract

In this paper, we experimentally investigate the effect of public consumer ratings on market outcomes in credence goods markets. Contrary to search or experience goods, consumers cannot evaluate all dimensions of trade for credence goods, which may inhibit the information and reputation-building value of public rating systems. We implement a healthcare market frame in which physicians as experts have an informational advantage over patients with respect to the appropriate treatment. The rating system takes the form of a five-star rating system as is common on online rating websites. The value of this rating system is compared in two different expert market settings: First, one in which patients cannot rely on information from personal experience with the expert, reflecting markets in which consumer-expert interactions are often first-time and infrequent (e.g. specialist visits). Second, one in which patients have personal experience with the expert, reflecting markets in which consumer-expert interactions are frequent and repeated (e.g. general practitioner visits). We find that the public rating system significantly improves market outcomes. Furthermore, a public rating system is a good substitute for personal experience information in terms of market efficiency and consumer surplus. Combined, however, we find no complementarity between public ratings and personal experience information, mainly due to the already high market efficiency in the presence of either one.

Keywords: Credence goods, expert behavior, ratings, feedback, laboratory experiment

JEL classification: C91, D82, I11, L15

1. Introduction

In 2019, OECD countries, headed by the United States and Germany, spent, on average, some 8.8 percent of their GDP on healthcare (OECD, 2021). According to Brown & Clement (2018), sizable parts of these expenditures are unnecessary¹ and can be attributed to physician misconduct (FBI, 2011). One problem in healthcare markets that may be contributing to the above is *informational asymmetries* between patients and physicians: While physicians are experts concerning the appropriate quality of service, patients typically do not know which treatment they need. Often, patients cannot verify the adequacy of the provided service even ex-post.

Services (or goods) with these properties are referred to as “*credence goods*”, as customers heavily rely on the advice of experts (Darby & Karni, 1973; Dulleck & Kerschbamer, 2006). Credence goods markets, such as financial services, repair services, legal advice, and healthcare services, can result in significant inefficiencies depending on the financial incentives and market institutions. In such markets, experts may have incentives to *overtreat* by providing unnecessary services, *undertreat* by providing insufficient services, or *overcharge* by billing for services that were not provided.²

Healthcare services, in particular, have garnered significant attention due to their societal and economic significance. One of the first papers in the field by Gruber & Owings (1996) demonstrated that healthcare providers respond to financial incentives. This assertion has been further corroborated by a mounting body of empirical evidence, indicating that physicians and other healthcare professionals react to financial incentives with potentially adverse welfare effects (Baker, 2010; Iizuka, 2007; Anthun et al., 2017; Barros & Braun, 2017; Batty & Ippolito, 2017; Clemens & Gottlieb, 2014; Dafny, 2005; Dunn & Shapiro, 2014; Geruso & Layton, 2019; Januleviciute et al., 2016; Parkinson et al., 2019; Shigeoka & Fushimi, 2014; Dai et al., 2017; Chao & Larkin, 2022). Undertreatment, for instance, has been shown in the area of pain management (Pasero & McCaffery, 2001), for the introduction of a fixed-price prospective payment system

¹ Brown & Clement (2018) categorize 1.52 million healthcare services administered between July 2015 and June 2016 in Washington state into 3 categories (necessary, likely wasteful, and wasteful) and conclude that 44% of those are deemed wasteful, amounting to excess spending of \$258 million (33% of the total \$785 million spent on health care services).

² Balafoutas & Kerschbamer (2020) provide a comprehensive review of the recent literature on credence goods.

(Cutler, 1995) as well as for uninsured patients visiting a hospital after a severe car accident (Doyle, 2005). Evidence for overtreatment is provided by Gottschalk et al. (2020) in a recent field experiment in the dental care market, where every fourth dentist visit resulted in the recommendation of unnecessary fillings. Overcharging happens for instance through upcoding in DRG-based hospital reimbursement systems³ (Cook & Averett, 2020; Jürges & Köberlein, 2015). Further field experimental support for biased expert decisions in healthcare markets is provided by Chen & Goldman (2016), Currie et al. (2014), Currie et al. (2011), Das & Hammer (2007), Das et al. (2016), and Lim et al. (2002).

As asymmetric information is the source of inefficiency in credence goods markets, providing information to customers can potentially alleviate these inefficiencies (Domenighetti et al., 1993). However, this depends on the nature of information in light of the fundamental problem that certain dimensions of expert and service quality cannot be judged even after consumption. This paper analyzes the effects of an important and increasingly prominent form of information in credence goods markets, a public rating system of experts.

Feedback platforms like Yelp, Google, TripAdvisor, Uber, etc., where consumers can rate their experiences with an expert, gain more and more popularity in recent years. To give an impression, Yelp counts approximately 28 million monthly users and has accumulated over 214 million customer reviews since its introduction in 2004, nine percent of them in the area of healthcare (Yelp, 2020). These platforms provide consumers with relevant information when choosing experts such as physicians (Xu et al., 2021). The majority of people in developed countries are aware of physician rating websites and many of them have already used them, to rate and find (new) physicians (Emmert & Meszmer, 2018; Hanauer et al., 2014; McLennan et al., 2017; Hedges & Couey, 2020). Given the widespread utilization of physician rating websites and the sparse empirical evidence on their effectiveness to improve market outcomes, studying this particular form of information—previous consumers’ feedback in the form of an expert rating—is of importance in healthcare credence goods markets.

Public rating information is however not the only information available to a patient before

³ Diagnosis-related group (DRG) is a case classification system for the reimbursement of inpatient care.

deciding to visit an expert. The patient may have consulted the expert before and thus have some previous experience with a particular expert. For instance, patients typically are in a repeat interaction with general practitioners. On the other hand, numerous specialized medical appointments occur rarely or only once, which means that patients may not be personally acquainted with the specialists and may solely rely on publicly available rating information, if any.⁴ In this paper, we analyze the value of a public rating system of experts in these two different market environments, when consumers have access to personal experience information with an expert, and when this is not the case. In particular, the experimental design allows for comparing these different types of information as well as analyzing their interaction.

We do so in a credence good laboratory experiment with a healthcare market frame. Experts (labeled as physicians in the experiment) and consumers (patients) interact over 16 periods in a classic credence goods market set-up in which experts have short-term incentives to undertreat and overcharge patients.⁵ In particular, a patient has a problem that needs to be treated but does not know the severity of it. Experts can costlessly diagnose the problem and provide and charge for either a minor or a major problem. In this setting, information about past expert behavior may allow reputation for quality equilibria to emerge. The focus of this paper is in particular how information in the form of a public rating system provides these reputation incentives, on a stand-alone basis and in comparison to personal experience information.

To keep the set-up simple and focus on the reputational effects of ratings, we fixed the prices and therefore mark-ups for the treatments in the experiment.⁶ Besides shutting down the potentially confounding effects of price competition, this is also in line with the fact that prices are heavily regulated in healthcare markets. The public rating system is implemented as the patient's choice to give feedback on a zero to five-star scale. In particular, after having received treatment from an expert and being charged a price, patients observe their payoff and can decide to provide a rating. These ratings are then averaged and provided to patients before their

⁴ The degree to which a physician has more repeated interactions compared to first-time or one-shot interactions depends among others on the specialty of the physician. Physicians performing rare examinations (e.g. radiologists doing MRI or CT scans) will have more first-time or one-shot interactions compared to GPs for instance.

⁵ They could also overtreat, but given the parametrization, this is dominated by simply overcharging instead of overtreating.

⁶ This follows from the result of [Mimra et al. \(2016\)](#), namely that competition for prices undermines reputation-building incentives for experts in credence goods markets.

decision to visit an expert in the next period in the rating conditions.⁷ To distinguish between markets with and without personal experience information, experts can be either identified by a fixed ID (personal experience conditions) or not. Thus, in the latter cases, personal experience—payoffs from previous interactions—cannot be attributed to given experts and thus not used to select and thereby incentivize particular experts.

We find that a public rating system significantly improves market efficiency and consumer surplus: Compared to a baseline in which neither a public rating system nor personal experience is available, both undertreatment and overcharging decrease significantly. The latter result is particularly interesting, as contrary to undertreatment, overcharging cannot be detected by patients. Furthermore, we find that a public rating system is a good substitute for personal experience information: Market efficiency and consumer surplus are on the same levels in markets with a public rating system compared to personal experience markets. Thus, in expert markets that are characterized by many first-time or infrequent interactions such as specialist visits in which consumers cannot rely on their own past experience to choose experts, a public rating system proves to be a well-functioning information alternative even in credence goods markets. Finally, we do not find complementarity between public rating and personal experience information when combined: Market outcomes do not improve further. However, this might be due to the fact that efficiency is already at a very high level when either type of information is available to choose and incentivize experts.

Our main contribution is to provide causal evidence on the effectiveness of a public rating system in credence goods markets. To the best of our knowledge, there exists no study systematically investigating the effect of public rating systems on expert behavior in a credence goods setting and no study that disentangles the effect of public rating systems for the two different market environments and forms of information. Recent research on experience goods suggests that, while public rating systems are beneficial in the first situation ([Tadelis, 2016](#)), they do not carry many additional benefits when market participants draw on personal relationships ([Cai et al., 2014](#)). Little is known, however, on the effectiveness of public rating

⁷ Throughout, we use the term condition for experimental treatments so as to not create confusion with the standard credence goods terminology of a treatment given by the expert to solve the consumer's problem.

systems in credence goods markets in general and healthcare settings in particular.

In contrast to observational data, a controlled laboratory experiment provides the advantage to observe the patient's "true" health problem and therefore unambiguously classifying expert behavior. Additionally, it allows testing the effect of introducing a public rating system on market outcomes such as efficiency and consumer surplus. Even though the setting does not take into account all factors of an expert-patient relationship, the laboratory offers a testbed for introducing institutions without putting the health of real-world patients at risk. Furthermore, disentangling reputational incentives in the two different market environments is difficult using observational data as these are not cleanly separated, which is another motivation to take the problem to the lab.

2. Related literature

Following the pioneering works on credence goods markets by [Darby & Karni \(1973\)](#), [Dulleck & Kerschbamer \(2006\)](#), and [Dulleck et al. \(2011\)](#), several studies set out to analyze the impact of different institutions such as competition, reputation, second opinions, price regulations, insurance coverage, new media, or monitoring. The papers conclude that several institutions could potentially mitigate inefficiencies in credence goods markets ([Angerer et al., 2021a](#); [Balafoutas et al., 2013](#); [Balafoutas & Kerschbamer, 2020](#); [Balafoutas et al., 2017](#); [Huck et al., 2016a](#); [Kerschbamer et al., 2016, 2017](#); [Liu et al., 2021](#); [Mimra et al., 2016](#); [Rajgopal & White, 2019](#)). In what follows, we shortly introduce and discuss studies investigating the impact of reputation.

Following the seminal papers by [Klein & Leffler \(1981\)](#), [Kreps et al. \(1982\)](#), and [Shapiro \(1982\)](#), a large and growing body of literature has investigated the effects of direct and indirect reputation in experience goods markets (e.g., [Bar-Isaac & Tadelis, 2008](#); [Bohnet & Huck, 2004](#); [Bolton et al., 2004](#); [Ely et al., 2008](#); [Ely & Välimäki, 2003](#); [Tadelis, 2016](#)). There are three papers on experience goods closely related to our present work by [Bohnet & Huck \(2004\)](#), [Huck et al. \(2012\)](#), and [Huck et al. \(2016b\)](#). Letting subjects play a binary-choice trust game for 20 periods,

Bohnet & Huck (2004) find that direct reputation is more effective in promoting trust than indirect reputation. Extending their model, allowing for competition between trustees, Huck et al. (2012) conclude that competition, coupled with direct reputation, helps eliminate market misconduct completely. However, Huck et al. (2016a) show that incentives for reputation-building are diminished once trustees start competing over prices.

The key difference between trust games and markets for credence goods is that, although participants in trust games have asymmetric information ex-ante, information is symmetric ex-post, whereas credence goods markets are characterized by persistent information asymmetries. Due to this, reputation-building may be impeded in credence goods markets, as experts have no way of unambiguously signaling trustworthiness to potential customers. The notion of credence goods was first introduced by Darby & Karni (1973). In their seminal paper, Dulleck & Kerschbamer (2006) provide a unifying theoretical framework and investigate the effectiveness of different institutions in markets for credence goods, among others (direct) reputation and competition, tested experimentally in Dulleck et al. (2011) under flexible prices. They find that, while competition drives down prices, therefore benefitting customers, it does not enhance overall market efficiency as undertreatment, overtreatment, and overcharging rates do not improve, compared to a situation without competition. Neither (direct) reputation nor a combination of (direct) reputation and competition influences relevant market outcomes under flexible prices. Conducting a field experiment in the U.S. market for auto repairs, Schneider (2012) concludes that reputation does not improve market outcomes in credence goods. In a recent literature review, Balafoutas & Kerschbamer (2020) find that the impact of competition and reputation on expert behavior in credence goods markets is at best ambiguous. The paper on credence goods closest to the present study is by Mimra et al. (2016). They experimentally investigate the role of reputation in markets under different price regimes (price competition and fixed prices) and with two forms of reputation mechanisms (private and public histories). Under private histories, customers receive information on posted prices, charged prices, whether undertreatment occurred, and their period payoff for their own previous interactions with an expert. Under public histories, customers receive this information for all previous interactions of an expert including their own. Note, that no environment without the possibility to build

a direct reputation is studied. The authors find that, regardless of the underlying reputation mechanism, undertreatment is significantly higher in markets with price competition compared to those under fixed-price regimes. Reputation through public histories has no impact compared to private histories in either of the price regimes. They conclude that price pressure undermines reputation-building, explaining why regulating prices may increase patient welfare in credence goods markets.

Our main contribution to the existing literature on institutions in credence goods is that we experimentally test how a public rating system of experts, where customers can rate interactions with experts on a five-star rating scale, influences outcomes under a fixed-price regime. We can thereby distinguish the effect in two relevant market settings, markets of first-time interactions without personal experience information and those in which customers have personal experience information.

More recently and following the rise in online markets (such as [ebay](#), [Amazon](#), etc.), there has been an increased interest in electronic reputation systems ([Bolton et al., 2004](#); [Resnick & Zeckhauser, 2002](#); [Resnick et al., 2006](#); [Rice, 2012](#); [Cabral & Hortaçsu, 2010](#); [Moreno & Terwiesch, 2014](#); [Ba & Paul, 2002](#); [Dellarocas, 2006, 2003](#)). Online markets lacked traditional reputation, but electronic reputation systems were designed to enhance trust and cooperation and to facilitate the exchange of information about the quality and reliability of market participants. Consumers can provide feedback on sellers' goods/services, creating aggregated ratings that reflect the seller's past performance and allow them to build a reputation. There is a growing body of research on electronic reputation mechanisms in experience goods markets, with studies examining their effects on market outcomes such as prices ([Ba & Paul, 2002](#); [Moreno & Terwiesch, 2014](#); [Resnick et al., 2006](#)), trading volume ([Cabral & Hortaçsu, 2010](#); [Moreno & Terwiesch, 2014](#)), and seller performance ([Rice, 2012](#); [Bolton et al., 2004](#)). Some studies have shown that reputation systems can reduce information asymmetry and increase trust ([Dellarocas, 2003](#)), and increase competition among sellers ([Cabral & Hortaçsu, 2010](#)). The findings in this literature are mainly based on laboratory experiments where students play a trust game ([Rice, 2012](#); [Bolton et al., 2004](#)), field experiments on online trading platforms such as ebay.com ([Resnick & Zeckhauser, 2002](#); [Resnick et al., 2006](#)), analyzing observational data from such plat-

forms (Dellarocas, 2005; Cabral & Hortaçsu, 2010; Ba & Paul, 2002). Over the past few years, many rating platforms, were introduced for *offline* markets which enable consumers to provide feedback and rate the expertise of providers across various goods and services markets. These platforms have become particularly relevant in credence goods markets, such as healthcare, repair, and legal services.⁸

While feedback systems have been shown to have a positive impact on experience goods markets, it remains an open question whether ratings will be as effective in credence goods markets. This is due to the fact that consumers are unable to determine whether the quality of the product or service provided was suitable. A recent study by Kerschbamer et al. (2019) suggests that consumers benefit from rating platforms in the computer repair market. However, their results are based on observational data and cannot account for reverse causality.

Our main contribution to the literature on electronic reputation and feedback systems is that we expand it to credence goods markets and experimentally test the value of a public rating system of experts. In addition, we can do so in two different market settings, one with, and one without personal experience.

Another strand of literature related to the underlying study is the emerging literature on experimental health economics. The seminal paper by Hennig-Schmidt et al. (2011) compares physician behavior under different payment schemes. Medical students act as physicians, choosing treatment quantities, while patients are not present in their laboratory experiments.⁹ Their results and several additional laboratory experiments that followed investigating different payment schemes (fee-for-service, capitation, pay-for-performance, or mixtures of them) indicate that physicians respond to financial incentives, as they overtreat under fee-for-service and undertreat under capitation (Brosig-Koch et al., 2016, 2013, 2017b; Green, 2014; Lagarde & Blaauw, 2017; Brosig-Koch et al., 2017c). These results are similar for real physicians, medical-, and non-medical students (Brosig-Koch et al., 2016). Other laboratory experiments in the context of health economics look at the impact of insurance (Huck et al., 2016a), performance

⁸ See for example www.jameda.de, www.yelp.com, or www.lawyers.com.

⁹ Physicians' choices have consequences for real patients outside the lab, as the money corresponding to the benefits of the lab-patients was given to Christoffel Blindenmission charity, caring for real patients.

disclosure ([Godager et al., 2016](#)), non-monetary incentives ([Kairies & Krieger, 2013](#)), professional norms ([Kesternich et al., 2015](#)), competition between healthcare providers ([Brosig-Koch et al., 2017a](#); [Han et al., 2017](#)), and whether teams of decision-makers decide differently than individuals ([Han et al., 2020](#)). For a comprehensive review of behavioral experiments in health economics see ([Galizzi & Wiesen, 2018](#)).

The main difference between our study and these earlier studies is the active decisions of subjects in the role of patients. Therefore, we can study interactions between patients and experts and the dynamics. Patients in our experiment can (i) decide whether to consult an expert and, in the rating conditions, (ii) rate interactions with experts on a five-star rating scale. This allows us to investigate the impact of ratings on expert behavior in a controlled laboratory experiment.

Lastly, we relate to the evolving literature on the value and reliability of (online) rating mechanisms in healthcare markets. A considerable amount of studies looked at the association between online physician ratings and other quality measures. While some find associations between them ([Lu & Rui, 2018](#)), others don't ([Saiffee et al., 2019, 2020](#)). Conducting a systematic literature review, [Hong et al. \(2019\)](#) conclude that the relationship between physician ratings and clinical outcomes is at best weak. Interestingly, [Saiffee et al. \(2020\)](#) argue that they perform poorly, especially in disciplines characterized by extensive credence goods nature (e.g., chronic disease care) because there it is particularly difficult for patients to assess the effectiveness of a particular physician accurately, given the long treatment-horizon.

3. Experiment

The experimental design is based on the credence goods framework of [Dulleck & Kerschbamer \(2006\)](#) and the seminal experiment by [Dulleck et al. \(2011\)](#). [Dulleck et al. \(2011\)](#) employed a neutral frame, in our experiment we chose to implement the framing of one of the most important credence goods markets, health care markets. The experimental instructions thus referred to expert sellers as physicians and consumers as patients, and the service for which

there is asymmetric information is a treatment for a health problem. This framing is applied based on the insights of [Kesternich et al. \(2015\)](#), [Kairies-Schwarz et al. \(2017\)](#), [Reif et al. \(2020\)](#), and [Angerer et al. \(2021b\)](#) who explore the effect of different framings in economic laboratory experiments. Throughout the paper, we will use the wording of ‘expert’ on the one side and interchangeably ‘patient’ or ‘consumer’ for the other market side.

3.1. The basic set-up and parameterization

In the basic setup, experts and patients are grouped in a market of eight subjects, four patients, and four experts. Patients suffer from a major health problem with probability $h = 0.5$ and a minor one with probability $(1 - h)$. The probability h is common knowledge. Patients decide whether to consult an expert knowing that they suffer from some problem in every period. They do not get information about the severity of their problem. Experts diagnose their patients’ problems with certainty and at zero costs. They provide one of two treatments, a major treatment (q_H) or a minor treatment (q_L). The cost for the expert to provide the major treatment is 6 ECU.¹⁰ The cost for the minor treatment is 2 ECU. Treatment prices, paid by the patients, are either 8 ECU (p_H) or 3 ECU (p_L) respectively. The major treatment cures both, the major and the minor health problem, while the minor treatment only cures the minor one. Patients obtain 10 ECU (v) if cured, and zero if treated insufficiently. The payoff for patients consulting an expert is the difference between the obtained value and the price charged (p_H or p_L). For experts, the payoff is the spread between the price charged (p_H or p_L) and the cost for the chosen treatment.¹¹ In case a patient decides against consulting any expert, the patient receives an outside option of (-4) ECU (o_{Pat}). Experts receive $o_{Exp} = 0$, if they do not interact with any patient in a given period. Compared to the framework of [Dulleck et al. \(2011\)](#), our basic model differs in two dimensions. First, the outside option of patients is negative ($o_{Pat} = -4$) illustrating the disutility of an uncured (health) problem.¹² Second, p_H and p_L are exogenously fixed, which is common in many expert markets, notably in highly regulated healthcare markets. Through-

¹⁰ Experimental Currency Unit (ECU)

¹¹ Following [Dulleck et al. \(2011\)](#), we assume large economies of scope between diagnosis and treatment. Hence, patients who decide to consult a physician commit to undergo treatment by this physician.

¹² This negative outside option ensures market interaction in order to investigate the effect of ratings.

out the experiment, there is neither verifiability nor liability, allowing us to investigate both undertreatment and overcharging.

The structure of the stage game is as follows (see Figure A1 in Appendix A for an illustration of the game in extensive form):

1. For each patient, nature draws the type of problem. With probability h patients have a major problem, and with probability $(1 - h)$ patients have a minor problem.
2. Patients decide whether to consult an expert. If patients decide not to visit an expert, the period ends. Otherwise, they choose one expert from a list of four.¹³
3. Experts costlessly diagnose the problem, provide a treatment (q_H or q_L), and charge a price (p_H or p_L).
4. Patients and experts observe their payoff in the respective period.
5. In the conditions with a public rating system after learning the payoff for the respective period, patients decide whether to rate the interaction with the expert. If they decide to rate the interaction, they choose the rating on a scale between 0 and 5 stars which is shown to the expert.

The stage game is played for 16 periods in all experimental conditions.

3.2. Experimental conditions

We employ a 2×2 factorial design to test the effect of a public rating system. The four experimental conditions are displayed in Table 1.

The value of a public rating scheme is analyzed and compared in two different expert market environments: First, a market environment in which patients can, over time, rely on their personal experience with a particular expert. Second, a market environment in which patients

¹³ Depending on the experimental condition, experts can be identified through a personal ID (in the personal experience conditions) and/or the average rating from previous periods is displayed at this stage for each expert (in the rating conditions).

Table 1: Experimental Conditions

Market Environment:			
Personal Experience with Expert			
		No	Yes
Public Rating	No	<i>Baseline</i>	<i>Experience</i>
	Yes	<i>Rating</i>	<i>Exp+Rating</i>

Note: In all our experimental conditions physicians compete for patients, i.e., patients choose one expert from a list of four if they decide to visit an expert.

cannot rely on their personal experience with a particular expert. The latter represents markets in which patient-expert interactions are often first-time and infrequent (such as specialist visits), whereas the former represents markets in which patient-expert interactions are more frequent and repeated (such as general practitioner visits). These two different market environments are implemented in the experiment as follows: In the experimental conditions without personal experience with experts, in each period patients choose one expert from a list of four without being able to identify them. All players are informed beforehand that patients have no means of identifying experts from previous periods. Thus, although patients observe their payoffs in each period and can partially infer expert behavior, they cannot attribute it to a particular expert and therefore cannot build up personal experience with a particular expert. In the experimental conditions with personal experience, patients can on the contrary identify experts by a fixed ID (physician 1, physician 2, physician 3, and physician 4) and decide whether to interact with a particular identified expert. Over the 16 periods of play, they can thus learn from their personal experience (payoffs) with a particular identified expert.

In the conditions with the public rating system (***Rating*** and ***Exp+Rating***), patients can choose to rate interactions with experts on a five-star rating scale after receiving their payoff in a given

period.¹⁴ This rating is shown to the respective expert at the end of the period.¹⁵ Subsequently, ratings for each expert over all treated patients are aggregated, averaged, and displayed to patients. Patients see these public ratings for all experts when they decide whether to interact and which expert to choose starting in period 5. In the condition without personal experience (**Rating**), as highlighted before, patients cannot identify a particular expert and only see the public ratings. The public ratings of all experts are displayed to experts when they decide on the type of treatment and which price to charge in a given period (see Appendix C for the screenshots showing the feedback information provided to patients and physicians).

In addition to the main experimental conditions shown in Table 1, we also ran four further conditions to be able to separate the role of expert competition, personal experience in the absence of expert competition, and private ratings (for a detailed description of the experimental conditions and the results see Appendices B and C). These control conditions will be explained in the corresponding results sections whenever they are used to disentangle effects in the main conditions.

3.3. Main outcome variables

Our main outcome variables describe expert behavior, patient decisions, and market efficiency. Table 2 lists these outcomes and provides their description and measurement for the results section.¹⁶

Expert behavior On the expert side, given the experimental set-up and incentives, undertreatment and overcharging are the relevant expert decisions. Undertreatment is defined as the consumer (patient) needing the major treatment q_H , but the expert providing the minor

¹⁴ In essence we model a single-dimension rating systems where patients can give one overall rating for every interaction. Note that many platforms have adopted multidimensional rating systems where patients can rate multiple dimensions, like waiting times, office environment, or physician knowledge, which seems to enhance rating informativeness (Chen et al., 2018).

¹⁵ We decided to inform the expert about the private rating to have full information provision about the rating to all participants irrespective of the history of play. To disentangle the effect of providing this information privately from the effect of the public disclosure of the average rating, see the results on the private feedback condition in Appendix B.

¹⁶ Section 3.5 lays out in more detail which expert and patient behavior can be supported in equilibrium in the different experimental conditions.

treatment q_L . An expert might have incentives to do so since the costs for the major treatment are higher (6 ECU versus 2 ECU) and the expert can always charge the price of the major treatment (8 ECU). In the results section, undertreatment will be reported in % of the expert-patient interactions in which patients need the major treatment. In terms of information, patients can detect undertreatment in a period ex-post via their payoff, as the problem is not cured. In particular, if the expert charged p_H , the patient payoff from undertreatment is -8 ECU.

Overcharging is defined as the expert charging the price of the major treatment (p_H) while only providing the minor treatment to a patient who has a minor problem. Overcharging is accordingly reported in % of the expert-patient interactions in which patients need minor treatment in the results section. In terms of information, patients cannot infer ex post whether they have been overcharged, as they might have had a major problem requiring the major treatment charged at p_H . Thus, an expert can 'hide' behind a major treatment problem when overcharging.

In principle, there is also scope for overtreatment, which would be providing the major treatment (q_H) and charging for it to a patient with a minor problem, but overtreatment is strictly dominated by overcharging for the parametrization. In particular, instead of providing the major treatment with costs of 6 ECU, for a patient with a minor problem, the expert can always only provide the minor treatment (costs of 2 ECU) and just (over)charge for the major treatment.

Patient decisions On the patient side, we record whether they choose to interact, and in the rating conditions whether they choose to provide a rating (captured by the variable feedback) and what the rating is (captured by variable rating). Given the low outside option, except for very high-risk aversion, patients should always choose to interact, which is intentional in this study to mimic credence good markets realistically. Our main focus of patient decisions will therefore be the ratings themselves.

Market outcomes We use two measures of market outcomes, overall market efficiency and patient surplus. Market efficiency is driven by interaction (allowing surplus generation) and

whether there is undertreatment, as undertreatment does not generate patient value. Given our parametrization, we expect high levels of interactions, such that market efficiency is primarily determined by undertreatment. We normalize market efficiency, with 0% for no interaction and 100% for an interaction with the correct treatment. Consumer surplus incorporates the prices paid by patients and is thereby influenced by overcharging, which is not the case for market efficiency. Consumer surplus is reported in absolute value.

3.4. Experimental protocol

We ran our experiment with 48 subjects in each condition. The sessions were conducted in the laboratory for experimental economic research at the University of Innsbruck. Overall, including the additional experimental conditions, 384 students participated. All sessions were run computerized using z-Tree ([Fischbacher, 2007](#)) and students were recruited using hroot ([Bock et al., 2014](#)). The project was approved by the internal review board of the University of Innsbruck. To ensure our target attendance of 24 participants (some sessions were run with 16 participants only), we invited 30 people to each session, however, dismissed all but 24 participants before starting the experiment. Those who did not get the chance to participate received a show-up fee of 4 Euros. At the beginning of each session, we explained the market setup to the participants, following a standardized protocol. An experimenter presented brief instructions to all subjects, covering the main features of the decision problem. Afterward, we asked subjects to read detailed instructions of the game and to answer a set of incentivized control questions (see Appendix E for the instructions and control questions). Once all subjects correctly answered the control questions, they were informed of their randomly assigned roles and played the credence goods game for 16 periods. At the end of the game, subjects participated in an individual risk preference task, a dictator game, a lying task, and a trust game. Finally, participants filled out a questionnaire (see Appendix F for the additional instructions and the questionnaire). The payment subjects received at the end of the session consisted of their profits from the credence goods game (4 randomly selected periods), one randomly selected additional task, and a lump sum payment of 2 Euros for answering the questionnaire.

Table 2: Main Outcome Variables

Expert behavior	Definition	Measurement	Notes
Undertreatment (UT):	Patient needs major treatment q_H , but expert provides minor treatment q_L .	As % of the expert-consumer interactions in which consumers need the major treatment.	Patient can detect undertreatment ex-post in a given period by a low payoff (-8 ECU if experts charged p_H)
Overcharging (OC):	Expert charges price of major treatment (p_H) but provides minor treatment q_L to a patient needing only the minor treatment q_L .	As % of the expert-patient interactions in which patients need minor treatment.	Cannot be identified ex-post by the patient via payoff
Overtreatment (OT):	Patient needs minor treatment q_L , but expert provides major treatment q_H .	As % of the expert-patient interactions in which patients need minor treatment.	For the expert, overtreatment is dominated by overcharging
Consumer decisions			
Interaction:	At the beginning of every period, consumers decide whether to visit an expert.	Relative frequency of consumer-expert interactions.	Given the chosen low outside option, patients should prefer to interact even when undertreated for the major problem unless they are strongly risk averse.
Feedback:	After every interaction with an expert, consumers decide whether they want to give feedback.	Relative frequency of giving feedback calculated as $\frac{\# \text{ of ratings}}{\# \text{ of interactions}}$.	
Rating:	Given that consumers give feedback, the rating given to expert on a scale from zero to five stars.	Average (expert) rating calculated as $\frac{\text{sum of ratings}}{\# \text{ of ratings}}$.	In a given period, the rating is calculated using all ratings up to this period. We distinguish public and private ratings: The public rating of an expert uses ratings from all patients, a private rating of a patient for an expert is using only the ratings of this patient
Market outcomes			
Market efficiency (EFF):	Overall realized market surplus	per possible interaction: 0% if there was no interaction, 100% if the patient was treated correctly, 0.25 (0.67) if the patient was undertreated (overtreated). Aggregated by averaging.	
Consumer surplus (CS):	Overall realized consumer surplus	in absolute value, per possible interaction: consumer value of provided treatment - charged price, or outside option. Aggregated by averaging.	

Note: Explanation of main outcome variables.

Subjects earned 24.54 Euros on average and sessions lasted approximately 120 minutes. For an overview of the sample characteristics, Table B1 (in Appendix B) provides descriptive statistics on the background information collected by experimental conditions.

3.5. Predictions and research questions

In this section, we discuss the main theoretical predictions and formulate our research questions. The analysis is based on the assumptions of rationality and, for simplicity, risk neutrality of experts and patients. The benchmark is condition **Baseline** in which patients can choose an expert, but cannot use information about past expert behavior in their expert choice as they can neither identify a given expert nor use rating information. This condition implements repeated first-time expert-patient interactions where patients choose between experts about whom no information is available.

The outside option of remaining untreated in **Baseline** and all other conditions is such that a patient prefers to interact: Even when undertreated in the case of the major problem, and always charged the high price p_H , the expected payoff from interacting (-3) is higher than the outside option (-4). This is different from other credence good experiments and reflects the important fact that in many credence goods markets, patients are better off seeing the expert in expectation. Without other-regarding preferences of experts, the unique equilibrium in the stage game is then that patients always interact and experts always undertreat and overcharge. As reputation-building of experts is not possible, the equilibrium over all periods is the repeated stage game equilibrium. If experts have social preferences such as altruism and efficiency concerns, they might however not always undertreat/overcharge. The results from **Baseline** allow to have an aggregate measure of these social preferences of experts given the market set-up.

Conditions **Rating**, **Experience** and **Exp+Rating** then allow patients to use information about past expert behavior, albeit through different channels. The basic patient information takes the following form: patients observe their payoff from an interaction with an expert, and can infer whether this expert undertreated them. Furthermore, if they are only charged price p_L (and not undertreated), they can even infer that the expert did not overcharge them. We will henceforth

call either of these two basic forms of information (no undertreatment, no overcharging) a positive patient experience.

In ***Experience***, patients can identify experts and have their own past experiences as information about the behavior of an identified expert. Punishing (rewarding) an expert by not visiting (re-visiting) the expert based on this personal experience information then allows for reputation equilibria to exist in condition ***Experience***. In these, experts build up a reputation for quality¹⁷ in early periods based on the following patient strategies: patients stay with an expert for which their belief about a positive experience is sufficiently high in early periods. Conversely, negative personal information leads to punishment by not (re)choosing the corresponding expert. In late periods, experts who provided a positive experience in early periods milk their reputation and are rewarded by patients staying with them and thereby allowing them to make (high) profits at/until the end. The reputation incentives for experts are thus a back-loaded remuneration, and this works as patients have earlier period personal experience information.

In ***Rating***, a reputation for quality equilibria may exist as well, albeit through a different channel. Patients cannot choose experts based on their own experience, but they have access to aggregated, indirect information from other patients' experiences. Interpreting this information requires a belief about the rating strategies of other patients. If they are such that patients believe that other patients give a high rating (on the 0-5 star scale) when they had a positive patient experience, then a higher rating leads to a higher belief about the expert providing a positive patient experience in early periods. The following patient strategies may then sustain a reputation for quality equilibria: patients give a high rating to an (unidentified) expert when they had a positive patient experience, and in the next period choose an (unidentified) expert with a high rating. In late periods, for rewarding experts with high ratings in early periods, patients continue to choose experts with the highest ratings in the reputation-milking phase, in order for these experts to keep their customers.

In terms of outcomes, the following types of a reputation for quality equilibria can be sup-

¹⁷ As the design is one of pure moral hazard, by reputation we mean reputation for quality and not a reputation for type.

ported.¹⁸ **Equilibria without undertreatment** in early periods, as well as **equilibria without undertreatment and without full overcharging** in early periods. These differ in whether patients punish experts by not re-visiting them (*Experience*) or giving low ratings (*Rating*) in early periods only when they receive a negative payoff (undertreatment) or also when they are charged the high price (p_H). The latter case is more complex as patients cannot distinguish between being overcharged or not: p_H is not overcharging when the patient had a major problem. Nevertheless, punishing when charged p_H can sustain equilibria without full overcharging in which experts undercharge in early periods.¹⁹

Thus, although via different channels, reputation equilibria may exist in both *Experience* and *Rating* as well as the combination (*Exp+Rating*). Whether they emerge and are more likely to prevail in *Experience* or *Rating*, or require the combination of both, is an empirical question and the motivation for taking the problem to the lab. The indirect information about expert behavior in credence goods markets from *Rating* might thereby be perceived as noisier and less reliable, as it depends on the rating strategies of other patients. In particular, the belief about expert behavior (reputation) depends on the beliefs about other patients' rating behavior. Conversely, the public rating might also be considered as containing more information and being more salient compared to personal experience. The 2×2 experimental design can provide results both on the effectiveness of a public rating system and whether public rating and personal experience are complements or substitutes for reputation-building. The focus of the analysis will thus be on the following research questions:

Research Question 1 *What is the impact of public rating information on expert behavior and market efficiency in markets without personal experience? (Baseline vs. Rating)*

Research Question 2 *Is public rating information a good substitute for personal experience information in terms of market outcomes? (Rating vs. Experience)*

¹⁸ The structure of these equilibria is described below. Appendix D shows the construction. There is a multiplicity of equilibria with e.g. different switching periods between reputation-building and reputation-milking. Of course, no reputation equilibria with full undertreatment and overcharging exist as well.

¹⁹ Compared to equilibria without undertreatment, those with additionally less than full overcharging have a shorter period of good expert behavior and a longer period of rewards as expert profits when undercharging are substantially lower.

Research Question 3 *Is public rating information a complement to personal experience information in terms of market outcomes? (**Experience** vs. **Exp+Rating**)*

Research Question 4 *Do patients react less strongly to public rating than to personal experience information? (Analysis of patient decisions in **Exp+Rating**)*

4. Results

In Section 4.1, we will first start with the comparison of aggregate results to answer the main research questions. To better understand the dynamics behind the aggregate results, and to confirm whether and how these results confirm our simple hypothesis of reputation-building for quality, we will analyze in turn ratings and patient decisions in more detail in Section 4.2.

Table 3 reports the aggregate results for the main experimental conditions averaged over markets and periods, with the corresponding non-parametric tests for the relevant experimental condition comparison. To complement the non-parametric results, Table 4 reports on the results from multilevel mixed-effects probit and linear regressions.²⁰ We ran two different models: The first model shows the effect of our experimental conditions when controlling only for time trends. In the second model, we control for economic preferences and personal characteristics relevant in a credence goods setting by adding experimental measures for social preferences, lying, trustworthiness as well as measures for personality traits alongside the standard socio-demographic covariates. Figure 1 displays the main results averaged over markets throughout the 16 periods.

4.1. The effects of a public ratings system

Column 1 of Table 3 shows that there is substantial undertreatment and overcharging in a market without either personal experience or public rating (**Baseline**): experts undertreat their

²⁰ Multilevel mixed-effects models are designed specifically to account for dependencies between observations on different hierarchical levels. In our case, we use a three-level mixed-effects model to account for the dependency of observations at the subject and/or market levels.

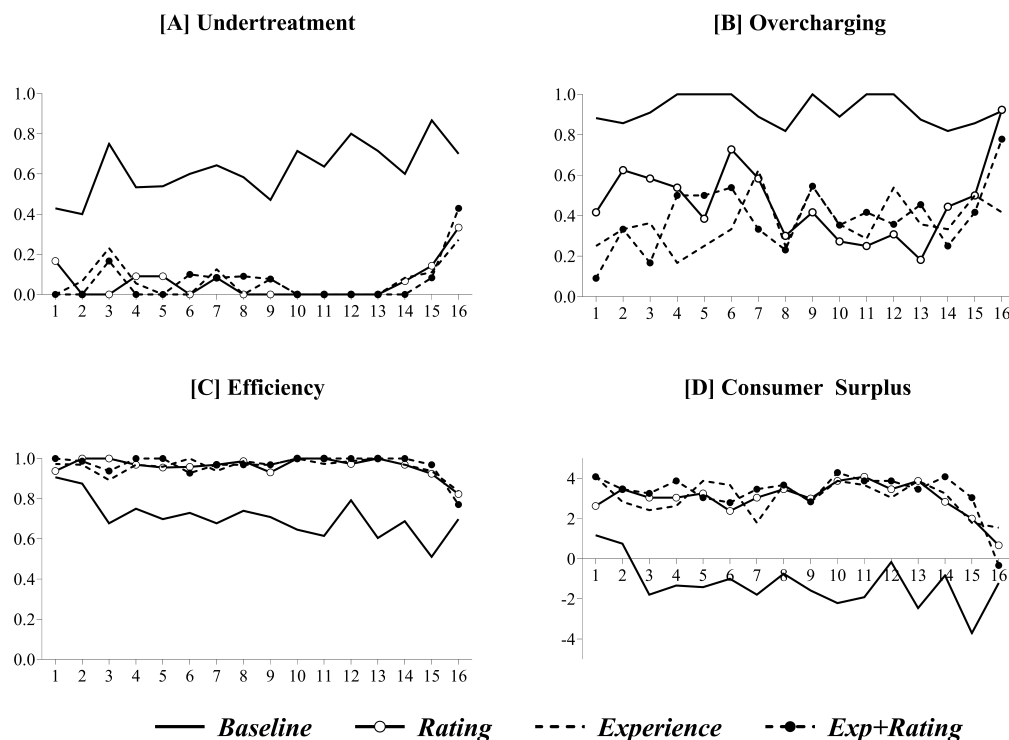


Figure 1: Rate of undertreatment, overcharging, efficiency, and consumer surplus by experimental conditions.

patients in 64.7% of all cases and overcharge them in 92% of all cases. Market efficiency, which is determined by undertreatment and interactions, is at only 70.7%. As a benchmark, full undertreatment with full interaction would lead to a market efficiency of 62.5%, as there is no efficiency loss for patients with the minor problem, and only no interaction could lead to a market efficiency below 62.5%.

The introduction of a public rating system (Column 2 of *Table 3*) leads to a sharp and significant decline in undertreatment, dropping from 64.7% in **Baseline** to only 5.8% in **Rating**. This induces a significant increase in market efficiency from 70.7% to 96.2%. Furthermore, overcharging also significantly decreases from 92% to 47.9%.²¹ This is a particularly interesting finding, as overcharging—contrary to undertreatment—cannot be directly observed by the patient. Despite the possibility to hide behind the probability of a major problem for the treatment of which the high price can be reasonably charged, the disciplining effect of ratings reduces this overcharging behavior. The reduction in both undertreatment and overcharging leads

²¹ These results also hold when restricting the comparison between conditions to the first eight, respectively the last eight periods.

Table 3: Overview of results (means).

	Markets without personal experience		Markets with personal experience		<i>p</i> -values of MWU ¹		
	<i>Baseline</i>	<i>Rating</i>	<i>Experience</i>	<i>Exp+Rating</i>	<i>Baseline</i> vs <i>Rating</i>	<i>Rating</i> vs <i>Experience</i>	<i>Experience</i> vs <i>Exp+Rating</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Expert behavior							
Undertreatment (in %)	64.72	5.81	6.81	6.89	0.002	0.794	0.777
Overcharging (in %)	92.03	47.94	36.89	38.07	0.002	0.093	0.974
Overtreatment (in %)	0.00	4.30	6.47	0.56	0.455	0.546	0.303
Consumer decisions							
Interaction (in %)	93.75	98.96	99.74	99.48	1.000	0.424	1.000
Feedback (in %)	-	93.62	-	88.786			
Star-rating	-	3.66	-	4.06			
Market outcomes							
Efficiency (in %)	70.70	96.21	95.42	96.85	0.002	0.849	0.959
Consumer Surplus (in ECU)	-1.27	3.01	3.05	3.30	0.002	0.937	0.485
Observations	384	384	384	384			

Note: We analyze six independent markets in every experimental condition. In each market, four patients and four experts interact. The experimental conditions are: **Baseline**, **Experience**, **Rating**, and **Exp+Rating**. Please refer to Section 3.2 for a description of the experimental conditions. See Table 2 for a description of the outcome variables.

¹ Mann-Whitney U-tests for pairwise differences between conditions with matching groups of 8 subjects as one independent observation (note that there are no significant differences between conditions **Rating** and **Exp+Rating** in any of our main outcome variables (see Table A1 for all pairwise comparisons)). *p*-values are adjusted for the small sample size, using Fisher's exact test.

to a substantial increase in patient surplus. These as well as all the following results from the nonparametric analysis on experimental condition comparison are confirmed in the regression analysis.

Result 1 *Introducing a public rating system into a credence good market (without personal experience) significantly decreases both undertreatment and overcharging and significantly increases patient surplus and market efficiency.*

One important question relating to the above result is whether it is the reputation-building

Table 4: Average Treatment Effects

	Undertreatment		Overcharging		Efficiency	Consumer Surplus
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted levels in <i>Baseline</i>	0.694 (0.092)	0.648 (0.130)	0.947 (0.039)	0.941 (0.032)	0.707 (0.068)	-1.266 (0.642)
<i>Marginal Treatment Effects</i>						
<i>Rating</i>	-0.601*** (0.100)	-0.607*** (0.134)	-0.399*** (0.051)	-0.415*** (0.039)	0.255*** (0.069)	4.271*** (0.686)
<i>Experience</i>	-0.600*** (0.104)	-0.617*** (0.136)	-0.492*** (0.056)	-0.553*** (0.080)	0.253*** (0.070)	4.318*** (0.718)
<i>Exp+Rating</i>	-0.575*** (0.100)	-0.583*** (0.133)	-0.512*** (0.072)	-0.529*** (0.060)	0.262*** (0.069)	4.563*** (0.676)
Period	+	+	+	+	-	-
<i>Additional Games</i>						
Amount donated to charity		-		-		
Liar (yes)		not sig.		not sig.		
Trustworthiness		not sig.		not sig.		
Covariates		✓		✓		
<i>p-values from post-estimation Wald-Test</i>						
<i>Rating</i> vs <i>Experience</i>	0.991	0.697	0.069	0.101	0.894	0.907
<i>Rating</i> vs <i>Exp+Rating</i>	0.402	0.505	0.093	0.089	0.572	0.365
<i>Experience</i> vs <i>Exp+Rating</i>	0.567	0.273	0.772	0.781	0.580	0.524
Observations	770		735		1536	1536
Number of Groups	24		24		24	24

Note: The table presents results from multilevel models with random effects at the market and individual levels (undertreatment & overcharging: columns 1-4) or at the market level for market efficiency (column 5) and consumer surplus (column 6). See Table 2 for a description of the outcome variables. We report effects as marginal effects, calculated as differences in the expected probabilities between the experimental condition in question and the baseline condition (Refer to Table A2 for the original regression output). Please refer to Section 3.2 for a description of the experimental conditions. All regressions include time trends. **Covariates:** Gender, age, BIG 5 personality traits (extraversion, agreeableness, conscientiousness, neuroticism, openness) measured with a 10-item BIG 5 questionnaire, whether the participant is a business/economics major, self-reported frequency of practicing religion, number of physician visits in the past 12 months, an indicator for experience with incorrect physician behavior, an indicator for experience with physician recommendations, relative school performance as a proxy for IQ, a measure for altruism (the amount donated to charity in a dictator game), an indicator whether the participant is classified as a liar (if reporting 4 or more correct dice rolls out of 12 in a lying task), and trustworthiness measured in a standard trust game. Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

mechanism that drives this result. While we will analyze ratings and rating dynamics in more detail in the next section, the comparison of *Rating* with a control condition in which patients provide a rating to the expert but in which ratings are not aggregated and publicly displayed

(condition **Rating-Priv**) shows that the reputation-building mechanism is very important: In the control condition, undertreatment goes down compared to **Baseline** (43% versus 64.7%) but is substantially higher than in **Rating** (5.8%). Furthermore, overcharging with 87% stays almost at the level of **Baseline** (92%) in the control condition **Rating-Priv**.

In the condition with personal experience but without a public rating system, undertreatment is at only 6.8% and overcharging is at 36.9% (Column 3 of *Table 3*). Compared to **Baseline**, we find a significant decrease in both undertreatment and overcharging which lead to significant increases in patient surplus and efficiency (p-value MWU: <0.01 all). When comparing **Rating** to **Experience**, we find that aggregate results on expert behavior are very similar: The undertreatment rate at 5.8% in **Rating** is almost the same as that in **Experience**. The overcharging rate in **Rating** at 48% is a bit higher than overcharging in **Experience** (36.9%), but this difference is significant only at the 10% level. There is no significant difference in either patient surplus or market efficiency. Thus, overall, a public rating system appears to implement similar market outcomes as a market in which patients can rely on personal experience information about experts.

Result 2 *There is no significant difference in market outcomes between **Rating** and **Experience**. With respect to overall market outcomes, the public rating system is a good substitute for personal experience information in the studied credence goods markets.*

We now turn to the question of the effect of a public rating system when patients can rely on information about expert behavior from their personal experience with the expert, in particular, whether there is a complementarity of personal experience and public rating information. Columns 3 and 4 of *Table 3* show that markets with personal experience, both without and with a public rating system, have a low level of undertreatment (6.8% and 6.9% respectively) and moderate levels of overcharging (36.9% and 38.1% respectively). For all variables of expert behavior as well as patient surplus and efficiency, there are no statistically significant differences between **Exp+Rating** and **Experience**. A crucial observation is that the scope of improvement of market outcomes by adding public rating information to personal experience

is quite limited, as outcomes—in particular undertreatment—are already close first best levels. Similarly, using Result 2, the vice versa observation for adding personal experience information to a public rating system is analogous. Taken together, we do not find a complementarity of public rating and personal experience information in our experiment, but this can be explained by an already high level of market performance in a market with either personal experience or public rating, which reduces the scope for complementarity.

Result 3 *Introducing a public rating system into a market in which patients have personal experience with experts neither improves (nor worsens) market outcomes. We find no complementarity between public rating and personal experience information with respect to overall market outcomes.*

In addition to the differences between conditions, the coefficient for the time trend shows that undertreatment and overcharging increase, while market efficiency decreases over time. Moreover, the regression results of the second model reveal that participants who are willing to give more money to a charity in a dictator game engage significantly less often in undertreatment or overcharging. We do however not find statistically significant effects for subjects classified as a liar in our lying task nor trustworthiness.

4.2. Patient ratings and expert selection

In this section, we explore patient behavior with a special emphasis on the ratings and expert choice.

In **Rating**, the large majority (93.6%) of interactions are rated and the average rating is 3.7 stars. Similarly, in **Exp+Rating** 88.7% of interactions are rated with an average rating of 4.1 stars. These average star-ratings hide a substantial differentiation by the patient payoff. *Figure 2* shows the average star-rating by patient payoff for the two experimental conditions with ratings.

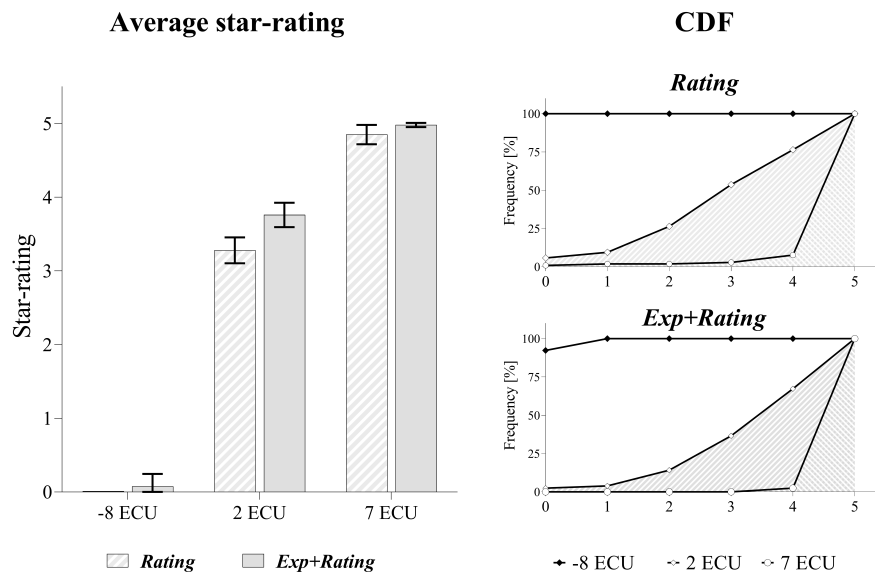


Figure 2: Rating behavior of consumers. On the left side, we see the means and 95%-CI of ratings for each of the possible payoffs of patients. The right side shows the cumulative distribution function (CDF) of given star-ratings, separately for possible payoffs of patients. If patients are undertreated, the payoff is -8ECU, whereas if they have a minor health condition and are treated appropriately, the payoff is 7ECU. In the case of a minor health problem and appropriate treatment but overcharging, or in the case of a major health problem and appropriate treatment with charges, the payoff is 2ECU.

When experiencing a negative payoff (-8 ECU) in a period, patients can infer that they were undertreated in this period.²² Figure 2 reveals that this leads to a rating of 0 stars. In fact, this was the case for all interactions except for a single one, in which the patient gave a rating of 1 star. Thus, undertreatment, which can be observed ex-post, leads to an unambiguous punishment with 0 stars that is symmetric across patients (Table 5).

When receiving a payoff of 7 ECU, patients can infer that they were neither undertreated nor overcharged.²³ In that case, 95% of interactions were rated with five stars. The most interesting part is the rating given for a payoff of 2 ECU: This payoff is generated when the patient either had a major problem and was appropriately treated and charged or when the patient had a minor problem and was overcharged. Thus, the expert can 'hide' behind a major problem and overcharge in case of a minor problem. The distribution of ratings for this case is more dispersed, with ratings in the two conditions ranging from 0 (4.25%) to 5 (27.74%) and a median

²² While undertreatment can be observed ex-post, this is not an experience good characteristic but remains a credence good characteristic as the severity of the problem is drawn randomly in each period.

²³ They do not know whether they had a severe or minor problem, so they cannot infer whether the expert might have even undercharged them.

rating of 4 stars. Interestingly, there is a statistically significant difference in the distribution of ratings for the 'ambiguous' payoff of 2 ECU between the conditions **Rating** and **Exp+Rating**: The ratings are better in **Exp+Rating** when patients can learn from personal experience with an expert (Two-sample Kolmogorov-Smirnov test: $D = -0.19$, $p < 0.01$).

Table 5: Ratings response

	Star-Rating (1)
Predicted star-rating if patient payoff is 7ECU	4.89
	<i>Marginal effects if ...</i>
... payoff is 2ECU	-1.393*** (0.142)
... payoff is -8ECU	-4.894*** (0.159)
Observations	695
Number of groups	12

Note: The table presents marginal treatment effects of multilevel models with random effects at the market and individual levels. The dependent variables are star-ratings following an interaction with an expert. Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Apart from ratings, it is essential for reputation as an indicator of quality to be effective that patients visit experts whom they anticipate will provide a high quality of care (positive patient experience). As highlighted in *Section 3.5*, the channel is staying with an expert in the personal experience conditions, and going to experts with the highest ratings in the rating conditions, conditional on symmetric strategies by the other patients. The latter, symmetric strategies where undertreatment is clearly punished with a bad rating and no undertreatment/no overcharging is clearly rewarded with a good rating seems to be the case. In **Baseline**, on the contrary, patients do not have information that can (re)direct them to experts for which they can expect high quality.

Figure 3 shows the frequency of a change in expert by the patient payoff. The line corresponds to the expected frequency associated with random assignment among the 4 experts in a market (75%). *Figure 3* nicely illustrates that in **Baseline**, as experts cannot be identified, patients cannot change intentionally and thus cannot provide incentives via their expert visit decisions.

In **Rating**, while patients cannot identify experts, they can decide to stay with best-rated experts when they receive a high payoff, and this can explain the lower frequencies of change in **Rating** compared to **Baseline** for patient payoffs 2 ECU and 7 ECU. However, and in line with intuition, the reaction of patients is strongest in markets with personal experience where experts are identified.

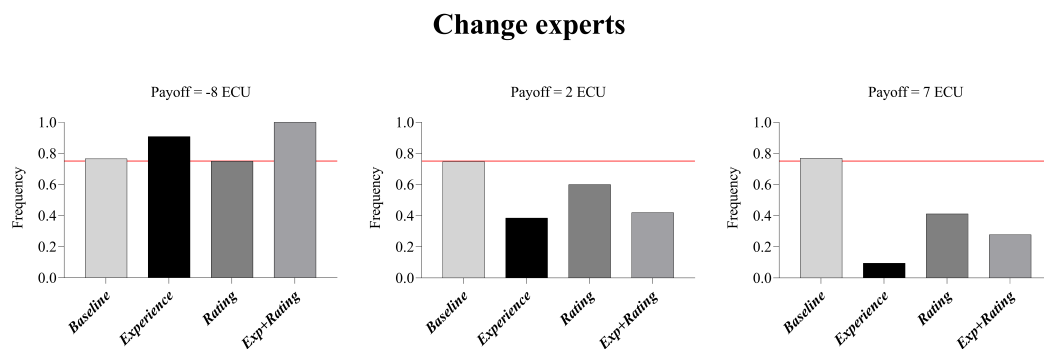


Figure 3: Frequency of a change in expert by realized consumer payoff in a given period.

The regression analysis shown in *Table 6* confirms that the decision to change the expert depends strongly on the patient payoff, and the direction of this effect is consistent with what was expected. While we show the frequency of change also for **Baseline** and **Rating** in *Figure 3*, the results on the decision to change reported in *Table 6* are based only on conditions **Experience** and **Exp+Rating** to account for the fact that patients can only fully intentionally leave a given expert in these two experimental conditions.

Figure 4 shows expert visits depending on their ranking with respect to both public and private ratings, and *Table 7* provides the corresponding regression results. The private rating of a patient is the average rating that the patient gave to the expert up to the corresponding period. We distinguish four categories of visits: whether the expert visited was the highest ranked in both public and private rating, had the highest public but not private rank, the highest private but not public rank, or did not fall in either of the previous categories (other). For interpretation, it is important to note that both public and private ratings are only explicitly available to a patient in **Exp+Rating**. For this reason, we speak of realized expert visits but not choice/selection. In **Rating**, the private rating is implicit as patients cannot attribute it to

Table 6: Associations between payoffs of patients and their decision to change the expert.

	Change Expert (1)
Predicted frequency if payoff of patient is 7ECU	0.19
<i>Marginal effects if ...</i>	
... payoff is 2ECU	0.213*** (0.038)
... payoff is -8ECU	0.680*** (0.092)
Observations	719
Number of groups	12

Note: For this analysis, only the treatments **Experience** and **Exp+Rating** are considered. The table presents results from three-level a model with random effects at the market and individual levels. The dependent variable is a binary indicator of whether a patient changed the expert. We report effects as marginal effects, calculated as differences in the expected probabilities between the payoff in question and the maximum payoff of 7ECU. Robust standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

a given expert. For the condition **Experience** in which patients do not rate experts, we have constructed a hypothetical private and public rating of experts given their choices based on the corresponding average ratings for the same choices in **Rating**.²⁴ We also display the results for the control condition **Exp+Rating-Priv**, which is the same as **Experience** except that patients give a private rating to experts which can be used as a private rating and aggregated to a public rating. It is reassuring to see that the distribution of patients' expert visits according to ranks looks almost identical in both conditions.²⁵

The first observation from *Figure 4* is that the majority of patients select the publically best-rated experts in **Rating** and **Exp+Rating**. Interestingly, in all conditions, visits with experts that had both the top public and private rating ranks are the most frequent.²⁶ Thus, even though both private and public rating are not available in all conditions, the feedback information available in the respective condition effectively channels patients to the individually and publicly best-rated experts. Furthermore, *Table 7* shows that the shares of expert visits for which the private rank but not public rating rank is highest are significantly higher in all

²⁴ To calculate the average rating of various patient-payoffs, we use the decisions of patients in **Rating**, where realized payoffs correspond to 0 stars (-8 ECU), 3.28 stars (2 ECU), and 4.85 stars (7 ECU).

²⁵ Note that we reconstruct hypothetical ranks in **Experience**, while ranks are based on actual ratings in **Exp+Rating-Priv**.

²⁶ We also do not find differences in these visit shares between **Rating**, **Experience**, and **Exp+Rating**.

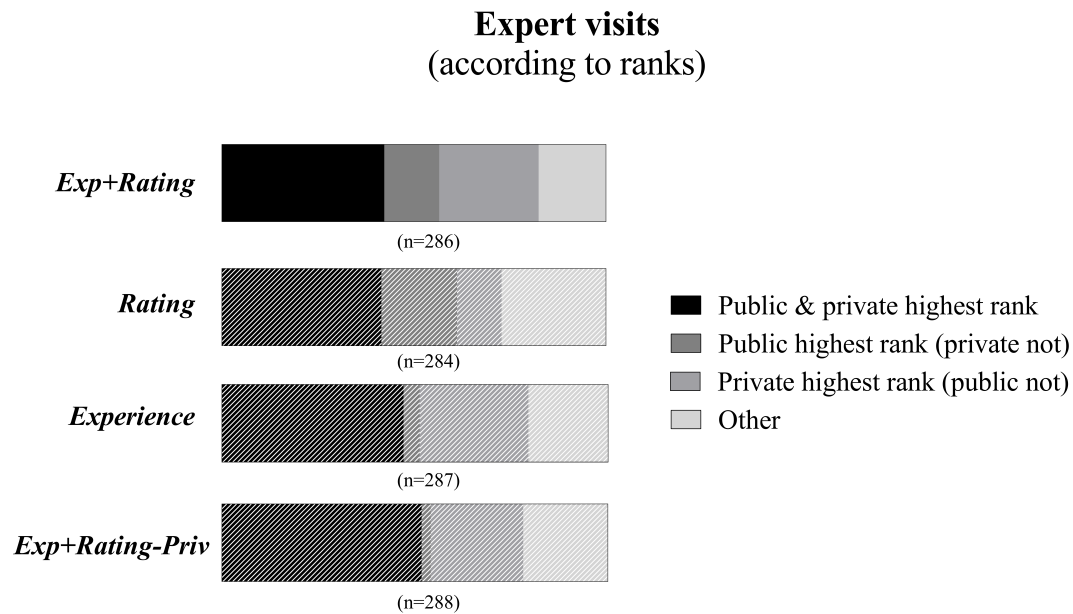


Figure 4: Distribution of patients' realized expert visits according to relative private and public rankings of experts. Results for *Rating*, *Experience* and control condition *Exp+Rating-Priv* are shaded as patients do not have full information on both private and public rankings in these conditions.

conditions with personal experience compared to *Rating*. Similarly, the shares of expert visits with the best public but not private rank are (weakly) significantly lower in the conditions with personal experience compared to rating.

An important question is which type of information is more relevant for patients' selection of experts when they have both personal experience information and public rating information available. To interpret the relative importance of personal experience vs. public rating information in expert choice, we look at *Exp+Rating* in more detail. Figure 5 shows the distribution of selected experts by the spread in public-private rank (left) and rating (right). Both distributions are left-skewed (left: -0.667 , $p < 0.01$ ²⁷; right: -1.410 , $p < 0.01$), revealing that patients, when selecting experts put more weight on their private experience than on the public rating.

Result 4 *The majority of patients select the best-rated experts in *Exp+Rating* and *Rating*.*

When there is a discrepancy between the private experience and the public ratings, patients seem to

²⁷ For this analysis, we exclude the 140 visits where ranks were equal and only considered interactions where there was a discrepancy between the private and the public ranking.

Table 7: Expert visits according to rank.

	<i>Public and private best</i>	<i>Public best (private not)</i>	<i>Private best (public not)</i>	<i>Other</i>
	(1)	(2)	(3)	(4)
<i>Levels in Rating</i>	0.415	0.197	0.116	0.271
<i>Marginal Treatment Effects</i>				
<i>Exp+Rating</i>	0.008 (0.041)	-0.054* (0.031)	0.143*** (0.032)	-0.096*** (0.035)
<i>Experience</i>	0.055 (0.042)	-0.155*** (0.026)	0.166*** (0.033)	-0.066* (0.036)
<i>Exp+Rating-Priv</i>	0.102** (0.041)	-0.173*** (0.025)	0.123*** (0.032)	-0.052 (0.036)
<i>p-values from post-estimation Wald-Tests</i>				
<i>Exp+Rating vs Experience</i>	0.254	0.000	0.527	0.348
<i>Exp+Rating vs Exp+Rating-Priv</i>	0.023	0.000	0.596	0.185
<i>Experience vs Exp+Rating-Priv</i>	0.259	0.240	0.244	0.699

Note: The table presents results from a multinomial logistic regression. We report the predicted frequencies of choosing experts based on public and private ranks in **Rating** and the difference between **Rating** and other experimental conditions as marginal effects. E.g. patients in **Rating** choose the private (but not public) best-rated expert in 11.6% of cases (column 3), while patients in **Exp+Rating** do so significantly more often, in 25.9% of the cases. See Figure 4 for an illustration of the distribution of physician choices.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

put more weight on information from private experiences when choosing experts in Exp+Rating.

4.3. Ratings and expert profits

We can now turn to the impact of public ratings on expert outcomes. Table 8 shows the relationship between average public expert ratings and the sum of interactions, starting in period five, for all the conditions with public ratings pooled.²⁸ Aggregating and confirming the previous results, we find that higher public ratings are rewarded with a significantly higher number of interactions.

²⁸ We exclude the first four periods from our analysis since patients could only see public ratings from the fifth period onwards

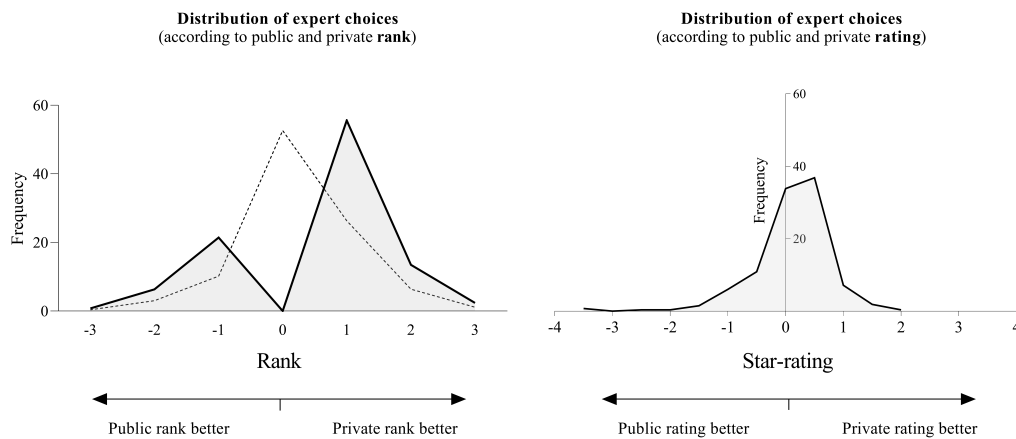


Figure 5: Distribution of selected experts by the spread in public-private rank (left) and rating (right). **Left:** The dashed line shows the distribution of choices according to ranks including equal ranks. We observed 266 interactions where patients chose an expert for whom they had both, private experiences, and a public rating. Of those, patients selected an expert with equal ranks in 53% (140 interactions). The solid line (shaded area) only shows the distribution of selected experts when there was a discrepancy between the private and the public ranking. Testing for normality reveals that the distribution is significantly skewed to the left (-0.667 , $p < 0.01$). **Right:** we show the distribution of selected experts according to differences between the private and public average ratings (private average rating - public average rating). Hence, positive (negative) numbers indicate that the expert had a better private (public) rating. The distribution is significantly skewed to the left (-1.410 , $p < 0.01$).

Table 8: Associations between interactions and ratings.

	Sum of Interactions ²
Public Average Rating ¹	0.668*** (0.127)
Observations	558
Number of groups	12

Note: Two-level model with random effects at the market level. Robust standard errors in parentheses.

¹ Average expert rating at the beginning of a period calculated as $\frac{\text{sum of ratings}}{\# \text{ of ratings}}$

² Sum of interactions of an expert in a given round, starting in period 5.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Good ratings not only lead to more interaction but also weakly higher expert profits. The first model of Table 9 reports the results of a multilevel mixed-effects linear regression, where the experimental profit (sum of all profits generated by physicians over the course of the experiment) is regressed on the overall rating of expert (average of all ratings obtained over the course of the experiment). The second model of Table 9 explores the association between the profit an expert generated from an individual interaction and the corresponding rating the patient left for this interaction. As expected, a higher rating for an individual interaction is associated

with significantly lower profits for the interaction. However, the results show a weakly significantly positive relationship between overall ratings and the sum of experimental profits. Thus, at least weakly, building a good reputation pays off for experts.

Table 9: Associations between profits and ratings of experts.

	Overall Profit³	Profit from interaction⁴
	(1)	(2)
Overall Rating ¹	11.131*	
	(5.022)	
Rating ²		-0.615***
		(0.078)
Observations	48	695
Number of groups	12	12

Note: The table presents results from two-level models with random effects at the market levels. The first column explores the association between the total profit generated by experts and their average rating over all 16 periods. The second column explores the association between the profits of a expert and ratings for individual interactions. The second column includes time trends. Robust standard errors in parentheses.

¹ Average expert rating over the course of the 16 periods calculated as $\frac{\text{sum of all ratings}}{\# \text{ of all ratings}}$

² Rating for an expert from an interaction with a patient

³ Sum of profits generated by the experts over the course of 16 periods.

⁴ Profit for an expert from an interaction with a patient

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To disentangle the dynamics of the relation between ratings and expert profit over the course of the experiment, we regress the sum of experimental profits on the average ratings of experts in distinct blocks of periods (Table 10). We find that high average ratings generated in the first four periods are associated with high total experimental profits. This association weakens throughout the experiment until it does not reach statistical significance in the last four periods.

5. Conclusion

Online rating platforms have become increasingly common in recent years. Nevertheless, there is a lack of studies that investigate the causal effect of public feedback systems on market outcomes. In this paper, we experimentally investigate the effect of the prominent five-star rating system on expert behavior in a healthcare credence good market. The experiment thereby distinguishes between two different market environments: Those in which consumers (patients)

Table 10: Associations between profits and ratings of physicians generated in distinct phases of the experiment.

	Overall Profit			
	(1)	(2)	(3)	(4)
Average Rating ¹ (Per. 1-4)	10.626** (3.990)			
Average Rating ¹ (Per. 5-8)		4.160* (2.107)		
Average Rating ¹ (Per. 9-12)			4.290* (1.694)	
Average Rating ¹ (Per. 13-16)				0.082 (2.694)
Observations	44	40	41	41
Number of groups	12	12	12	12

Note: The table presents results from two-level models with random effects at the market levels. The dependent variables are the sums of profits generated by an expert over the course of the entire experiment (16 periods). Robust standard errors in parentheses.

¹ Average Rating obtained in the distinct blocks of periods (period 0-4, period 5-8, period 9-12, period 13-16).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

can base their expert choice on their own previous experience with a particular expert, and those in which this is not the case. The results show that even though patients cannot judge all relevant quality aspects even ex-post, a public rating system significantly improves market efficiency and consumer surplus. Even overcharging, which cannot be detected, decreases significantly. When it comes to expert behavior, ratings can influence it essentially in two ways—directly, through a signal sent by the consumer to an expert, and indirectly, through the reputational effect of those ratings. Our results suggest that the reputational effect is the driving force.

Furthermore, we find that a public rating system is a good substitute for personal experience information to enhance market outcomes. Market efficiency and consumer surplus are on the same levels in markets with a public rating system compared to personal experience markets. When patients have both personal experiences as well as public ratings to base their decisions on, they tend to prioritize the former over public ratings. This aligns with previous research in experience goods suggesting that public feedback systems are particularly helpful for those interacting for the first time, like tourists and travelers (Fang, 2022), but do not offer additional benefits when market participants can rely on personal relationships (Cai et al., 2014).

Considering ratings, we see that patients use them effectively to reward or punish experts, which allows reputation equilibria to emerge. We find that consumers symmetrically punish experts with a *zero* rating when being undertreated, give low ratings (albeit more dispersed) when being charged the high price, and reward—again symmetrically—experts for which they know that they did neither undertreat nor overcharge with a five-star rating. These rating strategies appear to be well understood by all market participants as they then strongly direct subsequent expert choice. Experts that follow a reputation for quality strategy in early periods thus generate higher ratings which in turn lead them to interact with more patients over the course of the experiment. Furthermore, we find that higher overall ratings are associated with higher experimental profits, suggesting that experts indeed benefit from building up a good rating/reputation for themselves.

Despite the potential benefits of public rating mechanisms, there are also concerns about their accuracy and reliability. Ratings are affected by various factors other than the genuine quality of the service provided (Doing-Harris et al., 2016; López et al., 2012; Okike et al., 2016), in particular in credence goods markets. Compared to the experimental setup, ratings tend to be more subjective and the information provided about the quality of expert decisions becomes less reliable. Review fraud is another potential concern. Studying the resilience of public rating platforms in the face of growing levels of noise and reduced reliability of feedback is a vital area for future research.

References

- Angerer, S., Glätzle-Rützler, D., & Waibel, C. (2021a). Monitoring institutions in healthcare markets: Experimental evidence. *Health Economics*, 30(5). <https://doi.org/10.1002/hec.4232>
- Angerer, S., Glätzle-Rützler, D., & Waibel, C. (2021b). Trust in health care credence goods: Experimental evidence on framing and subject pool effects. Working papers, Faculty of Economics and Statistics, University of Innsbruck.
- Anthun, K. S., Bjørngaard, J. H., & Magnussen, J. (2017). Economic incentives and diagnostic coding in a public health care system. *International Journal of Health Economics and Management*, 17(1), 83–101. <https://doi.org/10.1007/s10754-016-9201-9>
- Ba, S. & Paul, A. P. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 26(3), 243–268. <https://doi.org/10.2307/4132332>
- Baker, L. C. (2010). Acquisition of mri equipment by doctors drives up imaging use and spending. *Health Affairs*, 29(12), 2252–2259. <https://doi.org/10.1377/hlthaff.2009.1099>. PMID: 21134927
- 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, 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. *The Economic Journal*, 127(599), 1–18. <https://doi.org/10.1111/eoj.12260>

- Bar-Isaac, H. & Tadelis, S. (2008). Seller reputation. *Foundations and Trends® in Microeconomics*, 4(4), 273–351. <https://doi.org/10.1561/07000000027>
- Barros, P. & Braun, G. (2017). Upcoding in a national health service: the evidence from portugal. *Health Economics*, 26(5), 600–618. <https://doi.org/10.1002/hec.3335>
- Batty, M. & Ippolito, B. (2017). Financial incentives, hospital care, and health outcomes: Evidence from fair pricing laws. *American Economic Journal: Economic Policy*, 9(2), 28–56. <https://doi.org/10.1257/pol.20160060>
- Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review*, 71, 117–120. <https://doi.org/10.1016/j.euroecorev.2014.07.003>
- Bohnet, I. & Huck, S. (2004). Repetition and reputation: Implications for trust and trustworthiness when institutions change. *American Economic Review*, 94(2), 362–366. <https://doi.org/10.1257/0002828041301506>
- Bolton, G. E., Katok, E., & Ockenfels, A. (2004). How effective are electronic reputation mechanisms? an experimental investigation. *Management Science*, 50(11), 1587–1602. <https://doi.org/10.1287/mnsc.1030.0199>
- Brosig-Koch, J., Hehenkamp, B., & Kokot, J. (2017a). The effects of competition on medical service provision. *Health Economics*, 26, 6–20. <https://doi.org/10.1002/hec.3583>
- Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., & Wiesen, D. (2013). How effective are pay-for-performance incentives for physicians? - a laboratory experiment. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2278863>
- 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>

- Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017b). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, 26(2), 243–262. <https://doi.org/10.1002/hec.3292>
- Brosig-Koch, J., Kairies-Schwarz, N., & Kokot, J. (2017c). Sorting into payment schemes and medical treatment: A laboratory experiment. *Health Economics*, 26(S3), 52–65. <https://doi.org/https://doi.org/10.1002/hec.3616>
- Brown, D. L. & Clement, F. (2018). Calculating health care waste in washington state: first, do no harm. *JAMA Internal Medicine*, 178(9), 1262–1263. <https://doi.org/10.1001/jamainternmed.2018.3516>
- Cabral, L. & Hortaçsu, A. (2010). The dynamics of seller reputation: evidence from ebay*. *The Journal of Industrial Economics*, 58(1), 54–78. <https://doi.org/https://doi.org/10.1111/j.1467-6451.2010.00405.x>
- Cai, H., Jin, G. Z., Liu, C., & Zhou, L.-A. (2014). Seller reputation: From word-of-mouth to centralized feedback. *International Journal of Industrial Organization*, 34, 51–65. <https://doi.org/10.1016/j.ijindorg.2014.03.002>
- Chao, M. & Larkin, I. (2022). Regulating conflicts of interest in medicine through public disclosure: Evidence from a physician payments sunshine law. *Management Science*, 68(2), 1078–1094. <https://doi.org/10.1287/mnsc.2020.3940>
- Chen, A. & Goldman, D. (2016). Health care spending: Historical trends and new directions. *Annual Review of Economics*, 8(1), 291–319. <https://doi.org/10.1146/annurev-economics-080315-015317>
- Chen, P.-Y., Hong, Y., & Liu, Y. (2018). The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments. *Management Science*, 64(10), 4629–4647. <https://doi.org/10.1287/mnsc.2017.2852>

- 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>
- Cook, A. & Averett, S. (2020). Do hospitals respond to changing incentive structures? evidence from medicare's 2007 drg restructuring. *Journal of Health Economics*, 73, 102319. <https://doi.org/10.1016/j.jhealeco.2020.102319>
- Currie, J., Lin, W., & Meng, J. (2014). Addressing antibiotic abuse in china: An experimental audit study. *Journal of development economics*, 110, 39–51. <https://doi.org/10.1016/j.jdeveco.2014.05.006>
- Currie, J., Lin, W., & Zhang, W. (2011). Patient knowledge and antibiotic abuse: Evidence from an audit study in china. *Journal of Health Economics*, 30(5), 933–949. <https://doi.org/10.1016/j.jhealeco.2011.05.009>
- Cutler, D. M. (1995). The incidence of adverse medical outcomes under prospective payment. *Econometrica*, 63(1), 29–50. <https://doi.org/10.2307/2951696>
- Dafny, L. S. (2005). How do hospitals respond to price changes? *American Economic Review*, 95(5), 1525–1547. <https://doi.org/10.1257/000282805775014236>
- Dai, T., Akan, M., & Tayur, S. (2017). Imaging room and beyond: The underlying economics behind physicians' test-ordering behavior in outpatient services. *Manufacturing & Service Operations Management*, 19(1), 99–113. <https://doi.org/10.1287/msom.2016.0594>
- Darby, M. R. & Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law & Economics*, 16(1), 67–88. <http://www.jstor.org/stable/724826>
- Das, J. & Hammer, J. (2007). Money for nothing: The dire straits of medical practice in delhi, india. *Journal of Development Economics*, 83(1), 1–36. <https://doi.org/10.1016/j.jdeveco.2006.05.004>

- Das, J., Holla, A., Mohpal, A., & Muralidharan, K. (2016). Quality and accountability in health care delivery: Audit-study evidence from primary care in india. *American Economic Review*, 106(12), 3765–99. <https://doi.org/10.1257/aer.20151138>
- Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), 1407–1424.
- Dellarocas, C. (2005). Reputation mechanism design in online trading environments with pure moral hazard. *Information Systems Research*, 16(2), 209–230.
- Dellarocas, C. (2006). Reputation mechanisms. *Handbook on Economics and Information Systems*, 1–38. <https://doi.org/10.1287/isre.1050.0054>
- Doing-Harris, K., Mowery, D. L., Daniels, C., Chapman, W. W., & Conway, M. (2016). *Understanding patient satisfaction with received healthcare services: A natural language processing approach*. <http://europepmc.org/abstract/MED/28269848>
- Domenighetti, G., Casabianca, A., Gutzwiller, F., & Martinoli, S. (1993). Revisiting the most informed consumer of surgical services: The physician-patient. *International Journal of Technology Assessment in Health Care*, 9(4), 505–513. <https://doi.org/10.1017/S0266462300005420>
- Doyle, J. J. (2005). Health insurance, treatment and outcomes: Using auto accidents as health shocks. *The Review of Economics and Statistics*, 87(2), 256–270. <https://doi.org/10.1162/0034653053970348>
- 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. *The American Economic Review*, 101(2), 526–555. <http://www.jstor.org/stable/29783682>
- Dunn, A. & Shapiro, A. H. (2014). Do physicians possess market power? *The Journal of Law and Economics*, 57(1), 159–193. <https://doi.org/10.1086/674407>

- Ely, J., Fudenberg, D., & Levine, D. K. (2008). When is reputation bad? *Games and Economic Behavior*, 63(2), 498–526. <https://doi.org/10.1016/j.geb.2006.08.007>
- Ely, J. C. & Välimäki, J. (2003). Bad reputation. *The Quarterly Journal of Economics*, 118(3), 785–814. <http://www.jstor.org/stable/25053923>
- Emmert, M. & Meszmer, N. (2018). Eine dekade arztbewertungsportale in deutschland: Eine zwischenbilanz zum aktuellen entwicklungsstand. *Das Gesundheitswesen*, 80(10), 851–858. <https://doi.org/10.1055/s-0043-114002>
- Fang, L. (2022). The effects of online review platforms on restaurant revenue, consumer learning, and welfare. *Management Science*, 68(11), 8116–8143. <https://doi.org/10.1287/mnsc.2021.4279>
- FBI (2011). Financial crimes report to the public - fiscal year 2009. Report 9781611225440, Federal Bureau of Investigation.
- 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. (2018). Behavioral experiments in health economics. *Oxford Research Encyclopedia of Economics and Finance*. <https://doi.org/10.1093/acrefore/9780190625979.013.244>
- Geruso, M. & Layton, T. (2019). Upcoding: Evidence from medicare on squishy risk adjustment. *Journal of Political Economy*, 128(3), 984–1026. <https://doi.org/10.1086/704756>
- Godager, G., Hennig-Schmidt, H., & Iversen, T. (2016). Does performance disclosure influence physicians' medical decisions? an experimental study. *Journal of Economic Behavior & Organization*, 131, 36–46. <https://doi.org/https://doi.org/10.1016/j.jebo.2015.10.005>
- 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>

- Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization*, 106, 367–378. <https://doi.org/https://doi.org/10.1016/j.jebo.2014.05.009>
- Gruber, J. & Owings, M. (1996). Physician financial incentives and cesarean section delivery. *The RAND Journal of Economics*, 27(1), 99–123. <http://www.jstor.org/stable/2555794>
- Han, J., Kairies-Schwarz, N., & Vomhof, M. (2017). Quality competition and hospital mergers-an experiment. *Health Economics*, 26, 36–51. <https://dx.doi.org/10.1002/hec.3574>
- Han, J., Kairies-Schwarz, N., & Vomhof, M. (2020). Quality provision in competitive health care markets: Individuals vs. teams (no. 839). *Ruhr Economic Papers*. <https://doi.org/10.4419/86788972>
- Hanauer, D. A., Zheng, K., Singer, D. C., Gebremariam, A., & Davis, M. M. (2014). Public awareness, perception, and use of online physician rating sites. *JAMA*, 311(7), 734. <https://doi.org/10.1001/jama.2013.283194>
- Hedges, L. & Couey, C. (2020). How patients use online reviews. *Software Advice*, (24.2.2023). <https://www.softwareadvice.com/resources/how-patients-use-online-reviews/>
- Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, 30(4), 637–646. <https://doi.org/https://doi.org/10.1016/j.jhealeco.2011.05.001>
- Hong, Y. A., Liang, C., Radcliff, T. A., Wigfall, L. T., & Street, R. L. (2019). What do patients say about doctors online? a systematic review of studies on patient online reviews. *J Med Internet Res*, 21(4), e12521. <https://doi.org/10.2196/12521>
- Huck, S., Lünser, G., Spitzer, F., & Tyran, J.-R. (2016a). Medical insurance and free choice of physician shape patient overtreatment: A laboratory experiment. *Journal of Economic Behavior & Organization*, 131, 78–105. <https://doi.org/https://doi.org/10.1016/j.jebo.2016.06.009>

- Huck, S., Lünser, G. K., & Tyran, J.-R. (2012). Competition fosters trust. *Games and Economic Behavior*, 76(1), 195–209. <https://doi.org/https://doi.org/10.1016/j.geb.2012.06.010>
- Huck, S., Lünser, G. K., & Tyran, J.-R. (2016b). Price competition and reputation in markets for experience goods: an experimental study. *The RAND Journal of Economics*, 47(1), 99–117. <https://doi.org/https://doi.org/10.1111/1756-2171.12120>
- Iizuka, T. (2007). Experts' agency problems: Evidence from the prescription drug market in japan. *RAND Journal of Economics*, 38(3), 844 – 862. <https://doi.org/10.1111/j.0741-6261.2007.00115.x>
- Januleviciute, J., Askildsen, J. E., Kaarboe, O., Siciliani, L., & Sutton, M. (2016). How do hospitals respond to price changes? evidence from norway. *Health Economics*, 25(5), 620–636. <https://doi.org/10.1002/hec.3179>
- Jürges, H. & Köberlein, J. (2015). What explains drg upcoding in neonatology? the roles of financial incentives and infant health. *Journal of Health Economics*, 43, 13–26. <https://doi.org/https://doi.org/10.1016/j.jhealeco.2015.06.001>
- Kairies, N. & Krieger, M. (2013). How do non-monetary performance incentives for physicians affect the quality of medical care? - a laboratory experiment. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2278866>
- 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/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 of the United States of America*, 113(27), 7454–7458. <https://doi.org/10.1073/pnas.1518015113>

- Kerschbamer, R., Neururer, D., & Sutter, M. (2019). Credence goods markets and the informational value of new media: A natural field experiment. *MPI collective goods discussion paper*, (2019/3).
- Kerschbamer, R., Sutter, M., & Dulleck, U. (2017). How social preferences shape incentives in (experimental) markets for credence goods. *The Economic Journal*, 127(600), 393–416. <https://doi.org/https://doi.org/10.1111/ecoj.12284>
- 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>
- Klein, B. & Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89(4), 615–641.
- Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, 27(2), 245–252. [https://doi.org/https://doi.org/10.1016/0022-0531\(82\)90029-1](https://doi.org/https://doi.org/10.1016/0022-0531(82)90029-1)
- Lagarde, M. & Blaauw, D. (2017). Physicians' responses to financial and social incentives: A medically framed real effort experiment. *Social Science & Medicine*, 179, 147–159. <https://doi.org/https://doi.org/10.1016/j.socscimed.2017.03.002>
- Lim, T. O., Sorays, A., Ding, L. M., & Morad, Z. (2002). Assessing doctors' competence: application of cusum technique in monitoring doctors' performance. *International Journal for Quality in Health Care*, 14(3), 251–258. <https://doi.org/10.1093/oxfordjournals.intqhc.a002616>
- Liu, M., Brynjolfsson, E., & Dowlatabadi, J. (2021). Do digital platforms reduce moral hazard? the case of uber and taxis. *Management Science*. <https://doi.org/10.1287/mnsc.2020.3721>
- Lu, S. F. & Rui, H. (2018). Can we trust online physician ratings? evidence from cardiac surgeons in florida. *Management Science*, 64(6), 2557–2573. <https://doi.org/10.1287/mnsc.2017.2741>

- López, A., Detz, A., Ratanawongsa, N., & Sarkar, U. (2012). What patients say about their doctors online: A qualitative content analysis. *Journal of General Internal Medicine*, 27(6), 685–692. <https://doi.org/10.1007/s11606-011-1958-4>
- McLennan, S., Strech, D., & Reimann, S. (2017). Developments in the frequency of ratings and evaluation tendencies: A review of german physician rating websites. *Journal of Medical Internet Research*, 19(8), e299. <https://doi.org/10.2196/jmir.6599>
- Mimra, W., Rasch, A., & Waibel, C. (2016). 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>
- Moreno, A. & Terwiesch, C. (2014). Doing business with strangers: Reputation in online service marketplaces. *Information Systems Research*, 25(4), 865–886.
- OECD (2021). *Health at a Glance 2021*. <https://doi.org/https://doi.org/https://doi.org/10.1787/ae3016b9-en>
- Okike, K., Peter-Bibb, T. K., Xie, K. C., & Okike, O. N. (2016). Association between physician online rating and quality of care. *Journal of medical Internet research*, 18(12), e324.
- Parkinson, B., Meacock, R., & Sutton, M. (2019). How do hospitals respond to price changes in emergency departments? *Health Economics*, 28(7), 830–842. <https://doi.org/10.1002/hec.3890>
- Pasero, C. & McCaffery, M. (2001). The undertreatment of pain: Are providers accountable for it? *AJN The American Journal of Nursing*, 101(11). https://journals.lww.com/ajnonline/Fulltext/2001/11000/The_Undertreatment_of_Pain
- Rajgopal, S. & White, R. (2019). Cheating when in the hole: The case of new york city taxis. *Accounting, Organizations and Society*, 79, 101070. <https://doi.org/10.1016/j.aos.2019.101070>
- 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>

- Resnick, P. & Zeckhauser, R. (2002). *Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system*, volume 11 of *Advances in Applied Microeconomics*, 127–157. Emerald Group Publishing Limited. [https://doi.org/10.1016/S0278-0984\(02\)11030-3](https://doi.org/10.1016/S0278-0984(02)11030-3)
- Resnick, P., Zeckhauser, R., Swanson, J., & Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2), 79–101. <https://doi.org/10.1007/s10683-006-4309-2>
- Rice, S. C. (2012). Reputation and uncertainty in online markets: An experimental study. *Information Systems Research*, 23(2), 436–452. <https://doi.org/10.1287/isre.1110.0362>
- Saifee, D. H., Bardhan, I. R., Lahiri, A., & Zheng, Z. (2019). Adherence to clinical guidelines, electronic health record use, and online reviews. *Journal of Management Information Systems*, 36(4), 1071–1104.
- Saifee, D. H., Zheng, Z. E., Bardhan, I. R., & Lahiri, A. (2020). Are online reviews of physicians reliable indicators of clinical outcomes? a focus on chronic disease management. *Information Systems Research*, 31(4), 1282–1300. <https://doi.org/10.1287/isre.2020.0945>
- 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://doi.org/https://doi.org/10.1111/j.1467-6451.2012.00485.x](https://doi.org/10.1111/j.1467-6451.2012.00485.x)
- Shapiro, C. (1982). Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*, 13(1), 20–35.
- Shigeoka, H. & Fushimi, K. (2014). Supplier-induced demand for newborn treatment: Evidence from japan. *Journal of Health Economics*, 35, 162–178. <https://doi.org/10.1016/j.jhealeco.2014.03.003>
- Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8, 321–340. <https://doi.org/10.1146/annurev-economics-080315-015325>

Xu, Y., Armony, M., & Ghose, A. (2021). The interplay between online reviews and physician demand: An empirical investigation. *Management Science*, 67(12), 7344–7361. <https://doi.org/10.1287/mnsc.2020.3879>

Yelp (2020). *Fast facts*. <https://www.yelp-press.com/company/fast-facts/default.aspx>

Appendix A Additional Figures and Tables

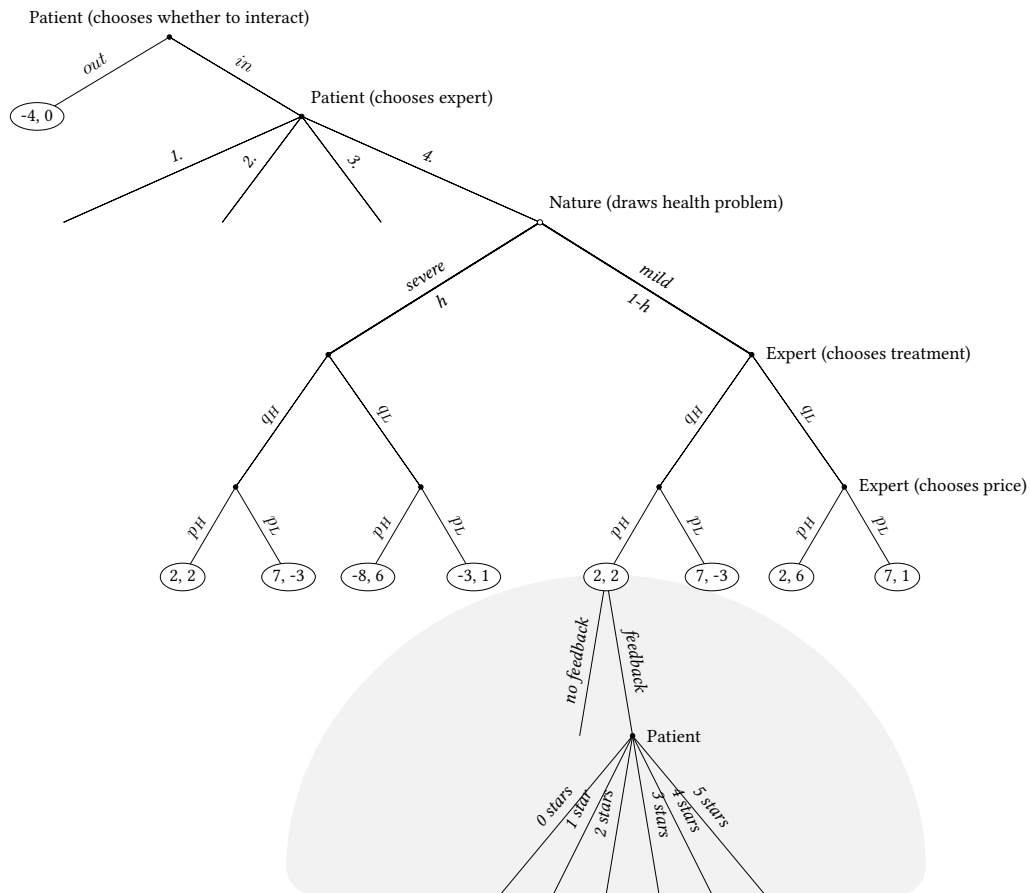


Figure A1: Game tree for one period. The shaded sub-game shows the rating decision of patients in the feedback conditions. After observing their payoffs, patients can rate their interaction with the physician on a five-star rating system. Note that in this game tree, we only draw the feedback decision for one particular outcome $(2, 2)$. It is worth mentioning that patients (in experimental conditions with feedback) have the ability to rate each interaction with an expert, irrespective of the realized payoff.

Table A1: Overview of results (means).

	Markets without personal experience		Markets with personal experience		<i>p-values of MWU</i> ¹					
	Baseline	Rating	Experience	Exp+Rating	Baseline vs Rating	Baseline vs Experience	Baseline vs Exp+Rating	Rating vs Experience	Rating vs Exp+Rating	Experience vs Exp+Rating
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Expert behavior										
Undertreatment (in %)	64.72	5.81	6.81	6.89	0.002	0.002	0.002	0.794	0.667	0.777
Overcharging (in %)	92.03	47.94	36.89	38.07	0.002	0.002	0.004	0.093	0.180	0.974
Overtreatment (in %)	0.00	4.30	6.47	0.56	0.455	0.182	1.000	0.546	0.727	0.303
Consumer decisions										
Interaction (in %)	93.75	98.96	99.74	99.48	1.000	0.424	0.424	0.424	0.546	1.000
Feedback (in %)	-	93.62	-	88.68	-	-	-	-	0.407	-
Star-rating	-	3.66	-	4.06	-	-	-	-	0.485	-
Market outcomes										
Efficiency (in %)	70.70	96.21	96.42	96.85	0.002	0.002	0.002	0.849	0.725	0.959
Consumer Surplus (in ECU)	-1.27	3.01	3.05	3.30	0.002	0.002	0.002	0.937	0.513	0.485
Observations	384	384	384	384						

Note: We analyze six independent markets in every experimental condition. In each market, four patients and four physicians interact. The experimental conditions are: **Baseline**, **Experience**, **Rating**, and **Exp+Rating**. Please refer to Section 3.2 for a description of the experimental conditions. See Table 2 for a description of the outcome variables.

¹ Mann-Whitney U-tests for pairwise differences between conditions with matching groups of 8 subjects as one independent observation. *p*-values are adjusted for the small sample size, using Fisher's exact test.

Table A2: Average Treatment Effects (regression output)

	Undertreatment		Overcharging		Efficiency	Consumer Surplus
	(1)	(2)	(3)	(4)	(5)	(6)
Rating	-3.005***	-2.894***	-2.075***	-1.773***	0.255***	4.271***
	(-4.99)	(-4.47)	(-4.18)	(-5.11)	(3.70)	(6.22)
Experience	-3.000***	-3.065***	-2.398***	-2.188***	0.253***	4.318***
	(-4.48)	(-4.66)	(-4.68)	(-4.82)	(3.62)	(6.02)
Exp+Rating	-2.777***	-2.585***	-2.471***	-2.114***	0.262***	4.563***
	(-5.36)	(-4.59)	(-4.73)	(-5.91)	(3.81)	(6.74)
Period	0.086**	0.087**	0.052**	0.051**	-0.006***	-0.076**
	(3.08)	(3.23)	(3.04)	(3.06)	(-3.48)	(-2.91)
Male (yes)		-0.332		-0.416*		
		(-1.11)		(-2.38)		
Age (in years)		-0.025		-0.049		
		(-0.59)		(-1.95)		
Amount donated to charity in a dictator game		-0.162**		-0.098**		
		(-3.10)		(-3.02)		
Liar (yes)		0.877		0.432		
		(1.73)		(1.73)		
Trustworthiness		0.544		-0.163		
		(0.81)		(-0.39)		
Extraversion		-0.298		-0.0712		
		(-1.30)		(-0.60)		
Agreeableness		-0.404*		0.101		
		(-2.35)		(0.65)		
Conscientiousness		0.0837		-0.221*		
		(0.39)		(-2.20)		
Neuroticism		-0.179		0.171		
		(-0.86)		(1.15)		
Openness		0.001		-0.131		
		(0.00)		(-1.32)		
Frequency of practicing religion (sometimes)		-0.661		-0.008		
		(-1.81)		(-0.04)		
Frequency of practicing religion (often)		-0.059		-0.765*		
		(-0.17)		(-2.00)		
Relative school performance		-0.001		0.000		
		(-0.09)		(0.01)		
Experience with incorrect physician behavior (yes)		-0.831*		0.016		
		(-2.29)		(0.07)		
Number of physician visits in the past 12 months		0.005		-0.032		
		(0.23)		(-1.50)		
Experience with incorrect physician behavior (yes)		1.173**		0.224		
		(2.81)		(0.70)		
Economics/business major (yes)		-0.060		-0.091		
		(-0.18)		(-0.30)		
Constant	0.101	2.695	1.807***	3.737**	0.754***	-0.618
	(0.24)	(1.02)	(3.74)	(3.01)	(11.38)	(-0.89)
Observations	770		735		1536	1536
Number of Groups	24		24		24	24

Note: The table presents results from multilevel models with random effects at the market and individual levels (undertreatment & overcharging: columns 1-4) or at the market level for market efficiency (column 5) and consumer surplus (column 6). **Dependent Variables:** The level of undertreatment: patient needs q_h , but the physician provides q_l , the level of overcharging: patient needs q_l , the physician provides q_l but charges for q_h , and market efficiency, defined as *zero* if there was no interaction, *one* if the patient was treated correctly, 0.25 (0.67) if the patient was undertreated (overtreated). Consumer surplus is the payoff of patients in a given period. **Covariates:** Gender, age, BIG 5 personality traits (extraversion, agreeableness, conscientiousness, neuroticism, openness) measured with a 10-item BIG 5 questionnaire, whether the participant is a business/economics major, self-reported frequency of practicing religion (never is the reference category), number of expert visits in the past 12 months, an indicator for experience with incorrect expert behavior, an indicator for experience with expert recommendations, relative school performance as a proxy for IQ, a measure for altruism (the amount donated to charity in a dictator game), an indicator whether the participant is classified as a liar (if reporting 4 or more correct dice rolls out of 12 in a lying task), and trustworthiness measured in a standard trust game. Robust standard errors in parentheses.

Appendix B Additional Information and Results

In addition to the four conditions discussed in the paper, we ran four further conditions to investigate the role of competition (patients can choose among 4 experts), personal experience in the absence of competition, and private ratings on market outcomes. Descriptive statistics of our sample for all experimental conditions separately are shown in Table B1.

To disentangle the effect of personal experience and competition, we ran two conditions in which patients were randomly matched with an expert in each round and thus, there was no competition between experts. In one condition, experts are identifiable and patients can thus attribute personal information to a given expert, (*ExpNoComp*) and in the other condition physicians are not identifiable (*NoComp*). Comparing condition *Experience* (*Baseline*) with *ExpNoComp* (*NoComp*) shows the effect of adding competition in a market with (without) personal experience information, whereas the comparison of *ExpNoComp* with no-Comp shows the effect of adding personal experience information into a setting without competition between experts.

To disentangle the effect of providing a private rating to experts from the reputational effect of a public rating system, we ran two additional rating conditions in which patients can rate the interaction with the physician without showing the rating to other market participants (*Rating-Priv* and *Exp+Rating-Priv*). Comparing condition *Baseline* with *Rating-Priv*, respectively *Experience* with *Exp+Rating-Priv* shows the effect of providing feedback (cheap talk) to the expert, whereas the comparison between *Rating-Priv* with *Rating*, respectively *Exp+Rating-Priv* and *Exp+Rating* shows the effect of reputational incentives of the ratings.

Table B2 reports the aggregate results of our main outcome variables for all experimental conditions averaged over markets and periods. The following discussion is based on the results from the regression analysis, which are largely in line with the non-parametric tests.

Competition: In markets without personal experience information, we find that competition (*NoComp* vs. *Baseline*) does not alter market outcomes, except for an unexpected and weakly significant increase in undertreatment. However, the introduction of competition does not result in significantly different levels of market efficiency. If instead, personal experience information is available, allowing patients to choose among experts significantly improves market outcomes. Undertreatment- and overcharging rates are significantly lower in *Experience* compared with *ExpNoComp*, leading to higher overall market efficiency. This finding is in line with Huck et al. (2012) who find that only if some form of reputation-building is coupled with

competition, market outcomes are enhanced. Similar findings were reported by [Brosig-Koch et al. \(2017a\)](#) and [Han et al. \(2017\)](#), who show that competition among healthcare providers results in higher patient well-being.

Private Feedback: Comparing *Baseline* with *Rating-Priv*, respectively *Experience* with *Exp+Rating-Priv* allows analyzing the impact of private ratings from patients to experts. In markets without reputation-building (*Baseline*) the mere fact that patients can send private ratings (*Rating-Priv*) to experts significantly decreases undertreatment, whereas there is no effect on overall efficiency. Allowing patients to give a private rating to experts in markets with personal experience information (*Exp+Rating-Priv*) leads to an unexpected but significant increase in overcharging rates while overall market efficiency is not affected.

Reputation effect of ratings: We saw that private ratings seems not to improve market outcomes by and large, except for a reduction in undertreatment in markets with first-time interactions. We have seen, however, that the possibility to rate experts enhances market outcomes when it enables experts to build up a reputation for quality as in condition *Rating*. Comparing *Rating-Priv* with *Rating*, allows us to analyze the reputational effect of public rating mechanisms. We find highly significant decreases in undertreatment- and overcharging rates, which translate into significantly higher efficiency levels when ratings are made public, allowing patients to guide their choice of experts.

Table B1: Descriptive Statistics

	Markets without personal experience				Markets with personal experience			
	<i>Baseline</i>	<i>Rating</i>	<i>NoComp</i>	<i>Rating-Priv</i>	<i>Experience</i>	<i>Exp+Rating</i>	<i>ExpNoComp</i>	<i>Exp+Rating-Priv</i>
Male [%]	41.7	52.1	56.3	43.8	43.8	54.2	50.0	39.6
Age (in years)	22.3 (2.82)	22.8 (3.07)	22.1 (2.43)	22.5 (2.63)	22.8 (4.08)	22.9 (2.85)	23.4 (5.69)	23.7 (5.23)
Relative School Performance	0.72 (0.180)	0.71 (0.222)	0.76 (0.143)	0.73 (0.182)	0.70 (0.201)	0.76 (0.141)	0.73 (0.184)	0.74 (0.185)
Number of Physician Visits last year	4.92 (4.78)	5.83 (6.56)	4.19 (3.02)	4.54 (4.05)	4.94 (5.19)	4.73 (3.58)	5.96 (14.23)	4.88 (3.72)
Exp. with incorrect physician behavior [%]	56.3 (49.6)	45.8 (49.8)	33.3 (47.1)	60.4 (48.9)	27.1 (44.5)	37.5 (37.5)	50.0 (50.0)	39.6 (48.9)
Exp. with physician recommendation [%]	81.3 (39.4)	75.0 (43.8)	-	79.2 (41.0)	77.1 (42.5)	83.3 (37.7)	87.5 (33.4)	75.0 (43.8)
Business/Economics major [%]	29.2 (45.9)	52.1 (50.4)	39.6 (49.4)	33.3 (47.6)	33.3 (47.6)	47.9 (50.5)	50.0 (50.5)	37.5 (48.9)
Frequency of practicing Religion [%]								
Never	54.17	58.33	50	56.25	77.08	50	58.33	52.08
Rarely	33.33	33.33	41.67	37.5	18.75	43.75	27.08	41.67
Often	12.5	8.33	8.33	6.25	4.17	6.25	14.58	6.25
Extraversion	3.48 (0.97)	3.43 (1.01)	3.55 (0.86)	3.65 (1.00)	3.32 (0.92)	3.46 (0.81)	3.53 (0.83)	3.40 (0.96)
Agreeableness	3.26 (0.83)	3.12 (0.96)	3.26 (0.78)	3.36 (0.80)	3.22 (0.98)	3.26 (0.89)	3.12 (0.88)	3.15 (0.91)
Conscientiousness	3.56 (0.77)	3.43 (0.79)	3.37 (0.87)	3.88 (0.80)	3.33 (0.86)	3.52 (0.87)	3.56 (0.85)	3.63 (0.87)
Neuroticism	2.87 (0.95)	2.96 (0.84)	2.95 (1.02)	2.82 (1.09)	2.95 (0.92)	2.91 (0.92)	2.75 (1.09)	2.93 (1.05)
Openness	3.81 (0.95)	3.52 (0.93)	3.50 (1.14)	3.64 (0.92)	3.50 (1.16)	3.55 (1.06)	3.60 (1.08)	3.30 (0.92)
DGKeep	8.78 (3.18)	8.03 (3.41)	7.87 (3.55)	8.20 (3.41)	7.11 (4.18)	7.48 (3.56)	5.74 (3.98)	7.42 (3.87)
Risk Aversion	11.71 (2.81)	12.46 (3.86)	12.19 (3.52)	11.27 (3.00)	12.33 (3.02)	12.56 (2.92)	13.23 (3.43)	13.31 (3.38)
Trustworthiness	0.29 (0.26)	0.31 (0.21)	0.31 (0.26)	0.38 (0.23)	0.40 (0.28)	0.43 (0.21)	0.43 (0.24)	0.32 (0.22)
Lying	9.77 (2.90)	8.42 (3.67)	9.15 (3.24)	9.25 (3.17)	8.85 (3.12)	7.40 (3.71)	7.27 (3.85)	8.81 (3.46)
Experimental Payoff (physicians)	75.25 (23.51)	42.33 (18.91)	69.33 (17.02)	70.46 (27.55)	40.33 (21.49)	38.50 (22.90)	61.17 (13.72)	43.46 (25.88)
Experimental Payoff (patients)	-20.25 (30.27)	48.08 (16.27)	-1.83 (20.29)	-1.54 (23.06)	48.83 (18.05)	52.75 (13.97)	8.75 (19.50)	46.38 (18.08)

Note: We analyze six independent markets in every experimental condition. In each market, four patients and four physicians interact. Means (standard deviations). DG-Keep is the amount kept in a dictator game (DG) as a measure of altruism, trustworthiness is measured as the share sent back to the first-mover in a trust game, risk aversion is the number of safe choices in a choice list with 20 binary decision problems between a risky prospect and a safe option, lying ranging from 0 to 12 measured in the lying task, the BIG 5 personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness) measured with a 10-item BIG 5 questionnaire and the other demographic background variables, as well as the frequency of practicing religion (never, often, rarely), experience with incorrect physician behavior, number of physician visits, self-reported relative school performance as a proxy for IQ, and experience with physician recommendations (absent in no-Comp), are measured in the post-experimental questionnaire (see Appendix F for a detailed description of all these measures). The experimental payoff is the sum of payoffs in ECU generated by participants over the 16 periods (not the payout they received at the session's end).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Overview of all results (means).

	Markets without personal experience				Markets with personal experience			
	<i>Baseline</i>	<i>Rating</i>	<i>NoComp</i>	<i>Rating-Priv</i>	<i>Experience</i>	<i>Exp+Rating</i>	<i>ExpNoComp</i>	<i>EXP+Rating-Priv</i>
Expert behavior								
Overtreatment (in %)	0.00	4.3	0.60	1.61	6.47	0.56	0.49	0.49
Undertreatment (in %)	64.72	5.81	39.90	42.93	6.81	6.89	24.82	11.76
Overcharging (in %)	92.03	47.94	91.02	86.65	36.89	38.07	88.54	48.51
Consumer decisions								
Interaction (in %)	93.75	98.96	96.88	98.70	99.74	99.48	91.41	100.00
Feedback (in %)	-	93.62	-	84.57	-	88.68	-	77.60
Star-rating	-	3.66	-	3.02	-	4.06	-	3.73
Market outcomes								
Efficiency (in %)	70.70	96.21	80.77	81.06	95.42	96.85	82.53	95.42
Consumer Surplus (in ECU)	-1.27	3.01	-0.11	-0.10	3.05	3.30	0.55	2.89
Observations	384	384	384	384	384	384	384	384

Note: We analyze six independent markets in every experimental condition. In each market, four patients and four experts interact. The experimental conditions are: **Baseline**, **Rating**, No Competition (**NoComp**), Private Feedback (**Rating-Priv**), **Experience**, **Exp+Rating**, Experience without Competition (**ExpNoComp**), and Experience with private Feedback (**Exp+Rating-Priv**). Please refer to Section 3.2 for a description of the main experimental conditions and to Appendix B for a description of the additional conditions.

Appendix C Detailed information about the rating conditions and screenshots

In **Rating** physicians see at the end of each period the private rating²⁹ for patients they treated, and who decided to rate them. Besides, physicians (Figure C1) and patients (Figure C2) observe the public average rating³⁰ of all physicians over all treated patients when they make their decisions starting in period five. The reason for displaying the public average rating only from period five onwards is to render direct reputation-building impossible in the first couple of rounds, where not many ratings have been submitted so far and identification might be possible via those ratings.

In **Rating-Priv**, physicians receive at the end of each period a private rating from patients they treated, and who decided to rate them. Neither physicians (Figure C3), nor patients (Figure C4) see any ratings when taking their decisions. Their decision screens look the same as in **Baseline**.

In **Exp+Rating** physicians see at the end of each period the private rating for patients they treated, and who decided to rate them. In line with **Rating**, patients (Figure C5) and physicians (Figure C6) observe the public average rating of each physician over all treated patients when they make their decisions (from period 2 onwards). Figure C5 shows the decision screen for a patient. However, unlike in condition **Rating**, physicians also see the private average ratings³¹ received from each patient separately, and patients see their own private average ratings from previous interactions for each physician separately on top of the average public ratings over all patients. It is important to distinguish between the two average ratings. While the *public average rating* is the average rating for a physician from all patients, the *private average rating* is the average rating for a physician from one patient.

In **Exp+Rating-Priv**, physicians receive at the end of each period a private rating from patients they treated, and who decided to rate them. Patients see the private average rating for all physicians they rated so far when they decide whether, and which of the physicians to visit (Figure C7). Additionally, physicians observe their private average rating per patient when they decide about treatments and prices (Figure C8).

²⁹ Private rating from one patient to one physician in any given round on a five-star rating scale. Note that rating a physician is optional in our experiment. Patients may decide not to rate physicians they interacted with.

³⁰ The public average rating is calculated as the sum of all ratings for a given physician, divided by the number of ratings for this physician.

³¹ The private average rating is calculated as the sum of all ratings for a physician by a given patient, divided by the number of ratings the patient has given the physician so far.

Decision 1: Consult a physician?

As patient you can decide if you want to consult a physician in this period.

Do you want to consult a physician in this period? ☐ yes ☐ no

If so, which physician do you chose? (please chose only one)





Physician	Public Rating (Number) <small>(public average rating per physician from pervious interactions)</small>	Public Rating (Stars) <small>(public average rating per physician from pervious interactions)</small>	Consult Physician
1 st Physician	3.80		<input type="radio"/> consult 1 st physician <input type="radio"/> clear selection
2 nd Physician	3.25		<input type="radio"/> consult 2 nd physician <input type="radio"/> clear selection
3 rd Physician	3.00		<input type="radio"/> consult 3 rd physician <input type="radio"/> clear selection
4 th Physician	4.20		<input type="radio"/> consult 4 th physician <input type="radio"/> clear selection

Figure C1: Decision screen of patients in *Rating*. Patients are asked whether to visit a physician or not. If they do, they may choose one from a list of four. Starting in period five, they see the public average rating once as a number (column 2) and once as a star rating (column 3). These two columns are absent in the first four periods.

Decision 2 & 3: Treatment and Prices

Chose the treatment for your patients and the price:

The price for the **mild treatment** is 3 points. The costs are 2 points.
The price for the **severe treatment** is 8 points. The costs are 6 points.

Patient	Type of illness	Choose a treatment	Choose a price
1 st Patient	mild illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
2 nd Patient	severe illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
3 rd Patient		<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
4 th Patient	severe illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment





	Other Physician	Other Physician	Your Rating	Other Physician
Public Rating (Number)	3.80	3.25	3.00	4.20
Public Rating (Stars)				

Figure C2: Decision screen of physicians in *Rating*. Physicians see the type of illness of patients visiting them (column 2). Starting in period five, they see the public rating of themselves and the other physicians. They have to choose a treatment and a price for every visiting patient.

Decision 1: Consult a physician
As patient you can decide if you want to consult a physician in this period.

Do you want to consult a physician in this period? ☐ yes ☐ no

If so, which physician do you chose? (please chose only one)

Patient	Consult Physician
1 st Physician	<input type="radio"/> consult 1 st physician <input type="radio"/> clear selection
2 nd Physician	<input type="radio"/> consult 2 nd physician <input type="radio"/> clear selection
3 rd Physician	<input type="radio"/> consult 3 rd physician <input type="radio"/> clear selection
4 th Physician	<input type="radio"/> consult 4 th physician <input type="radio"/> clear selection

Figure C3: Decision screen of patients in *Rating-Priv*. Patients are asked whether to visit a physician or not. If they do, they may choose one from a list of four.

Decision 2 & 3: Treatment and Prices
Chose the treatment for your patients and the price:

The price for the **mild treatment** is 3 points. The costs are 2 points.
The price for the **severe treatment** is 8 points. The costs are 6 points.

Patient	Type of illness	Choose a treatment	Choose a price
1 st Patient	mild illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
2 nd Patient	severe illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
3 rd Patient		<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
4 th Patient	severe illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment

Figure C4: Decision screen of physicians in *Rating-Priv*. Physicians see the type of illness of patients visiting them (column 2). They have to choose a treatment and a price for every visiting patient.

Decision 1: Consult a physician?
As patient you can decide if you want to consult a physician in this period.

Do you want to consult a physician in this period? ☐ yes ☐ no

If so, which physician do you chose? (please chose only one)

Physician	Public Rating (Number) <small>(public average rating per physician from previous interactions)</small>	Public Rating (Stars) <small>(public average rating per physician from previous interactions)</small>	Private Rating (Number) <small>(own average rating per physician from previous interactions)</small>	Private Rating (Stars) <small>(own average rating per physician from previous interactions)</small>	Consult Physician
Physician 1	3.80	★★★★★	4.60	★★★★★	<input type="radio"/> consult physician 1 <input type="radio"/> clear selection
Physician 2	3.25	★★★★★			<input type="radio"/> consult physician 2 <input type="radio"/> clear selection
Physician 3	3.00	★★★★★	2.25	★★★★★	<input type="radio"/> consult physician 3 <input type="radio"/> clear selection
Physician 4	4.20	★★★★★			<input type="radio"/> consult physician 4 <input type="radio"/> clear selection

Figure C5: Decision screen of patients in *Exp+Rating*. Patients are asked whether to visit a physician or not. If they do, they may choose one from a list of four. Starting in the second period, they see the public average rating once as a number (column 2) and once as a star rating (column 3). Additionally, they see their private average rating for those physicians they already rated as a number (column 4) and a star rating (column 5).

Decision 2 & 3: Treatment and Prices
Chose the treatment for your patients and the price:

The price for the **mild treatment** is 3 points. The costs are 2 points.
The price for the **severe treatment** is 8 points. The costs are 6 points.

Physician	Average Rating of Patient <small>(from previous interactions)</small>	Average Rating in Stars <small>(from previous interactions)</small>	Type of illness	Choose a treatment	Choose a price
Patient 1			mild illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 2	2.67	★★★★★		<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 3				<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 4	4.10	★★★★★	mild illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment

	Other Physician	Other Physician	Your Rating	Other Physician
Public Rating (Number)	3.80	3.25	3.00	4.20
Public Rating (Stars)	★★★★★	★★★★★	★★★★★	★★★★★

Figure C6: Decision screen of physicians in *Exp+Rating*. Physicians see the type of illness of patients visiting them (column 4). Starting in period two, they see the private average rating from patients (columns 2 and 3). Additionally, they see the public rating of themselves and other physicians. They have to choose a treatment and a price for every visiting patient.

Decision 1: Consult a physician?

As patient you can decide if you want to consult a physician in this period.

Do you want to consult a physician in this period? ☐ yes ☐ no

If so, which physician do you chose? (please chose only one)

Physician	Rating (Number) <small>(own average rating per physician from pervious interactions)</small>	Rating (Stars) <small>(own average rating per physician from pervious interactions)</small>	Consult Physician
Physician 1	0.00	☆☆☆☆☆☆	<input type="radio"/> consult physician 1 <input type="radio"/> clear selection
Physician 2	3.60	☆☆☆☆☆☆	<input type="radio"/> consult physician 2 <input type="radio"/> clear selection
Physician 3			<input type="radio"/> consult physician 3 <input type="radio"/> clear selection
Physician 4	4.00	☆☆☆☆☆☆	<input type="radio"/> consult physician 4 <input type="radio"/> clear selection

Figure C7: Decision screen of patients in *Exp+Rating-Priv*. Patients are asked whether to visit a physician or not. If they do, they may choose one from a list of four. Starting in the second period, they see their private average rating for those physicians they already rated as a number (column 2) and a star rating (column 3).

Decision 2 & 3: Treatment and Prices

Chose the treatment for your patients and the price:

The price for the **mild treatment** is 3 points. The costs are 2 points.
The price for the **severe treatment** is 8 points. The costs are 6 points.

Physician	Average Rating of Patient <small>(from previous interactions)</small>	Average Rating in Stars <small>(from previous interactions)</small>	Type of illness	Choose a treatment	Choose a price
Patient 1			mild illness	<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 2				<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 3				<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment
Patient 4	4.10	☆☆☆☆☆☆		<input type="radio"/> mild treatment <input type="radio"/> severe treatment	<input type="radio"/> price for mild treatment <input type="radio"/> price for severe treatment

Figure C8: Decision screen of physicians in *Exp+Rating-Priv*. Physicians see the type of illness of patients visiting them (column 4). Starting in period two, they see the private average rating from patients (columns 2 and 3). They have to choose a treatment and a price for every visiting patient.

Appendix D Predictions

In this section, we construct a reputation for quality equilibria for the experimental conditions *Rating*, *Experience* and *Exp+Rating*.

We assume that patients and experts are rational, risk-neutral, and maximize their own payoff. All information but the patients' health problem type in a given round is common knowledge. The repeated one-shot equilibrium in which experts provide the minor treatment and charge for the major treatment (price p_H) and patients always interact is an equilibrium in all experimental conditions. In the following, we construct further symmetric equilibria in which experts do not undertreat/build up a reputation in early periods. The reputation equilibria shown below are not unique, similar ones can be constructed in which the no undertreatment/reputation-building phase is for instance shorter. If necessary, we use masculine pronouns (he) for patients and feminine pronouns (she) for experts.

Condition *Experience*

Equilibrium without undertreatment in early periods

- Expert's strategy: Provide sufficient treatment (minor treatment for a minor health problem, major treatment for a major health problem) and charge for the major treatment in periods 1-15. Provide the minor treatment and charge for the major treatment in period 16.
- Patient's beliefs: Expert provides sufficient treatment and charges for major treatment (price p_H) in periods 1-15. Expert provides minor treatment (q_L) and charges for the major treatment (p_H) in periods 16.
- Patient's strategy: Visit an expert every period. Pick one at random in the first period. In periods 2-15, visit the same expert in the following periods as long as she never undertreated. In periods 1-15, if undertreated, randomly pick one of the experts that never undertreated the patient before. If there is no such expert, randomly select one. In period 16, choose the expert visited that never undertreated the patient in any period 1-15. If there is no such expert, choose an expert at random among those never visited before. If there is no expert never visited, randomly select one.

Verification: Patients' beliefs are consistent with experts' strategies. Next turning to patients' strategy: In every period it is rational for a patient to interact as the the lowest expected payoff from interaction (in periods 13-16), $0.5 \cdot 2 + 0.5 \cdot (-8) = (-3)$ is higher than the outside option (-4) . Furthermore, it is payoff-maximizing to stay with an expert that never undertreated them. Considering experts, we have to verify that there exists no profitable deviation. In period 16, an expert has no incentive to deviate from her strategy as providing a minor treatment

(q_L) and charging for the major treatment (p_H) to every patient, independently of the health problem of a patient and the number of patients, maximizes the expert's payoff. Next, we have to show that sticking to the strategy in periods 1-15 is optimal. In period 15, an expert with a major problem patient has a continuation payoff of $6 + 0$ from undertreating this patient, as the patient will not return given the above strategies, whereas the continuation period from not undertreating is $2 + 6$ since the patient will return in period 16. Thus, there is no incentive to deviate to undertreatment in period 15 (for one or more patients). In earlier periods, the continuation payoff of an expert from undertreatment remains the same, whereas the continuation payoff from sticking to the strategy is even higher as the patient returns in more periods. Hence, there is no deviation incentive for an expert.

Equilibrium without undertreatment and without full overcharging in early periods

- Expert's strategy: Provide sufficient treatment (minor treatment for minor health problem, major treatment for major health problem) in periods 1-15, charge for the minor treatment (p_L) in periods 1-11, and charge for the major treatment (p_H) in periods 12-15. Provide the minor treatment and charge for the major treatment in period 16.
- Patient's beliefs: Expert provides sufficient treatment in periods 1-15. Experts charge for the minor treatment (p_L) in periods 1-11 and charge for the major treatment (p_H) in periods 12-15. Expert provides minor treatment (q_L) and charges for the major treatment (p_H) in period 16.
- Patient's strategy: Visit an expert every period. Pick one at random in the first period. In periods 2-15, visit the same expert in the following periods as long as she never undertreated in any of the previous periods and always charged p_L in periods 1-11. In periods 1-11, if undertreated or charged p_H , randomly pick one of the experts that never undertreated or charged the patient p_H before. If there is no such expert, randomly select one. In period 16, choose an expert visited that never undertreated the patient in any period 1-15 and never charged $p - H$ in periods 1-11. If there is no such expert, choose an expert at random among those never visited before. If there is no expert never visited, randomly select one.

Verification: Patients' beliefs are consistent with experts' strategies. Next turning to patients' strategy: In every period it is rational for a patient to interact as the lowest expected payoff from interaction (in periods 13-16), $0.5 \cdot 2 + 0.5 \cdot (-8) = (-3)$ is higher than the outside option (-4) . Furthermore, it is payoff-maximizing to stay with an expert that never undertreated them. Considering experts, we have to verify that there exists no profitable deviation. In period 16, an expert has no incentive to deviate from her strategy as providing a minor treatment (q_L) and charging for the major treatment (p_H) maximizes the expert's payoff. Next, we have to show that sticking to the strategy in periods 12-15 is optimal. In period 15, an expert with a

major problem patient has a continuation payoff of $6 + 0$ from undertreating this patient, as the patient will not return given the above strategies, whereas the continuation period from not undertreating is $2 + 6$ since the patient will return in period 16. Thus, there is no incentive to deviate to undertreatment in period 15 (for one or more patients). In earlier periods for periods 12-15, the continuation payoff of an expert from undertreatment remains the same, whereas the continuation payoff from sticking to the strategy is even higher as the patient returns in more periods. Hence, there is no deviation incentive for an expert. In periods 1-11, the expert makes an expected loss per patient of -1 from not undertreating and always charging p_L , and a loss of -3 with a given major problem patient. The profit from deviating for a major problem patient in period 11 is $6 + (16 - 11) \cdot 0 = 6$ which is lower than the continuation payoff from sticking to the strategy with this patient which amounts to $-3 + (15 - 11)4 + 6 = 19$. As the expert makes losses in the first periods, deviation incentives are larger in period 1: The continuation profit from deviating for a major problem patient in period 1 is $6 + (15) \cdot 0 = 6$ which is lower than the continuation payoff from sticking to the strategy with this patient which amounts to $-3 - 1(11 - 1) + 4(15 - 11) + 6 = 9$. Hence, no expert has an incentive to deviate.

Condition Rating

Equilibrium without undertreatment in early periods

- Provision and charging strategy of an expert: Provide sufficient treatment (minor treatment for a minor health problem, major for a major health problem) and charge for the major treatment in periods 1-13. Provide the minor treatment and charge for the major treatment in periods 14-16.
- Patient's beliefs: Expert provides sufficient treatment and charges for major treatment (p_H) in periods 1-13. Expert provides minor treatment (q_L) and charges for the major treatment (p_H) in the periods 14-16.
- Patient's strategy: Visit an expert every period. Randomly pick one expert in periods 1 – 4. In periods 1-13, give a rating for every interaction following the rule: A rating of 5 stars if the payoff from interaction in the current period is positive, a rating of 0 stars otherwise. Starting in period 5 until period 13, choose randomly among experts with a five-star rating. If, there is no such expert, visit an expert that was never been rated before. If there is no expert with a five-star rating and no expert that was never rated before, pick the highest-rated expert. In periods 14-16, pick the highest-rated expert and do not rate interactions.

Verification: Patients' beliefs are consistent with the experts' strategy. Next turning to patients' strategy: In every period it is rational for a patient to interact as the lowest expected payoff from interaction (in periods 14-16), $0.5 \cdot 2 + 0.5 \cdot (-8) = (-3)$ is higher than the outside option

(−4). Furthermore, starting in period 5, it is rational for a patient to choose the expert with a five-star rating as any rating lower than 5, given the symmetric patient strategy, signals that this expert undertreated some patients in earlier periods. For experts, we have to verify that there exists no profitable deviation from the strategy stated above in any period. In periods 14-16, providing a minor treatment (q_L) and charging for the major treatment (p_H) to a visiting patient maximizes the expert's payoff. Next, we show that there is also no deviation incentive in any period 5-13: In period 13, an expert with the highest deviation incentives (four major problem patients) has a continuation payoff of $4 \cdot 6 + 0 = 24$ from undertreating her patients, as patients will give a rating of 0 stars and hence there will be no patients visiting in periods 14-16, as the expert will not have a 5-star rating anymore and given the above-specified strategies of patients (and other experts). The expected continuation period from not undertreating is $2 \cdot 4 + 6 \cdot 3 = 26$ since the patients will give a rating of 5 stars and, given the symmetric strategies, the expert will have in expectation one patient visiting in each of the periods 14-16. Thus, there is no incentive to deviate to undertreatment in period 13. In earlier periods 5-12, the continuation payoff of an expert from undertreatment remains the same, whereas the continuation payoff from sticking to the strategy is even higher as patients return in more periods. Hence, there is no deviation incentive for an expert in periods 5-13. It remains to show that experts do not deviate in periods 1-4 in which no public ratings are available for expert choice. The incentive to deviate is strongest in period 1, as patients do not adapt their expert choice in periods 2-4. The maximal deviation profit (undertreating four patients with a major disease in period 1 and undertreating any patient thereafter) is $6 \cdot 4 + 3 \cdot 6 = 42$, whereas the continuation payoff from sticking to the above strategy is $2 \cdot 4 + 12 \cdot 4 + 6 \cdot 3 = 74$. Thus, no expert has an incentive to deviate.

Equilibrium without undertreatment and without full overcharging in early periods

- Expert's strategy: Provide sufficient treatment (minor treatment for minor health problem, major treatment for major health problem) in periods 1-13, charge for the minor treatment (p_L) in periods 1-3, and charge for the major treatment (p_H) in periods 4-13. Provide the minor treatment and charge for the major treatment in periods 14-16.
- Patient's beliefs: Expert provides sufficient treatment in periods 1-13. Experts charge for the minor treatment (p_L) in periods 1-3 and charge for the major treatment (p_H) in periods 4-13. Expert provides minor treatment (q_L) and charges for the major treatment (p_H) in periods 14-16.
- Patient's strategy: Visit an expert every period. Randomly pick one expert in periods 1 – 4. In periods 1-3, give a rating for every interaction following the rule: A rating of 5 stars if the payoff from the interaction is 7 (correct treatment, p_L), a rating of 0 stars otherwise. In periods 4-13, give a rating for every interaction following the rule: A

rating of 5 stars if the payoff from the interaction is positive, a rating of 0 stars otherwise. Starting in period 5 until period 13, choose randomly among experts with a five-star rating. If, there is no such expert, visit an expert that was never been rated before. If there is no expert with a five-star rating and no expert that was never rated before, pick the highest-rated expert. In periods 14-16, pick the highest-rated expert and do not rate interactions.

Verification: Patients' beliefs are consistent with experts' strategies. Next turning to patients' strategy: In every period it is rational for a patient to interact as the lowest expected payoff from interaction (in periods 14-16), $0.5 \cdot 2 + 0.5 \cdot (-8) = (-3)$ is higher than the outside option (-4) . Considering experts, we have to verify that there exists no profitable deviation. In periods 14-16, an expert has no incentive to deviate from her strategy as providing a minor treatment (q_L) and charging for the major treatment (p_H) maximizes the expert's payoff. Next, we have to show that sticking to the strategy in periods 4-13 is optimal. In period 13, an expert with the highest deviation incentives (four major problem patients) has a continuation payoff of $4 \cdot 6 + 0 = 24$ from undertreating her patients, as patients will give a rating of 0 stars and hence there will be no patients visiting in periods 14-16, as the expert will not have a 5-star rating anymore and given the above-specified strategies of patients (and other experts). The expected continuation period from not undertreating is $2 \cdot 4 + 6 \cdot 3 = 26$ since the patients will give a rating of 5 stars and, given the symmetric strategies, the expert will have in expectation one patient visiting in each of the periods 14-16. Thus, there is no incentive to deviate to undertreatment in period 13. In earlier periods 4-13, the continuation payoff of an expert from undertreatment remains the same, whereas the continuation payoff from sticking to the strategy is even higher as patients return in more periods. In periods 1-3, the expert makes an expected loss per patient of -1 from not undertreating and always charging p_L , and a loss of -3 with a given major problem patient. The incentive to deviate is strongest in period 1, as patients do not adapt their expert choice in periods 2-4 and experts make losses in early periods. The maximal deviation profit, when facing four major problem patients, is $6 \cdot 4 + 3 \cdot 6 = 42$, whereas the continuation payoff from sticking to the above strategy is $-3 \cdot 4 - 1 \cdot 2 + 10 \cdot 4 + 6 \cdot 3 = 44$. Thus, no expert has an incentive to deviate.

Condition *Exp+Rating*

Reputation equilibria can be constructed as for *Experience*.

Appendix E Short- and long instructions for Rating and Exp+Rating and control questions

To save space, we report the instructions for ***Rating*** and show the variations for ***Exp+Rating*** in brackets and underlined.

Short-Instructions (without screenshots)

Problem

- 16 periods
 - 2 roles: **Physician** and **patient**
 - Random allocation of the role (remains the same over the entire 16 rounds)
 - The patient has an **illness** in every round
 - 2 types of illness: **minor** and **major** illness
 - **Illness** is randomly **re-determined** in **each round**
 - The physician may then freely choose from one of two treatment types: **minor** and **major treatment**
 - **NOTE: minor and major treatment cure minor illness, BUT only major treatment cures major illness**
- Each round consists of **max. 4 decisions** (see description below)

Information patient

- **The patient** does not know at any time whether he has a minor or major illness in the respective round
- The only information the patient receives is ...
 - ... his payoff after decision 2 and 3
 - ... if his illness was cured
 - ... starting in round 5, the public average rating per physician
 - [(after submission of a rating) his private average rating per physician and the public feedback per physician]

Information physician

- **The physician learns** what illness the patient has when the patient decides to go to the physician
- Furthermore, the physician receives information about ...
 - ... her payoff per patient according to her decision 3
 - ... decision 4 of her patients
 - ... starting in round 5, her own public feedback, as well as the public feedback of other physicians [her private average rating per patient (according to the given rating) and her own public feedback, as well as the public feedback of other physicians]

Decision 1 patient

- **Consult a physician? YES or NO**
- **YES:** Select a physician, then proceed to Decision 2
- **NO:** Round ends here:
 - Payoff patient: -4 points
 - Payoff physician: 0 points

Decision 2 physician

- **Type of treatment: minor or major?**
- **minor** treatment: costs (K) for physician:
 $K = 2$ points
- **major** treatment: costs (K) for physician:
 $K = 6$ points

Decision 3 physician

- **Price for the treatment (P) ?**
- **price** for **MINOR** treatment:
 $P = 3$ points
- **price** for **MAJOR** treatment:
 $P = 8$ points

Decision 4 physician

- **Rating the physician? YES or NO**
- **YES:** Rating of the physician with 0 - 5 stars
(0 = not at all satisfied, 5 = very satisfied)
- **NO:** no rating

Payoff patient

$$N - P$$

Illness cured: $N = 10$ points

Illness not cured: $N = 0$ points

Payoff physician

$$P - K$$

(Price chosen in decision 3 minus the costs of the treatment chosen in decision 2)

Long-Instructions

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 those 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.

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: **Physician** and **patient**. At the beginning of the experiment, you will be randomly assigned one of these roles and maintain this role for the entire experiment. On the first screen of the experiment, you can see which role is assigned to you. This role remains the same throughout all periods.

At the beginning of the experiment, you will be randomly assigned to a **group of 7** other players. This **group** remains **the same** for all periods and consists of **4 physicians** and **4 patients**. If you are a patient, the 4 physicians (1st, 2nd, 3rd, and 4th physician) [physician 1, physician 2, physician 3, and physician 4] in your group are your potential interaction partners. If you are a physician, then your potential interaction partners are the 4 patients (1st, 2nd, 3rd, and 4th patients) [patient 1, patient 2, patient 3, and patient 4] in your group. **Note:** The order of the physicians varies randomly from round to round, i.e. the first physician in round one does not necessarily have to be the first physician in round two. The order of the patients varies randomly from round to round as well. [The identification (1, 2, 3, 4) are fixed throughout the experiment, i.e.: A certain patient or physician always has the same identification number (physician 1 is the same person in every round, patient 1 is the same person in every round, etc.).]

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

Overview of the decision situation

Every **patient** is suffering from an **illness** in each period. There are 2 types of illnesses, a **minor** and a **major** illness. Which kind of illness a patient has is determined randomly **each new period**. 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 period – if the coin shows "head", then the patient suffers from a minor illness, if it shows "tails", the patient suffers from a major illness. At **no time** is the patient informed whether he has a minor or major illness in a particular round. The physician learns what illness a patient suffers from

only when the patient decides to consult the physician. The physician may then freely choose from one of two treatment types (**minor** or **major** treatment). However, a **major illness** is **only cured** by a **major treatment**. A **minor illness** is **cured** by a **minor** or a **major treatment**.

Overview of the decisions in a round

Each round consists of a maximum of 4 decisions, which are made consecutively. Decision 1 (consult the physician) is made by the patient; decision 2 (treatment) and 3 (price) are made by the physician; decision 4 (rating) is again made by the patient.

The sequence of the decisions of a round and presentation of their consequences

Decision 1

The **patient** decides **whether** he wants to consult **ONE** physician and **WHICH** of the **4 physicians** (1st, 2nd, 3rd, and 4th physician) he wants to visit (if and with which physician he wants to interact). The order of the physicians is random – at which position a physician appears (as first, second, third, or fourth physician) is determined randomly in each new round.

[The **patient** decides **whether** he wants to consult **ONE** physician and **WHICH** of the **4 physicians** (physician 1, physician 2, physician 3, and physician 4) he wants to visit (if and with which physician he wants to interact).]

If so, the physician in decision 2 and 3 chooses a treatment and sets a price (see below). However, the patient cannot observe which treatment the physician has chosen.

If not, this round **ends** for the **patient**. **If no patient visits a physician** in a given round, the round ends for her as well.

Decision 2

If the patient decides to consult a physician in decision 1, the **physician learns the nature of the patient's illness** *before* making her decision 2. Then the physician chooses a treatment. At **no time** is **the patient** informed about the treatment chosen by the physician.

The treatment incurs a cost for the **physician**.

The **minor treatment** costs the physician **2 points** (= experimental currency unit) and cures only a minor illness.

The **major treatment** costs the physician **6 points** (= experimental currency unit) and cures both, minor and the major illness.

Physicians can choose treatments **independently** of the type of illness.

Decision 3

The physician **charges a price** for the treatment. Two prices are available:

- The price for the **minor treatment** is **3 points**.
- The price for the **major treatment** is **8 points**.

The chosen price **need not** be equal to the price of the treatment chosen in decision 2; it may also be the price of the other treatment.

Decision 4

The patient receives information about his payoff in this round and whether his illness has been cured or not.

Now the patient decides whether he wants to evaluate the interaction with the physician. **If not**, this round ends for him. **If yes**, the patient rates the interaction between 0 (= not satisfied at all) and 5 (= very satisfied) stars.

Afterward, the physician receives information about her payoff and, in case the patient evaluated her, her rating from this round. The round ends then.

Note: The other physicians and patients also see the ratings: As soon as at least **one** physician was rated by at least **one** patient (i.e. at least one interaction with a physician has been rated), **from the fifth round onwards** [in the subsequent periods] all patients see the **public feedback** of that physician (i.e. the average value of the ratings from all patients per physician) when asked for their decision 1.

Furthermore, **starting in round five** [in the subsequent periods], physicians see their own **public feedback** and the public feedback of the other physicians in their group when asked for their decision 2 & 3.

[Note: The other physicians and patients also see the ratings: As soon as at least **one** physician was rated by at least **one** patient (i.e. at least one interaction with a physician has been rated), **all** patients see the **public feedback** of that physician (i.e. the average value of the ratings from all patients per physician) in the following rounds when asked for their decision 1. In addition to the public feedback, patients see their **private** average rating per physician (if at least one interaction with the respective physician has already taken place with subsequent rating).]

Physicians see their **private** average rating for each of their patients at decision 2 & 3 (if there has been at least one interaction with the respective patient) in the following rounds. Besides, physicians also see their own **public feedback** and the public feedback of the other physicians in their group.]

Payoffs

I) No interaction (Patient decides not to consult the physician)

If the **patient** ends the period in decision 1 (decision "no" of the patient), then he receives **-4 points** in this period, i.e. he makes a **loss** of 4 points. If **no patient** in a given round **consults a physician**, the round ends for her, and she receives a **payoff** of **zero** points.

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

II) Interaction (Patient decides to consult the physician)

The **physician** receives the **price** (in points) chosen in decision 3 **minus** the **costs** of the treatment chosen in decision 2 for each of her patients.

For the **patient**, the payoff depends on whether the treatment cured the patient's condition.

- a) The treatment has cured the disease. The **patient** receives **10 points minus** the **price** demanded in decision 3.
- b) The treatment has not cured the disease. The **patient** must **pay the price** demanded in decision 3.

Two examples to illustrate this:

Example 1:

- The patient decides to consult a physician (Do you want to see a physician in this round = “yes” in decision 1).
- The patient has a major condition.
- The physician chooses a major treatment and charges the price for the major treatment.

$$\text{Payoff patient: } \underbrace{10}_{\text{benefit treatment}} - \underbrace{8}_{\text{price major treatment}} = 2$$

$$\text{Payoff physician: } \underbrace{8}_{\text{price major treatment}} - \underbrace{6}_{\text{cost major treatment}} = 2$$

Example 2:

- The patient decides to consult a physician (Do you want to see a physician in this round = “yes” in decision 1).
- The patient has a minor condition.
- The physician chooses a major treatment and charges the price for the major treatment.

$$\text{Payoff patient: } \underbrace{10}_{\text{benefit treatment}} - \underbrace{8}_{\text{price major treatment}} = 2$$

$$\text{Payoff physician: } \underbrace{8}_{\text{price major treatment}} - \underbrace{6}_{\text{cost major treatment}} = 2$$

The patient and the physician will be informed at the end of each period about their respective payoffs in this period. Besides, the patient learns whether his illness has been cured.

At the beginning of the experiment, you will receive an **initial endowment of 11 points**. You will also receive another **5 points** for **answering the control questions**. From this initial endowment, you can pay for possible losses in individual rounds. Losses can be compensated by winnings from other rounds as well.

At the end of the experiment, four periods will be drawn randomly for payment. For the calculation of payoffs, the initial endowment and the profits or losses over the four payoff-relevant periods are added together. 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 that it is **always** possible to avoid losses in the experiment with certainty. The total number of points will be exchanged for cash at the end of the experiment using the following exchange rate:

1 point = 60 Euro-Cent

(i.e. 5 points = 3 Euro).

You find the experimental receipts on your table. At the end of the experiment, please insert your payoff from the experiment (which you can see on your final screen) on the receipt as well as your first and last name in block letters and sign the receipt.

Control Questions

Here we show the control questions for **Rating** and **Exp+Rating**. Only if the question in **Exp+Rating** differs from the one in **Rating** we reported report it and underline it.

It is important to make sure that all participants have fully understood the experiment. Should something has remained unclear, please ask the experimenter. You will receive 5 points (= 3 Euro) for answering the questions correctly. Please answer the following questions:

Question	Correct Answer
1. How many decisions does a patient maximally make per period?	2
2. How many decisions does a physician maximally make per period?	2
Assess whether the statements below are true or false.	
3. "The patient learns what illness he suffers from in a particular period."	F
4. "If the physician cures the patient's illness, the total payoff of the patient in this period is exactly 10 points. "	F
5. "Your initial endowment of 11 points is worth 6.60 euros."	T
6. "A physician can identify a patient through the order of line-up over the rounds. That means, for example, that the first patient in the line-up is always the same person."	F
<u>6. "The number of identification (1-4) of patients and physicians are fixed throughout the experiment, i.e. patient (physician) 1 is the same participant in every period."</u>	T
7. "A patient can identify a physician throughout the periods by the order in which they are presented to him, i.e. for example, that the first physician in the list is always the same person."	F
<u>7. Assume you are a patient and rate your interaction with a physician. Who sees your rating within your group (of 4 patients and 4 physicians)? (only one answer is correct)</u>	All physicians and patients
8. From the fifth period onwards, all physicians and all patients within the group (of 4 patients and 4 physicians) see the average rating of those physicians already rated as they make their decisions.	T
<i>There was no similar question to question 8 in Exp+Rating. The control questions proceeded with question 9 (as question 8)</i>	
Please calculate the payoffs for the patient and the physician in the following examples	
9. The patient chooses "No" in decision 1.	Patient: -4 Physician: 0
10. The patient chooses "Yes" in decision 1 and chooses a physician. The patient suffers from a minor illness. The physician chooses a minor treatment and charges the price for a minor treatment.	Patient: 7 Physician: 1
11. The patient chooses "Yes" in decision 1 and chooses a physician. The patient suffers from a minor illness. The physician chooses a minor treatment and charges the price for a major treatment	Patient: 2 Physician: 6
12. The patient chooses "Yes" in decision 1 and chooses a physician. The patient suffers from a major illness. The physician chooses a minor treatment and charges the price for a major treatment.	Patient: -8 Physician: 6

Appendix F Experimental Instructions for additional games and questionnaire

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 make a decision 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 with 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:

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

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 <input type="radio"/> B <input type="radio"/>	EUR 0.60
2		A <input type="radio"/> B <input type="radio"/>	EUR 1.20
3		A <input type="radio"/> B <input type="radio"/>	EUR 1.80
4		A <input type="radio"/> B <input type="radio"/>	EUR 2.40
5		A <input type="radio"/> B <input type="radio"/>	EUR 3.00
6		A <input type="radio"/> B <input type="radio"/>	EUR 3.60
7		A <input type="radio"/> B <input type="radio"/>	EUR 4.20
8		A <input type="radio"/> B <input type="radio"/>	EUR 4.80
9		A <input type="radio"/> B <input type="radio"/>	EUR 5.40
10		A <input type="radio"/> B <input type="radio"/>	EUR 6.00
11		A <input type="radio"/> B <input type="radio"/>	EUR 6.60
12		A <input type="radio"/> B <input type="radio"/>	EUR 7.20
13		A <input type="radio"/> B <input type="radio"/>	EUR 7.80
14		A <input type="radio"/> B <input type="radio"/>	EUR 8.40
15		A <input type="radio"/> B <input type="radio"/>	EUR 9.00
16		A <input type="radio"/> B <input type="radio"/>	EUR 9.60
17		A <input type="radio"/> B <input type="radio"/>	EUR 10.20
18		A <input type="radio"/> B <input type="radio"/>	EUR 10.80
19		A <input type="radio"/> B <input type="radio"/>	EUR 11.40
20		A <input type="radio"/> B <input type="radio"/>	EUR 12.00

OK

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.

Part 5

Assume you randomly chosen to be in the role of player A. How much of your endowment (EUR 4) are you willing to send to player B in this case?

Send to player B

Now assume you randomly chosen to be in the role of player B. You have an endowment of EUR 4.

In the table below you see all possible amounts you could get from player A. Decide for every situation how much you want to send back to player A, had you received this amount.

Assume you received ... from player A (already tripled amount)	then I send ... of it back to player A	Your payoff and payoff of player A
0.0	<input style="width: 100%;" type="text"/>	
1.5	<input style="width: 100%;" type="text"/>	
3.0	<input style="width: 100%;" type="text"/>	
4.5	<input style="width: 100%;" type="text"/>	
6.0	<input style="width: 100%;" type="text"/>	
7.5	<input style="width: 100%;" type="text"/>	
9.0	<input style="width: 100%;" type="text"/>	
10.5	<input style="width: 100%;" type="text"/>	
12.0	<input style="width: 100%;" type="text"/>	

How well do the following statements describe your personality?

I see myself as someone who ...	Strongly disagree	Rather disagree	Neither	Rather agree	Strongly agree
... is reserved	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... is generally trusting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... tends to be lazy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... is relaxed, handles stress well	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... has few artistic interests	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... is outgoing, sociable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... tends to find fault with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... does a thorough job	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... gets nervous easily	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... has an active imagination	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate your gender:

- ☐ Female
- ☐ Male

How old are you?

Which field of study are you in?

Which subject do you study?

(If you are doing several studies, please indicate all and write the study program in parenthesis)

What semester are you in?

What was your average monthly net income over the last year, taking into account all sources of income such as scholarships, student loans, earned income, parental financial support, et cetera? Please round to the nearest ten Euro.

How often do you practice your religion?

- ☐ often
- ☐ rarely
- ☐ never

Please enter here your (average) grade from the Matura/Abitur certificate.

Which grading scale was used in your Matura/Abitur certificate?

- ☐ 1-5 (5 worst rating)
- ☐ 1-6 (6 worst rating)
- ☐ 1-10 (10 best rating)
- ☐ 0 -15 (15 best rating)
- ☐ 0 -100 (100 best rating)
- ☐ other (please specify including explanation)

How do you rate your past average school achievements compared to your former classmates? Answer on a scale from 0 -100 (0 you are the worst student in the class, 100 you are the best student in the class).

How many times have you visited a physician in the last 12 months (including all routine check-ups at the general practitioner, dentist, etc.)?

Have you ever had the impression that a physician is performing more or fewer treatments than necessary or is charging for services that he has not provided?

- ☐ Yes
- ☐ No

Have you ever rated a physician or recommended one?

- ☐ No feedback/recommendation
- ☐ Private feedback to physician
- ☐ Feedback through rating platforms for physicians
- ☐ Recommendation to a friend

- o Other (please specify including explanation)

Have you ever requested a recommendation for a physician?

- o Never requested a recommendation
- o Private recommendation from a friend
- o Looked it up on rating platforms for physicians
- o Other (please specify including explanation)

Were the instructions clear and understandable for you? What could be improved?

Do you have any other comments for us?